

Building and Sustaining Community Online

An Exploration of the Early Life of Online Communities

John J. Robinson

A document submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

UNIVERSITY OF WASHINGTON

2026

Reading Committee:

Kate Starbird, Chair

Sean Munson

Mark Zachry

Program Authorized to Offer Degree:

Human-Centered Design and Engineering

©Copyright 2026

John J. Robinson

University of Washington

ABSTRACT

Building and Sustaining Community Online

John J. Robinson

Chair of the Supervisory Committee:

Kate Starbird

Human-Centered Design and Engineering

Many people participate in online communities. They ask questions in a Facebook group, learn makeup tips on TikTok, or share parenting advice on Reddit. People can connect to people across the globe or help mobilize and organize neighborhoods. However, many communities fail to sustain themselves over time. Creators, participants, and even the platforms have a vested interest in helping communities meet their individualized goals.

This work aims to expand our understanding of the participants' behaviors, which participants do various types of "work", and how communities evolve over time. I take a historical, mixed-methods approach to look at the life of online communities on Reddit, a popular platform that hosts many online communities. This approach enables me to look at longer time-scales of the evolution of these com-

munities, and explore how behaviors differ between subreddits and across time. I first explore different behaviors across eight subreddits, which I use to ground later analyses within the observed data. I then explore some of these behaviors more across a broader sample, and who does some of the behaviors. Lastly, I use some of the behaviors in a new context to compare subreddits used during a political campaign.

My findings make theoretical and empirical contributions to research on online communities in human-computer interaction. I describe behaviors that occurred across a sample of 8 subreddits. These behaviors serve as the basis for a theoretical framework which is useful for comparing and understanding the evolution of online communities. My empirical findings suggest that behaviors and community characteristics differentially predict commonly used success metrics. I find that both creators and participants perform this community-building work.

This work contributes to scholarship around behaviors and community characteristics that impact different measures of success. This work expands our understanding of some community-building behaviors and how these may be employed by creators, moderators, and everyday participants to shape or maintain their community.

1 ACKNOWLEDGMENTS

This meager dissertation has a long history that is the result of a large amount of encouragement and support throughout my life. I am an incredibly difficult person to support, and so I am even more thankful for the time and effort so many people invested in me. This came in many forms and from many different people, and I would not be writing this today without them. I will undoubtedly miss some people here, but know that I appreciate you nonetheless.

I want to extend my deepest thanks to the late Jane Lowe Davis. She provided some critical support, guidance, and a large amount of encouragement for many years. My current life, my career, and this dissertation would have been possible without her. I would never have left my town without her support.

To my advisor, Kate for whom I unintentionally made things more difficult (and also myself), and yet she still continued to support me and my work. Thank you for allowing me to explore, fail, learn, fail again, and grow. I am also thankful for endless advice and for trying to keep me scoped. I am also thankful for working through the hanging sentences and typos – If I had seriously edited this too early the entire document would have been deleted.

A sincere thank you to all of my committee members, past and present. I am incredibly thankful Sean for always being willing to talk and offer great, tailored advice and a number of good laughs. Thank you also to Mark for stretching my thinking about this work in a different direction. Thanks also to David Ribes, Benjamin Mako Hill, and David McDonald for the incredibly insightful feedback and for expanding my thinking and approach.

I am also incredibly thankful for several professors who helped me in this journey. First, Cecilia Aragon accepted me into her research group despite me having no experience. She helped teach me the basics of research, and encouraged me to think bigger. Susan Joslyn helped me learn the ropes of research, and tried to provide so many forms of support throughout this journey. Her advice helped guide me and keep me on track through being ABD. Jennifer Turns also offered very thoughtful advice, feedback, and encouragement.

Throughout my time at the University of Washington, I want to thank a number of other people for their encouragement and support. This includes Katie, Michael, Daniel P, Jim, Megan, Andrew, and Daniel E, all of whom helped me learn, provided some laughs, and encouraged me to continue. Dharma always had incredible advice and a realistic perspective on the world that helped keep me grounded. I will always remember the trips to Darrington and one of the most difficult interviews I've ever done. Mia always had great jokes to share, and provided many great conversations. Tom provided some new perspective that helped change my approach to problems.

I am also grateful for the support of my PhD cohort. Elena was always happy to listen and provide support, and for occasionally getting out of my comfort zone! Her friendship was invaluable during this. I learned so much from Rafal and enjoyed our conversations and laughs. Cindy always had invaluable insights, often a joke, and she helped me understand new ways to do research for good. Ahmer and Rida provided some of the most valuable conversations about research and life in general, and for giving me something to aspire to. I also want to thank Lucas, Andrew, and Kristin for the many great times and much needed breaks.

My chosen family in Seattle was critical for making it through this work and an endless amount of support. I am indebted to Emily for her unwavering support through this journey and beyond. She has always had my back, and supported me throughout the darkest times I've had, and has been a source of regular discussions and debates that offer much needed breaks and interesting perspectives. Similarly, thanks to Jon for his continued support, laughs, and keeping me well-fed with the best of food. I also want to thank Tim and Jamie for their support and friendship. The laughs, the food, and the wine have been highlights. Angel provided a regular source of chill laughs, honest advice, and a regular source of inspiration. The reality tv hangouts and quests for increasingly spicy food provided so many good times and much needed breaks.

I also want to thank the other side of my chosen family: Amanda, Shannon, Liz, Genya, Lee, Ina, and Mikaela. Each of you have provided so much incredible support for this work, and for me as a human. Even as I got somewhat distant throughout this journey, you were there to inspire and encourage me

to rally not just for this, but to *do something* about the state of everything. I appreciate the laughs, the insights, calm that you provide, and above all else, the amazing food we share. I strive every day to give back to world a fraction of the love and support I have received from each of you.

Kuan-Chung provided many conversations, even through some rough times. I have been inspired by your research and ability to focus, and treasure our conversations. I am thankful to you for years of helpful conversations, advice, and your support.

To Pam and George, I am thankful for all of your kindness, generosity, and support for so many years.

Lastly, I want to thank my family, especially the ones I've lost while working on this dissertation. First and foremost, my mom sacrificed so much more than I knew to help me get to where I am today. Her love and support made all of this possible. I love you. I also want to thank and acknowledge Will, Lisa, Bonnie, Kristina and my dad for their support.

CONTENTS

1	Acknowledgments	v
1	INTRODUCTION	3
1.1	The Birth of Subreddits	4
1.2	Online Communities	7
1.3	Space, Place, Placemaking, and other Terminology	10
1.4	Central Questions & Contributions	11
1.5	Reddit as a Site of Inquiry	13
1.6	Positionality	16
1.7	Chapter Outline	19
2	BACKGROUND & RELATED WORK	21
2.1	Online Community: A definition	21
2.2	Defining Community	23
2.2.1	The Space, Place, and Locality of Community	26
2.2.2	Criticism of Community and Online Community	29
2.2.3	Summary	32
2.3	Online Communities	32
2.3.1	Why Do People Participate in Online Communities?	33
2.3.2	Participating in Online Communities	35
2.3.3	Encouraging Contributions and Commitment	36

Contents

2.3.4	Membership	37
2.3.5	Governance	39
2.4	People, Bots, Alts, and Accounts	41
2.4.1	Multiple Accounts	41
2.4.2	Bots	41
2.5	Evolution of Online communities	42
2.6	Existing Frameworks (and Models and Characterizations)	45
2.6.1	Multi-Dimensional Frameworks & Models	46
2.6.2	Characterizing Attributes	47
3	METHODOLOGY & DATA	51
3.1	HCI & CSCW	51
3.2	Methodology	53
3.3	Grounded Theory	53
3.4	Sampling	55
3.5	Trace Ethnography	55
3.6	Data	57
3.6.1	Relative Months	58
3.7	Historical Analysis	59
3.8	Methodological Summary	60
3.9	Limitations	60
3.9.1	Data Limitations	60
3.9.2	Methodological Limitations	62
3.9.3	Data Verification & Mitigation	64
3.10	Ethical Considerations	65

Contents

4	STUDY 1 - BEHAVIORS DURING THE EARLY LIFE OF ONLINE COMMUNITIES	66
4.1	Methods	68
4.1.1	Data & Data Preparation	69
4.1.2	Subreddit Selection Criteria and Dimensions	70
4.1.3	Selected Subreddits	75
4.1.4	Analysis	76
4.2	Overview of the subreddits	77
4.3	Results: Thematic Characteristics from the Sampled Subreddits	80
4.3.1	Content Composition	80
4.3.2	Submission Types (Link vs Text)	81
4.3.3	Fostering	88
4.3.4	Promotion / Advertising	95
4.3.5	Interaction	105
4.3.6	Submission Interactions	106
4.3.7	Participation	114
4.3.8	Summary of Results	126
4.4	Discussion	127
4.4.1	A framework for online communities	128
4.4.2	When is community?	130
4.4.3	Reddit Communities as resources for questions	132
4.4.4	Fostering	134
4.5	Limitations	136
4.6	Conclusion	136
5	FOSTERING AND PROMOTION DURING EVOLUTION	137
5.0.1	Success in Online Communities	139
5.1	Data & Methods	140

Contents

5.2	Predicting Growth and Retention	141
5.2.1	Methods	141
5.2.2	Measures of Success	143
5.2.3	Predictive Features	144
5.2.4	Results	145
5.2.5	Discussion	147
5.3	Predicting Remaining Lifetime	149
5.3.1	Data & Survival on Reddit	149
5.3.2	Methods	150
5.3.3	Results	150
5.3.4	Discussion	151
5.4	Who does the work?	151
5.4.1	Data	152
5.4.2	Fostering	156
5.4.3	Promotion	158
5.5	Discussion	161
5.6	Limitations	164
5.7	Conclusion	165
6	PROMOTIONS AND FOSTERING DURING A POLITICAL CAMPAIGN	167
6.1	Introduction	167
6.2	Data	170
6.3	Characterizing r/The_Donald	172
6.4	When and How Did r/TD Grow?	174
6.5	Joins by Level of Tenure Engagement	176
6.5.1	Data	176
6.5.2	Results	177

Contents

6.6	Are Promotions Related to Account Joins?	178
6.6.1	Methods	179
6.6.2	Results	179
6.7	Where Were New Accounts Active Prior to Joining r/TD?	182
6.7.1	Methods	183
6.7.2	Results	183
6.8	Discussion	187
6.8.1	Limitations	188
6.8.2	Conclusion	189
7	CONTRIBUTIONS & DISCUSSION	191
7.1	A Framework for Characterizing Online Communities	191
7.2	Predicting "Success" & Who Does the Work	194
7.3	Promotion and Fostering in a Political Subreddit	196
7.4	Surfacing Invisible Work	197
7.5	Tailoring Success	199
7.6	Placemaking in Online Communities	201
7.7	Implications and Future Work	203
7.8	Reflections on this Work's Relevance Today	206
8	APPENDIX A	208
8.1	Qualitative Code book	209
9	APPENDIX B	216
9.1	Methodological Notes	216
9.1.1	Deleted and removed content	216
9.1.2	Scores	217
9.1.3	Mention and Link Regular Expressions	218

LIST OF FIGURES

4.1	Subreddit Sidebar Links Example	72
4.2	Network Graph of Subreddit Categories	74
4.3	Sampled Subreddit Monthly Activity	79
4.4	Subreddit Link and Text Submission Percentages	81
4.5	Submissions Created by selected accounts for r/UpliftingNews	89
4.6	Submissions Created by Selected Accounts for r/AskWomenOver30	91
4.7	Submissions Created by Selected Accounts for r/LinusTechTips	93
4.8	Example Cross-post Image	99
4.9	Network Graph of Cross-posts	102
4.10	Graph of Submissions in r/UpliftingNews with Engagement Types	107
4.11	Graph of Submissions with Replies for all Subreddits	108
4.12	Graph of Comment Depths for Subreddits	111
4.13	Graph of Retention Rates Over Time	117
4.14	Graph of Retention Rates for r/UpliftingNews	117
4.15	Graph showing Subreddit Tenures	121
4.16	Graph of Gini Coefficient Over Time	124
5.1	Subreddit meeting activity thresholds	142
5.2	AUC for retention and new users growth	146
5.3	Subreddit activity over time	152

List of Figures

5.4	Fostering Activity over time by account types	155
5.5	Time to reach 100th account comparison for fostering	157
5.6	Promotional content over time	159
6.1	Graph of r/The_Donald active accounts over time	171
6.2	Graph of promotions and fostered submissions for political subreddits	173
6.3	Graph of r/The_Donald daily joins	175
6.4	Graph of The_Donald joins by tenure	177
6.5	Graph of promotions and joins per day	179
6.6	Graph of average hourly joins relative to promotions	180
6.7	Graph of joins from conservative accounts	185

1 INTRODUCTION

Many people participate in online communities (M. Anderson and Jiang, 2018; Lieb et al., 2025; Noveck et al., 2021; Reddit and GlobalWebIndex, 2020). These online communities are places that exist on the Internet where people interact, and may feel an attachment to the space and the people (e.g. A. Blanchard and M. Markus, 2002). There are many different forms of online communities that serve a wide range of purposes. These include Facebook Groups, Nextdoor neighborhood groups, Reddit's subreddits, and Discord chats. People's participation in them can also take many forms including sharing information, discussing special interests, asking and receiving advice, providing support, reviewing products, creating content, and trading goods.

However, starting and maintaining online communities remains a challenge. Researchers have identified some interventions that have helped some communities to grow and survive over time (e.g. Kraut and P. Resnick, 2012), but challenges still remain. For example, on reddit, many subreddits may never achieve or sustain enough activity to be perceived by members as a community¹.

In this work, I present a historical analysis of behaviors that occurred across multiple communities on reddit. I cast some of these behaviors as placemaking behaviors, which are intended to develop a community from nothing. These behaviors may also stimulate the regular participation within the community, which may be important for the development of the feeling of a *sense of community* (McMillan and Chavis, 1986). My empirical findings here suggest that some placemaking behaviors may be effective in helping communities grow or sustain themselves. My research also suggests that everyday participants (e.g.

¹I explore this in more detail in Chapter 5.

not creators or moderators) may be effective in doing this work, and ultimately help some communities succeed.

1.1 THE BIRTH OF SUBREDDITS

In 2008, reddit added a new feature that enabled users to create their own subreddits, or topical spaces. At the time, reddit was a growing social news aggregation platform and subreddits were spaces within the platform focused around specific interests or topics. Prior to this feature, only platform administrators could create subreddits and as such the topics were generally broad (e.g. politics, programming, etc). This change subreddits enabled any user to create a community space for sharing and discussions about any topic or experience they wanted, which led to a rapid increase of newly created subreddits. Subreddits enabled the platform to become a multi-community platform, similar to Usenet - a distributed service for communicating in specific spaces called newsgroups. ².

The user-created subreddits had many different trajectories. Some went on to be wildly popular with millions of members, and others saw little or no activity. Users also created subreddits for many purposes. These varied not only in their growth patterns, but also in the diverse set of purposes for which they were created. For example, some subreddits were created simply as a place to share and discuss a particular topic like r/Science or to share specific types of content like r/UpliftingNews and r/Videos. Some allow only certain types of content like images, video, or text. A few subreddits eventually created new subreddits to separate specific content like questions into separate subreddits (eg. r/AskHistorians and r/AskScience). Users also created subreddits as a form of collective action like the r/FindBostonBombers which went horribly awry as users attempted to find the two bombers from the 2013 Boston Marathon bombing.

Some subreddits have developed their own norms and cultures (e.g. Fiesler, McCann, et al., 2018), and some of their behaviors and content have spilled into the rest of the internet. Redditors created r/The_Donald to support Donald Trump in the lead-up to the 2016 presidential election, which created

²Usenet is a distributed service for sharing messages created in the 1980s which has many different spaces, *newsgroups*, where people can interact through posting messages and replying. Like subreddits, some newsgroups evolved into online communities (e.g. Baym, 1993)

1 Introduction

a number of specific norms around posting and comments³, and was also a pocket for some of the culture created around Trump's candidacy and presidency that spanned multiple platforms and online communities (Peck, 2020; Timberg and Harwell, 2021).

The diverse outcomes of many subreddits gives rise to questions about their evolution. What factors affected how these subreddits developed? Are there certain behaviors which shaped how these subreddits developed? In particular, are there early behaviors that members engage in to help establish the community? If so, how do these affect the evolution of the community? Are there behaviors which subreddit creators can do to help communities achieve different goals for the community? And lastly, what role do different members play in the development of the community?

There has been a renewed interest among scholars in studying the formation of online communities (Harris et al., 2019), especially as technology, online platforms, behaviors, and the world continue to evolve. People are leveraging more loose-knit forms of community without formalized spaces, especially on platforms like TikTok and YouTube, where fewer affordances for community building may exist (Fiesler, 2022; Klassen et al., 2021) or when sharing less socially desirable content such that more anonymity may be important (Chancellor, Pater, et al., 2016; Jhaver, Chan, and Bruckman, 2018). Boundaries of these online communities may be blurry, and some communities can exist between spaces and platforms (Jhaver, Chan, and Bruckman, 2018; C. E. Smith et al., 2022), and even between online spaces and the real-world.

Throughout their lives, online communities can face many challenges. Many online communities are not active months after their creation (Kraut and Fiore, 2014) and many factors can affect the continued activity of the community including competition and topical overlap (Zhu, J. Chen, et al., 2014). Growth of the community presents a number of challenges as well, including managing unwanted, low-quality, and harmful content. This is especially challenging as moderators are typically unpaid volunteers, who face harassment in addition to their work (S. A. Gilbert, 2020).

³It should be noted that some of this also existed and in some cases pre-existed on 4chan. Some have claimed that reddit effectively helped bridge some of this content to the mainstream from 4chan (e.g. [World War Meme 2017](#)). The subreddit r/The_Donald had a number of specific norms and rules despite being a space to amplify memes and culture from 4chan.

1 Introduction

Online communities can differ both in terms of their structure and how they function (e.g D. Fisher, M. A. Smith, and Welser, 2006). Some communities are focused on supporting discussion, asking questions, sharing news, and producing content. The intended function of online communities may be set by the founders of online communities, which may influence the evolution of the community (Kraut and Fiore, 2014). In addition to the function and structure, founders of online communities like subreddits may have diverse goals which may not include becoming large, popular communities (Foote, Gergle, and Shaw, 2017; Hwang and Foote, 2021).

Questions about why some communities fail continue to be relevant. Changes to platforms like X/Twitter and Reddit have caused many people to look elsewhere for information and connection. In the wake of Elon Musk purchasing Twitter, many people once again tried to flee to other platforms like Mastodon, Bluesky, and Facebook's Threads with decentralized services being a recent technological goal⁴. Similarly, reddit members protested API changes few years ago that would have potentially incurred significantly more work for the volunteer moderators of many communities. A number of communities have attempted to migrate to new platforms, seemingly with limited success. This work is relevant for the challenges faced by these platforms, and the communities that may live within them.

Understanding different forms of community and the underlying processes of their development may lead to better design interventions and enable more communities to achieve their personalized goals. In this work, I explore questions about the early behaviors in these subreddits and how they differ across different communities. I am interested in how these behaviors affect the evolution of these spaces into online communities. Of particular interest here are the behaviors that attempt to transform the space into a place. This means a focus on the who and what of the work behind developing and growing an online

⁴as opposed to the centralized services popular during Web2.0, where companies like Meta/Facebook controlled every aspect of the platform and its access. Decentralized services are focused around open protocols such that any entity can create an instance that connects to the network. This works similar to an email server or former newsgroups. While this sounds ideal, there are logistical challenges with building and supporting servers, especially for social media. Platforms like Mastodon and Bluesky have had some success, but some of the distributed infrastructure these hope to establish can have restrictive barriers to entry for people to setup and maintain servers.

1 Introduction

community. I build upon existing research of online communities, and contribute to both theory and practice around the behaviors present and their value in contributing to evolving into a community.

I focus, in particular, on behaviors that can be described as placemaking behaviors. These are behaviors that may be effective in helping develop a space, such as a subreddit, into something people think of as a place and ultimately a community. One behavior I focus on is *fostering*. I define this a continued effort by an individual or group that attempts to build or maintain a sense of community among other participants. I focus on how content curation and content creation can be viewed as placemaking behaviors, but other examples include moderation or and platform design. These are similar to the act of building a house or decorating it, such that it may eventually become a home. I view these behaviors through a distinction between space and place as outlined within CSCW by Harrison and Dourish (1996), where we see the place is the “understood reality” of a particular space. These behaviors attempt to contribute to building a “recognizable and persistent social meaning” (Dourish, 2006) through the accounts continued effort of shaping and reshaping a subreddit space. Many communities fail to see activity, meaning that every subreddit does not develop into a community. However, these behaviors may facilitate the evolution and ultimately affect the social understanding of them.

1.2 ONLINE COMMUNITIES

Online communities are interesting socio-technical phenomena. At their core, online communities are a *place* for people to connect and *interact with other people*, and there are many forms of interaction that can *provide value* to people who visit them. For decades, some people have positioned them in a utopic light, focusing largely on the benefits that people can reap from online spaces focused around interests and unbothered by structural problems of the *real* world. In a 1968 paper, Licklider and Taylor imagined the possibility of online communities to connect large groups of people over similar interests (Licklider and Taylor, 1968). Some early Internet visionaries saw an opportunity to create spaces from tyrannies of the real world governments and corporations (Barlow, 1996).

1 Introduction

The growing popularity of online communities suggest users do find some value from participating in online communities. As of 2020, over 1.8 billion people using Facebook groups every month (Facebook, 2021). Reddit claimed over 121 million people are active daily and over 100k active communities as of May, 2026 (Reddit, 2026). The popularity has even led some to muse over the possibility of online communities being suitable replacements for traditional communities (Putnam, 2001), something which continues to be debated.

Two possible reasons for the growing popularity are the different benefits they provide to people and their ability to bring people together. The benefits people get can vary between individuals and communities based on their relationship and how they participate. Some people visit online communities to seek information about specific topics or interests (Nonnecke and Preece, 2001), some stay for the entertainment they provide (C. Lampe, Wash, et al., 2010), and some participate for the social support they receive within them (Fiore, Tiernan, and M.A. Smith, 2002). Many traditional groups and organizations have also created online spaces to further engage and keep people participating, like some local housing communities (Baborska-Narozny, Stirling, and Stevenson, 2016).

Although many people may find value in online communities, online communities have also faced increasing scrutiny over the years. Early literature often focused on positive aspects, and some made assumptions about the importance of pro-social values for growth. However, some communities may be less pro-social, sometimes eschewing some traditional values and yet still thrive (Bernstein et al., 2011). Anti-social behaviors have been observed in some highly visible online spaces in some recent years. This includes harassment, threats of violence, and organizing the January 6th storming of the US Capitol building. In 2020, reddit banned the subreddit r/The_Donald (Newton, 2020), which some had argued should have been shut down much sooner for amplifying hate and violence (u/spez, 2020). Facebook also banned a number of communities after the January 6th insurrection for violating many of their policies (Silverman et al., 2022). Kiwifarms, a forum known for targeting and harassing transgender people was dropped and blocked by one of their service providers, Cloudflare, because of escalating threats (Collins and Tenbarge, 2022).

1 Introduction

Online communities face many other challenges around attracting people and motivating participation (Kraut and P. Resnick, 2012). Building and managing online communities can be challenging work. Many of the spaces created on internet platforms are often not active after a few months (Kraut and Fiore, 2014). There are many factors which can affect how communities develop or evolve if at all. Many people creating communities may lack the knowledge of what they need to do to attract members and keep them engaged.

The design of the platform also plays a role in how a community develops (Kraut and P. Resnick, 2012; Preece, 2000). The underlying design helps build affordances for what users within the software, and how easy it makes it to manage the software. Many platforms rely on volunteer moderators, and the early founders are generally the first moderators of the community. When moderating or managing the community is difficult or time-consuming, the founder may not be able to keep up with the work necessary to keep the community focused and following the rules. Even within some successful communities, moderation is still challenging.

A body of work exists around high-level challenges for building online communities (e.g. Kraut and P. Resnick, 2012; Preece, 2000). A number of findings suggest gaps may still remain in our knowledge of how communities form, founders goals, and what success means to different community members. Work around founder goals suggests that some founders create online spaces without the intent of becoming a community or attracting other members (Foote, 2019) and some people prefer the intimacy of the smaller communities (Hwang and Foote, 2021). Open questions still remain around the goals both members and founders may have beyond growth, and what behaviors they can do to achieve or sustain those.

Some scholars have also been critical of how we conceptualize success in online communities. One study found that a number of popular success metrics for online communities are only weakly correlated, suggesting that success can look different even across communities (Cunha et al., 2019). Massimi et al. (2014) argue that we conceptualize success differently as some forms of community can grow and shrink over time, as people engage and disengage and that this is normal.

1.3 SPACE, PLACE, PLACEMAKING, AND OTHER TERMINOLOGY

This dissertation is focused on the work people do to transform an empty space into an online community. On reddit, someone would create a new subreddit and it has no content, no members, and no interactions. A core part of this work looks at this potential transformation from empty space into a community, and I need to use different words to help capture this change. This is especially important since subreddits likely vary in the amount of time it takes to become communities.

For this work, I define an online community as a social structure where interactions are primarily mediated through the Internet. Communities are inherently social structures, so I believe it's important that someone feels a sense of attachment towards it or perceives it to be a community⁵.

However, I do not measure how people's perception or feeling of the subreddit. Thus, for most of this work I will refer to the spaces as 'subreddits', with the understanding that some of these may be perceived as communities, and some may not. When I use the word *community*, it is in the more general or aspirational sense. So assuming that a specific subreddit has become a community, or speaking broadly about just those subreddits that have become communities.

A related set of concepts are that of space and place. In virtual environments, both would have some sort of location, like a URL or a page in an app. A place is a space where people have constructed social meaning around it through use of the *space*. This draws on a line of research (e.g. Harrison and Dourish, 1996), which notes that the distinction is similar to the difference between a house and a home.

Thus, *placemaking behaviors* are those that attempt to contribute or build social meaning through usage. I use *placemaking* and *community-building* interchangeably. as in this context, I believe that the behaviors build towards both, though the perception as a place and the perception as a community may not happen at the same time.

⁵It is worth noting that this is a working definition for this work, and that it is possible that a community may not be confined to a single subreddit. This means there could be a community between different subreddits, but this adds significant complication to both the studies and how I describe this. My writing is too poor for me to effectively communicate these nuances throughout the paper to a general audience.

1.4 CENTRAL QUESTIONS & CONTRIBUTIONS

This work began out of frustration from a prior project, where I struggled to understand whether the behaviors I documented were unique to the subreddits I was studying at the time. I also had challenges understanding if some temporal changes were unique to the communities I was studying. These challenges persisted into some of my industry work, where understanding meaningful differences in communities continued to be a challenge. So the contributions are largely the product of me trying to understand the processes underlying development over time of communities, and to better contextualize the multi-community site of reddit.

I contribute to our knowledge of online community formation and development with both empirical and theoretical contributions. This work may be of value to researchers and practitioners in industry who work on platforms that host online communities. Additionally, the theoretical and methodological contributions may be of interest to other scholars, particularly in the fields of Human-Computer Interaction and Computer-Supported Cooperative Work. Both audiences share an interest in understanding the work done by people to build and sustain online communities, which align with this work's contributions.

As socio-technical phenomena, online communities have been a popular “site” of study for researchers within both of these fields. Platforms like Facebook and reddit provide people with a space to seek information and build relationships, while many of their behaviors are logged as traces for other people to see and interact with. The behaviors in an online community are mediated by the underlying technology (Preece, Nonnecke, and D. Andrews, 2004), and the system designs and interventions studied have long been studied within these communities. Scholars have studied a variety phenomena that occur in online communities on reddit including privacy and differential identity reveals (Leavitt, 2015; Wohn, Yuan, and Siri Jr, 2021), collective action against the platform (Matias, 2016), political action (Mills, 2018), covering breaking news events (Leavitt and Robinson, 2017b), and many more.

The study of online communities is interdisciplinary, bringing many areas of social science to help understand the intersection of how people interact with technology and the effects of technology. This includes the use of social science to develop design principles for building online communities (Kraut

1 Introduction

and P. Resnick, 2012). Some areas of research look more critically at the online platforms, some of which host online communities including the political aspects of the platforms that host some of them (Gillespie, 2010).

To better understand my goals for this work and its contributions, it would help to understand the central research questions this work explores.

1. *How do online communities develop on reddit?* How do behaviors between different types of communities differ? Which behaviors do people engage with in an attempt to sustain or grow communities?
2. *Which behaviors within communities predict different forms of success?* Who typically performs these behaviors on reddit?
3. *How do political communities experience growth of engaged users?* Do promotions impact their growth?

One contribution of this work are empirical descriptions of behaviors during community formation on reddit. These behaviors include descriptions of two novel sets of behaviors which can be used to *promote* subreddits and *foster* the sustainability or growth of a community. I also observe and describe the prevalence of questions across all forms of communities I studied, which I characterize and describe in detail.

Another contribution is an analytical framework made by organizing behaviors into similar categories. I refer to each of the categories as perspectives, and each attempts to characterize a specific aspect of a community. The framework enabled me to compare different subreddits and to compare a single subreddit over time to understand how changes might be affecting the community. This provides analytical value in both contextualizing communities within a platform, and identifying areas for future exploration. A key portion of the contribution here is the description of these perspectives and how they manifest across the subreddits and over their early lifetimes. Two of these perspectives, fostering and promotion, have been less discussed within the existing literature.

1 Introduction

Leveraging the framework, I also contribute empirical findings on how both promotion and fostering behaviors may be useful for predicting three common success metrics used in industry and other scholarly work. These findings may be useful for community designers to build systems that help people understand, monitor, and improve their chances of meeting their collective goals for their communities.

I also explore who does some of the work of promoting and fostering both within and outside of communities. I found that while creators generally contributed to communities less over time, and more successful communities had moderators and average people step up. In some cases, the founders were absent entirely and community members stepped up to foster the growth and eventual success of communities. This finding opens new opportunities for platforms and their designers to create new interventions to help struggling or dead communities, by encouraging interested participants to become fostering participants in those communities.

In the last study, I apply the framework to a new type of online community – political campaign subreddits. I find the framework useful for identifying interesting differences between the largest, most well-known political campaign subreddit, *r/The_Donald*, and other campaign subreddits active in the same 2016 election. Some of the comparisons are also grounded against all of the subreddits within the dataset, to help understand how these subreddits compared to other subreddits.

Lastly, I analyzed the temporal effects of promotions on the growth of new users in the political subreddit. I found that, in general, promotions can precede the participation of new fostering members within a community, but also observe that both occur tend to increase around large media events. In this particular case, this provides an area of future study for other researchers to do causal analyses which at a larger scale and can control or minimize the co-occurrence.

1.5 REDDIT AS A SITE OF INQUIRY

In this work, my site of inquiry is *reddit.com*, a popular social media platform focused around content sharing and communities. It is one of the most popular websites on the Internet as of 2022⁶. In 2022, a

⁶The former internet traffic tracking site Alexa ranked reddit as the 19th most visited site in August, 2021 (Alexa, [2021](#))

1 Introduction

similar ranking site, Semrush, ranks reddit 9th globally and 6th in the US (Semrush, 2022) for internet traffic. It was founded in 2005 as a social link-sharing site, which enabled people to submit links to external content and vote on the submissions, so that every day there is a 'front page' of the most popular content that day. It has also seen an increasing interest from scholars who have situated more work there (Proferes et al., 2021).

It is also home to over 100,000 active subreddits, which the company refers to as communities, and over 121 million unique daily visits⁷. However, at the time of its founding the platform was a simple link-sharing platform, which had no subreddits or even comments. As the platform developed, it began to add new features, some of which were the result of requests from users. When users interact with the platform, the content is recorded and displayed to other users. This provides a historical record which can enable us to explore a number of research questions about the evolution of online communities. When combined with historical data about the evolution of the platform itself, we can also explore questions around the co-evolution of the platform and the user behaviors on the site.

On reddit, the primary forms of interacting are submitting content, voting, and commenting. For example, users can submit a funny picture, other users then *upvote* that image *submission* if they like it, or *downvote* the content if they don't like it. They can comment on any of the submissions. If a user really appreciated the image submission, they can choose to *comment* and tell the *original poster* (OP) of the submission that they liked it. Other users can also comment on their comment as 'replies'.

Over time, the platform has evolved to support submitting content other than links including text, images, and video. The site has also evolved from early link-sharing to support a many social features including commenting on submissions, messaging, real-time chat, and a simple friending system.

Most relevant to this work was the introduction of subreddits in 2006⁸. Subreddits are effectively digital spaces which moderators and users can share content around specific topics or experiences. Initially subreddits were only created by platform administrators at the request of users. In 2008, the

⁷This is according to the latest numbers released by the company in 2026 Reddit, 2026

⁸The first subreddit introduced was r/NSFW (Not safe for work), which separated pornography and other inappropriate content from the site's front page so people could browse reddit at work. (shuffman56, 2006b)

1 Introduction

platform enabled the ability of users to create their own subreddits. Reddit claims there were over 100,000 active communities as of January 2021 (reddit.com, 2015)⁹.

Thanks to work from reddit user u/StuckInTheMatrix, a large archive of reddit submissions and comments exists from some of the earliest posts in 2005 (Baumgartner et al., 2020). The archive enables us to explore answers to potentially answer questions about behaviors and how these behaviors evolve over time. The dataset is a promising source of many forms of analysis, but it is not without limitations¹⁰.

In addition to the availability of a historical archive, reddit as a platform has changed over its lifetime. This enables researchers like myself to analyze the co-evolution of user and community behaviors and the platform as it changed. Much of this work is focused in a period of time where the platform was undergoing significant changes, such as the introduction of AutoModerator, while some of the largest changes had occurred years prior (e.g. comments, user-created subreddits, etc.).

I used the historical data to explore the evolution of communities. To explore the evolution, I needed to look at long periods of time, while keeping the project scope manageable by a single individual with modest hardware. I used data from the creation of the platform through the end of 2016, which included over 3 billion comments and posts. This enabled me to select subreddits past the introduction of user-created subreddits to provide time for members to adapt to the feature, while providing a minimum of several years of data for the analysis.

I also leveraged external sources like archive.org and the reddit blog, to help understand the evolution of the platform. This was important for identifying some platform changes or feature introductions which might have biased certain forms of analyses. Prior to the availability of reddit, one of the few platforms where these types of analyses were readily available was from Usenet (e.g. Fiore, Tiernan, and M. A. Smith, 2002; D. Fisher, M. A. Smith, and Welser, 2006; Kang and Shin, 2006). Other multi-community platforms exist such as Facebook and NextDoor but access for researchers remains a problem.

⁹They defined active subreddits as those which received 5 or more posts or comments in a single day

¹⁰Most notably is that the dataset is missing many traces of data, which may bias some questions and specific analyses (Gaffney and Matias, 2018). I attempt to call out these limitations with the analyses that may be affected with them.

1 Introduction

In the final study, I do a case study of an infamous political community on reddit that was created in 2015. The community ultimately was banned in 2020 (Newton, 2020; Peck, 2020), but was creating problems as far back as 2016 (Koebler, 2016). The infamy of the community, which continued to grow after the 2016 election makes it an interesting community to explore in part because of some of its influence which brought some behaviors and memes to more mainstream audiences.

The community's rise was one of many examples of behaviors from reddit communities extending beyond the boundaries of the platform. These forms of collective action are interesting phenomena because in many ways they resemble more traditional forms of collective action. They illustrate how online communities can in some capacity influence popular culture and potentially influence change on the platform and beyond. Another example was the protest of reddit firing of Victoria Taylor in 2015 and the ongoing lack of moderator tooling and support (Matias, 2016), which yielded an increased commitment for better tooling for moderators. During a number of breaking news events, many redditors have also participated in collective sense-making of the ongoing events (Leavitt and Robinson, 2017a), including most notably the aftermath of the Boston Marathon Bombings where a crowd of redditors misidentified someone as a perpetrator.

In this dissertation, I focus on these behaviors and how they collectively appear at the subreddit level and how we can understand the creation and evolution of these subreddits. For example, when a person creates a subreddit, what activity do they do to attract new users and where do they do this? How do subreddits full of people react to sudden changes within the platform? Despite its limitations, this dataset provides an ability to understand the evolution of these communities from both a qualitative and quantitative perspective.

1.6 POSITIONALITY

Research is a journey, and at every turn decisions are made which impact aspects of the final work. Many of the decisions I have made throughout this work have been shaped by my life through this point. This is especially salient as I draw upon grounded methodology to guide much of the inquiry, and specifically the

1 Introduction

interpretive methods. I subscribe to constructionist beliefs – specifically that knowledge is constructed rather than discovered. As such, it is valuable to understand more about me and some of the aspects which may have affected this work and the many decisions within this paper.

This work is political. Every decision has been shaped by ongoing world around me and conversations I have had with friends, colleagues, and my committee. This work has been developed during a time where data privacy, research ethics, and critical views of platforms and their products have been fore-fronted and to which I care deeply. I am keenly aware of how this work and the data I present within this work may have a life beyond this dissertation and uses beyond what I imagine, and this has also strongly shaped this work. I have made many decisions to avoid or remove certain research questions, exclude analyses, avoid certain data, and select certain quotes over others. My goal was not to introduce additional bias into this work, but instead to limit possible contributions which may be used to find or otherwise target individuals or groups. These choices will ultimately limit the academic value of this work, and it may also introduce gaps around which I do not address. I have dedicated a substantial amount of time and thought to think through these problems and to address some of these gaps or holes through other means, and by leaning on other scholarly literature to fill these.

Related, part of that journey and the resulting work, as with all research, is the product of negotiation with members of my community and committee. While the overall goal is to strengthen this work, in some cases it has taken away elements that I find important. This is not dissimilar from a peer-review process of negotiation of edits to have papers published. Sometimes this may ultimately distract from the work, while others may strengthen the core arguments. I am aware of this, but I am also practical and thus have chosen some battles to fight while giving into others.

Historically, I have participated in many forms of online communities since the 1990s. I have made many friends and acquaintances throughout the world from communities, some of which I developed multi-year relationships with. As such, I have a disposition that community can exist or be created in online spaces. Throughout this work, I remained aware of my belief, and have worked to find and incorporate a diverse collection of perspectives on the topic. This has led me to actively seek out critical

1 Introduction

scholarship for the types of relationship and question whether these are communities. Throughout the development of this work, I have sought to challenge my assumptions through these voices and used this to focus some research questions.

For me, the allure of these online spaces lies in my ability to have experiences and learn in ways I would not have had locally. I grew up in a poor, rural town in the southeast United States. I felt constrained both geographically, financially, socially, and intellectually. I have a certain perception of social institutions, norms, and power structures in society that has been shaped both by this experience and by growing up a gay, white man in a rural town. This has further shaped my perspective on online communities and has shaped the scholarship I have read and cited within this work. I am aware of power structures that both promote and deny voices within social spaces, and although this work is built entirely on trace data, I remain aware of the holes within this data, and have reflected on what some absences may mean for the larger data.

It is also worth noting that I have worked as a software engineer. This has affected the methods I have chosen and my approach to analyses. Through this experience, I am able to approach larger amounts of data and learn new languages and toolsets relatively easily. I am also aware of many potential errors that can arise through the use of these tools, and in many cases have developed scripts to check and correct these errors. That is not to say there are no mistakes contained within, but I have made numerous attempts to check and account for errors in databases and data manipulation scripts.

I have also lived across the US and in Europe for a number of years. This has given me a very different perspective on many approaches to living, governance, and power. This has given me a unique perspective on many forms and perspectives on community and participation. This has also instilled a curiosity for the world around me beyond my immediate surroundings and an openness to other ways of being and knowing. Throughout this project, I have sought in several occasions perspectives on community from literature beyond.

1.7 CHAPTER OUTLINE

I will now describe the structure of the remaining part of this work.

1. **Chapter 2** provides a look at some existing research around communities and specifically online communities which may provide helpful context for understanding this work. It also provides short look at some of the existing frameworks that have been introduced, to better understand the niche that this work fills.
2. **Chapter 3** provides an overview of my methodology and approach for the analyses within the work. It also provides a look at the data and data preparation itself along with some of the limitations of using this data for analysis.
3. **Chapter 4** presents the first study, which explores behaviors during the formation of online communities on a micro-level and looks for characteristics and behaviors between the different forms of communities on reddit. It also explores some of the work people do for the communities in question. From these behaviors, I develop a five-perspective framework that captures a number of characteristic themes from which other scholars and practitioners might leverage in understanding online communities.
4. **Chapter 5** focuses on two specific perspectives of the framework and attempts to understand their role in the growth and how they may affect the survival of online communities. I further explore these dimensions by attempting to characterize who does these forms of work within reddit and theorize about the implications this may have for researchers and practitioners studying online communities.
5. **Chapter 6** leverages the framework to extend an existing work on political subreddits and collective action. I analyze a large community that arose during an election cycle to explore how the subreddit grew over time. This includes a particular focus on the relationship between promotion and growth of different levels of engaged accounts, including the more active accounts.

6. **Chapter 7** presents a summary of the contributions along with a discussion of some of the more important findings from this work and how they relate to other findings from online communities.

2 BACKGROUND & RELATED WORK

In this chapter, I provide background research around online communities and their evolution. This is intended to provide additional context to understand the contributions of this work. I have structured this chapter in the following way. First, I review definitions of communities with a particular focus on some of the more commonly agreed upon attributes of communities. I then shift focus to some of the research on online communities. Third, I cover some of the more commonly studied topics related to growth and evolution of online topics including motivation and commitment. Last, to contextualize the framework this work builds I want to provide some background on different dimensions that other scholars have used to characterize groups and communities.

2.1 ONLINE COMMUNITY: A DEFINITION

I want to begin with a definition of *online community* specific to this work. I think of *online community* as a social structure where interactions are primarily mediated through the Internet (and thus through some form of computing hardware). An online community is inherently social and psychological, thus requiring interaction between people which may eventually generate feelings of value, importance, or obligation. Being social, this can obviously lead to the development norms and forms of culture. I have made this definition intentionally broad in an attempt to be less prescriptive and avoid some of the debate around the community definition.

The definition here is less important than the empirical findings of this work. Similar to other researchers, I find the definition less important than the results, especially since I do not act as a community

2 Background & Related Work

arbiter. However, this definition aligns with works that focus on the interactivity aspect (e.g. Bruckman and Jensen, 2002), with some level of psychological significance or attachment such that someone views it as a community (Baym, 1993; A. Blanchard and M. Markus, 2002; D. B. Clark, 1973; McMillan and Chavis, 1986; Prinster et al., 2024). I also recognize that the community may not be defined by the subreddit, but instead may be the network that forms through interaction. Instead the community may actually exist both within and beyond the borders of the community, as other studies have found (TeBlunthuis, Kiene, et al., 2022; Wellman, 1979; Wellman, 1987).

My specific unit of focus in this work is a subreddit, which is a particular space on reddit that users can create around a topic, identity, experience, location, or even a purpose. I believe that not all subreddits are communities. Prior research has found that people do find some parts of reddit to be valuable, and think of some parts of reddit as forms of community (Hwang and Foote, 2021; TeBlunthuis, Kiene, et al., 2022). People feel a sense of belonging from continued participation in these communities. Many subreddits have their own forms of norms and rules, and there are some which extend across communities.

I use the term “online community” to align with existing work (Baym, 1993; Preece, 2000). This term invokes the social nature of traditional communities, while also highlighting that the community is mediated largely through the Internet. This work is squarely focused on places and communities that exist on Reddit, without exploration of traditional communities. In some places, I occasionally refer to a community in an informal way, and I generally imply that a network of people that have formed a community within the subreddit or more generally as some other online community. This is sometimes used aspirationally, with the hope that a subreddit does form a community.

In using this definition, I want to highlight several key points that I feel are important for online communities. First, the focus is on social interaction over a period of time through the Internet. I believe the most critical part of any community are the interactions between people and the shared experiences and history that creates, which can create a sense of belonging and shared identity. Second, the definition lacks the notion of physical boundaries or physical space which is present in some definitions of traditional communities. While some online communities may be connected or bound by physical space (e.g.

r/Seattle), members of online communities connecting across the globe. Lastly, the vagueness in the definition is intended to provide a level of inclusion – to potentially include forms of online community beyond those I study within this work. Online communities formed in massively multiplayer online games may differ in many ways to the reddit communities formed in subreddits. I believe this variation is at least, in part, responsible for the diversity of definitions of traditional communities. For example, cities can look fairly different from rural villages (e.g. Hillery, 1959). The reader may benefit from a brief overview of the contentious effort to define communities before the internet existed.

2.2 DEFINING COMMUNITY

Many definitions of *community* have been offered over the years, but there is no widely agreed upon definition. The lack of agreement on defining attributes of a community have been the topic of a scholarly writings in sociology for decades at least (e.g. D.B. Clark, 1973; Hillery, 1955; Homans, 1951). The usage of the word is further complicated by its use in other academic fields and in everyday language. An overview of the debate may provide some value for contextualizing my definition and this work.

My intention is to provide additional context into the debated attributes of the term, since this work is focused on the creation and evolution of online communities. I intend this to have value in two ways. First, the attributes used across definitions will help understand points of general consensus and disagreement. One possibility is that some differences arise because of different configurations of community which might vary along these attributes. Second, the most common attributes also provide a loose framework which can be used to compare different community configurations.

Hillery (1955) analyzed 94 different definitions of community from existing papers and noted many areas of disagreement, and in some cases contradictory. In total, he identified 16 different elements across the definitions. The largest point of agreement is that a community involved a group of people, with 91 of the 94 definitions noting this (The remaining three were ecologically focused). Social interaction was the next most common element of the definitions, followed by geographic area, and social ties. Of the remaining features, a few stand out including Self-sufficiency, common life, “consciousness of kind” and

2 Background & Related Work

the “possession of common ends, norms, and means”. One of his challenges was how the definitions for rural communities differed from other forms of community.

Later, Hillery (1959) provided a model of community providing five high-level attributes from an analysis of five different villages. His intention was to derive these categories from multiple communities instead of “conforming to the preconceived notions of the investigator.” Four of his proposed categories were drawn from Homan’s work (Homans, 1951) and were: interaction, sentiment, norms/institutions, and activities. His addition for the fifth element was *space*. He also has a number of sub elements, but the most relevant was the inclusion of “continuance” which he defines as the ability to persist in time, such that a person could live a lifetime within that community. The temporal element is interesting because it recognizes that communities may not exist indefinitely.

Other sociology scholars who summarized definitions have also noted common elements of locality, social interaction, and social ties as important elements across definitions. Clark’s categorized definitions into four themes: locality, social activity, structure, and sentiment (D. B. Clark, 1973), but he felt that the two most “fundamental communal elements” were a sense of solidarity and significance. These latter two may also be seen as a form of sentiment. Bender (1978) notes that “the most common sociological definitions used today tend to focus on community as an aggregate of people who share a common interest in a particular locality.” Wellman and Leighton noted 3 common elements in their summary: social ties, locality, and “solidarity sentiments and activities” (Wellman and Leighton, 1979).

Community was also a topic in other fields, like psychology, where they focusing on internal aspects including feelings, attitudes, bonds, and behaviors. Some early scholars noted this collectively through attributes like “sentiment” (e.g. Hillery, 1959; Homans, 1951). One particularly relevant line of research from psychology focused on the feelings of belonging and commitment that members of community felt. This includes the research on “sense of community” (Doolittle and MacDonald, 1978; McMillan and Chavis, 1986). Some of this work was focused on expanding the sentiment aspect of communities. This led to forming additional definitions that focus on the critical elements which contribute to feeling of community, like elements of emotional connections and influence (McMillan and Chavis, 1986).

2 Background & Related Work

With the adoption of computer networking, community was adopted by people to describe their experience interacting with others through computers, perhaps unaware of the long-running debate over the word (Baym, 1998). Howard Rheingold coined the term virtual community which he defined with a focus on social interaction over time: “Cultural aggregations that emerge when enough people bump into each other often enough in cyberspace. A virtual community is a group of people who may or may not meet one another face to face, and who exchange words and ideas through the mediation of computer bulletin boards and networks.” (Rheingold, 1993). The term later evolved into online community, which also continues to be redefined.

Many of the features or elements of online community were shared with traditional communities. These include people, places, interactions, ties, norms, sentiment, and identity. Interaction was common in a number of definitions. Sentiment was often a particular focus of some early work which showed people could develop closeness and attachment (e.g. Baym, 1998). Space sometimes became a background in some definitions (e.g. Preece, 2000; Whittaker, Isaacs, and O’Day, 1997), but others did include some aspect of space (Jones, 1997; Ridings, Gefen, and Arinze, 2002, e.g). Another change for some definitions was a focus on commonalities or purpose, which was especially salient for the many people studying Usenet where newsgroups are topically focused (Whittaker, Isaacs, and O’Day, 1997). Some definitions also included references to time or repeat interactions (e.g. Rheingold, 1993; Ridings, Gefen, and Arinze, 2002; Whittaker, Isaacs, and O’Day, 1997).

To summarize, there are a number of common themes from the many definitions. These include: interaction, space, social ties, and sentiment. There are a few other elements which may be considered important including norms, activities, a commonality, and persistence over time. These attributes may be seen as a framework for understanding different community configurations, including both many forms of traditional and offline communities. Perhaps the most contentious of these elements is space, which is worthy of some additional ink, if only to understand how what aspects of online spaces resemble physical spaces, and why those are important for community.

2.2.1 THE SPACE, PLACE, AND LOCALITY OF COMMUNITY

Many definitions of community include a concept of space, place, or locality. This includes pastoral towns and villages that are often conjured up when many people think of community. The concept has been brought into some online definitions as well (Jones, 1997; Ridings and Gefen, 2004, e.g). Given the apparent differences between online space and geographically space, it may be of benefit to unpack the function of space in a community. Community definitions that include space often focus on an interaction stage, boundaries, and the influence of space on the community's norms and culture.

The shared space of a community can act as a stage where people interact and experience community. In traditional communities, this is the public space including a “town square” or be as large as the town itself. In this capacity, the space is providing an opportunity through proximity which enables the presence and awareness of others (Harrison and Dourish, 1996). Some conceptualizations of community focus on the lived experience within a space (MacIver and Page, 1959), which became particularly important as people began comparing early online communities with traditional ones. Some focused on the richness of information in face-to-face communication, others focused on the feelings experienced which they felt online environments could not compare (Doheny-Farina, 1996; Weinreich, 1997).

Place may be important for generating the opportunity to interact and initially build connections, but relationships built may extend beyond the borders of the place. Research suggests that people have been quite mobile for a long time, interacting and maintaining relationships with people beyond their neighborhood or village (e.g. Sanderson and Polson, 1939; Wellman, 1979). In his study of East York, a suburb of Toronto, Wellman (1979) found many East Yorkers reported having only a small percentage of close ties within their neighborhood. Many residents had a majority of their ties in the greater Toronto area or further away. Those that reported having intimate ties with neighbors tended to report them as being weaker than other intimate ties. The widespread adoption of technologies like the automobile and telephone enabled people to build and maintain social ties well beyond their neighborhood or town much more easily.

2 Background & Related Work

Such space also traditionally provides boundaries. In the traditional, ideological village, these may appear to be well-defined boundaries that help provide an area of inclusion and exclusion. Local governments have power largely within a specific space, and taxes are largely spent on projects within its borders. These boundaries can also define membership in the form of residents and citizens, and can make people feel safe through exclusion. This can also give rise to both in-groups and out-groups, which can result in identity development as well as creating and maintaining power among the in-group. In communities these might be residents, while people who live outside the borders are tourists, guests, or immigrants.

However, boundaries can be malleable and fuzzy. Unclear boundaries have been studied in the context of collaborative groups (C. P. Lee, Dourish, and Mark, 2006) and even online spaces with very permeable boundaries (e.g. Carrasco and Kerne, 2018; Milton et al., 2023). Although boundaries may provide a sense of safety to members, this can occur even in situations where boundaries are unclear and mediated largely through algorithmic recommendations of public content (Milton et al., 2023). Social ties have also found that people within neighborhoods regularly interact and maintain ties with people beyond these borders, which raise the question about the necessity of boundaries for all communities (Wellman, 1979).

Research from Waterville, New York in the early 20th century also illustrated how fuzzy these boundaries can be (Sanderson and Polson, 1939). The village itself had defined borders, and exists on the border of two towns. Because of neighboring settlements, the aggregate business areas shifted with time and competition from nearby villages and did not align with official boundaries. Many farmers and people from smaller nearby townships lived outside the village borders were considered in surveys and made use of some of its services, but were often considered an out-group for residents. For example, school districts and the local library often had officially defined service areas which shifted over time with socio-economic conditions which led to closures and consolidations. Further, the borders imposed power imbalances between the residents within the village's official borders and those living outside the borders. In one example, he describes the local library which serviced those farmers, was forced to expand its borders to be in line with the school district after pressure from the State Department of Education which provided

2 Background & Related Work

funding to include the larger area (Sanderson and Polson, 1939, pg. 39-40). This suggests that hard borders may be less important for a community. The borders and community itself may shift over time, and what may matter more are specific places which enable the social interaction.

Lastly, some scholars also consider the ability to influence or shape behavior and culture. To some extent, many aspects of behavior can be influenced by a geographic area. Many geographic characteristics like climate affect what food might be cultivated and how people dress. Coastal and river communities may include ship travel and fishing more than communities further from large bodies of water. However, these influences may also change over time as a community grows, and its relationship to the land itself changes. The growth of a city typically comes with changes in lifestyle which enable citizens to rely less directly on the local resources (eg. farming or fishing for your own food).

There are many instances of culture being influenced by the local geography, but our relationship to this is both social and psychological. We develop meanings, feelings, and even identities around these, and those can also change with time as a community evolves. Research exploring perceptions of beach goers in the years after desegregation, found that the white visitors who had visited before and after had felt they had lost their connection to space (Dixon and Durrheim, 2004). This suggests that even if geographic locations remain largely unchanged, our attachment to them can change based on psychological and social perceptions of a place.

Some scholarship has begun to look beyond the typical borders of a locale. The rise in urbanization and the adoption of technologies like the telephone and automobile have certainly helped diminish the focus of locality. A number of scholars instead began focusing on other types on more psychological and social elements present within communities. Scholars have focused on other forms of community including psychological aspects (e.g. D. B. Clark, 1973; McMillan and Chavis, 1986) or to see communities as a network of social ties (e.g. Bender, 1978; Wellman, 1979).

One particular line of research focuses on our relationships to space. Drawing on several threads of research, Harrison and Dourish (1996) propose a distinction between space and place such that place is the social context in which we act, while space is the physical area. The social context is our social

2 Background & Related Work

understanding of a place. This includes the norms within or around a space that help us understand what behaviors or participation is expected or allowed.

While space and place are social constructs (Dourish, 2006), our intentional use of space over time and our understanding of this creates a place from it (Harrison and Dourish, 1996). The distinction here captures both psychological and social elements created as part of the social practices over time, similar to the shift in focus to these elements of communities, wherein communities are places. The institutions, cultures, and norms that arise from practice contribute to our understanding of community, and are the understood reality of what we think of as communities. Jones (1997) used a similar line of thought to argue for distinguishing between the separation of virtual communities and their *virtual settlements* – e.g. the place of these virtual communities.

The usage of community might conjure up images of quaint villages with well-defined borders in which people live their entire lives, but for many the lived experience may be very different. What many think of is the experiential aspect of community – our understood reality of their practices over time, which have given it its character. Their social ties may not be bound solely with just other inhabitants, but those living beyond its borders may also contribute. While these may be influenced to some extent by the geographic boundaries, this may not always be the case.

2.2.2 CRITICISM OF COMMUNITY AND ONLINE COMMUNITY

As one might imagine, the lack of agreement over the definition of community has been one source of criticism about the concept and its usage. There are many criticisms, but a brief overview of some of these may benefit the reader by adding more context into areas of interest on the topic. The first two concern the general definition of the word, and the latter focuses specifically on the validity of online communities as true communities.

2.2.2.1 THE VAGUENESS OF DEFINITION AND CONCEPT

A number of scholars have been critical of the vagueness of the word community in its usage. The vagueness includes defining elements of a community, and also the concept the word is intended to

2 Background & Related Work

describe (Stacey, 1969). Hillery's work often approached this from a rural communities, while a number of sociologists focused on large cities which can manifest very differently. Some definitions allow community to be as large as a nation (MacIver and Page, 1959), which others feel are implausible (Sorokin, 1969).

As Butterworth and Weir note, "community tends to be a God word. In many circumstances, when it is mentioned, we are expected to abase ourselves before it rather than attempt to define it." (Butterworth and Weir, 1970, cited in Bell and Newby, 1975). Scholars have made calls to stop using it entirely (e.g Sorokin, 1969). Poplin (1972) felt that using the word community implied a lack of "conceptual rigor" because of its vagueness. Some have suggested people use other words instead (Hillery, 1959; Stacey, 1969). Despite the frustration and calls for change, the word continues to be used and redefined (D. B. Clark, 1973).

Bruckman (2006) argued that the question of whether a unit was a community or not was counterproductive, especially given how value-laden the term had become. For her, these did little to advance our understanding of communities. Borrowing from cognitive science work on categories, she argues it may be more valuable to examine the similarity of any such community with a model or prototypical member. This can also encourage us to look at a number of different forms of community, and compare members between these different classes.

2.2.2.2 DECLINATION AND THE USAGE OF COMMUNITY

Another point of contention is that many uses of the word community are often done using a declinist perspective. As Bender (1978) notes, there's often a negative undercurrent in these uses implying that community is declining in some way and we're in danger of losing it. The declinism here suggests that community was better in some unknown past. Layman definitions and a line of scholarly work often views community with a positive spin, such that it's an ideal situation. Some of this perspective may be the result of a number of known biases which affect comparisons between situations in the past and in the present causing them to view history more positively and creating an illusion of decline (e.g. Mastroianni and D. T. Gilbert, 2023). A number of early sociological works on community seek to explain the effects of industrialization or modernization on popular issues like crime or neighborhoods.

This has even been the topic of popular works (e.g. Putnam, 2001), even when thoroughly researched. Works from this perspective may have many valid and important points, but these often problematize the process of change itself. We might be better served to ask how much of these concerns are the result of a flawed definition – especially those written around a static definition that allows for no change. As Bender notes, “if community is defined as a colonial New England town– and it is frequently defined that way– then the prospect of community today is indeed dim.”

2.2.2.3 ONLINE COMMUNITIES ARE NOT COMMUNITIES

Not all scholars agree with the usage of community in describing online communities. The criticisms roughly fall into three areas. These are the quality of social relationships the Internet can afford, the lived experience of a traditional community, and the lack of a geographic or physical space ¹.

Perhaps the most important criticism is about the ability to build relationships of a quality comparable to traditional off-line communities. Deconstructing the criticism, there are two important aspects. There's the overall question of whether a purely online-only relationship can be as meaningful as those built off-line. The second is the weaker version, wherein people may leverage the internet for day-to-day communication, but at some point this is augmented with an in-person relationship.

Baym (1998) proposed reframing the debate to explore whether online communities could substitute any part of a traditional community and focusing on what occurred that made people describe their experience as a community?

A separate but related concern raised by some scholars is community needing to be a lived experience. For some, there's an intangible quality that comes with the lived experience of community (Weinreich, 1997, Doheny-Farina, 1996). Since online communities are not physically experienced in the same way, and

¹For a more detailed comparison of the uses of virtual community, including criticisms see (Jones, 1997).

people can join and leave much more easily, it couldn't capture the full experience of a more traditional off-line community.²

Lastly, the issue of physically bounded space. Much of this criticism appears to come largely from descriptive definitions of online community. When scholars studied community in rural areas and remote villages, the jurisdictional borders likely seemed like clear boundaries. However, even in the real world these boundaries can be fuzzy for delineating membership (Sanderson and Polson, 1939; Wellman, 1979). Stone (1991) argued that even as far back as the 1600s, scholars had developed a scientific community while not living in the same physical space.

2.2.3 SUMMARY

In summary, there is no broadly agreed upon definition of *community*, and the debate over the essential elements of a community will likely continue. There are however a few areas of overlap worth reiterating. First, there is some amount of social interaction. Through continued social interaction, social ties or relationships may form. When a tie is strong enough, people may feel more connected to other people. Traditional definitions also highlight *area* as an important feature of a community, but there is a debate over a realistic scale of this area and what affects it may have. This area may also be approximated, at least to some extent, by some online spaces.

2.3 ONLINE COMMUNITIES

In this section, I focus on research specific to how and and people participate in online communities. This includes the motivations for why people seek out and participate, how people participate in these spaces, and why some participate more or less than others. I then briefly summarize some work on governance in online communities, as this has become an incredibly important topic for every community and especially

²In discussing the issue, Baym (1998) notes that she and her participants used the word community because it's how they experienced it. She believes the more important question is around what occurs within these spaces makes people feel that community.

as it scales. Lastly, I will briefly cover some limited research on how communities evolve over time looking including how the communities are formed and how they change in a few dimensions.

2.3.1 WHY DO PEOPLE PARTICIPATE IN ONLINE COMMUNITIES?

Previous research has found that people come to online communities for many reasons including seeking information, entertainment, support, socializing/relationships, and networking (e.g. Diakopoulos and Naaman, 2011). A number of factors can influence these motivations including the type of online community, their needs, and how they want to participate (C. Lampe, Wash, et al., 2010; Moore and Serva, 2007; Ridings and Gefen, 2004). Someone looking for support or information in dealing with breast cancer may find a discussion forum aligned more to their needs than a wiki.

Two common reasons for participating in an online community to seek and share information. Many online communities on reddit, Usenet, and Facebook are focused around specific topics, identities, or experiences making them a potential resource for people looking for information. People with questions may be motivated to create accounts to ask the community (C. Lampe, Wash, et al., 2010; Ridings and Gefen, 2004). If they already have accounts and their questions are sensitive, they may also create new accounts to ask questions to protect their privacy (Leavitt, 2015; Wohn, Yuan, and Siri Jr, 2021). For public or semi-public spaces, having information readily available may still be a reason to read and not actively participate (Nonnecke and Preece, 2001).

People also commonly seek out online communities to give or receive support, such as emotional support. This can be common for people experiencing identity challenges, medical illnesses, and general hardships. For example, LGBTQ+ identifying people and their relatives may face challenges in coming out, finding resources, and generally receiving support for their situations. This can be especially important in situations where people feel they do not have real world support. Turning to online spaces, people can share their stories, get advice, and provide emotional support (Blackwell et al., 2016; Carrasco and Kerne, 2018). People facing health difficulties or illnesses may also be motivated to seek information or support from online groups (Kingod et al., 2017; Massimi et al., 2014; Y.-C. Wang, Kraut, and Levine, 2012; Welbourne, A. L. Blanchard, and Wadsworth, 2013). Finding online communities for support or building relationships

2 Background & Related Work

can also be important for people when they feel their real-world connections may be strained (Kingod et al., 2017).

Like traditional groups and communities, some people may be motivated to participate to have social interaction and develop relationships, like meeting new people (B. Butler et al., 2002). Some people seek out specific communities with an interest in just communicating with people, especially if they have friends or family in those communities (Backstrom, Huttenlocher, et al., 2006). However, some people may be more motivated to just connect with a group-at-large rather than individual people (Hwang and Foote, 2021). Relationship development can also include professional or job-related types of networking (B. Butler et al., 2002; Shrestha, Barik, and Parnin, 2021). Some evidence suggests some people may even do this in communities predominately focused around other goals, like question-answer spaces (Tausczik and Pennebaker, 2012).

Some people may be motivated to participate by their desire to give or give back. This may include active ways like answering questions (B. Butler et al., 2002; Nov, 2007). Though, for some this answering questions this may not be a motivation for joining, but instead for continued participation after they have received information from others. However, Kollock (2001) argues that some may contribute with a future expectation of receiving support, noting that true altruism is rare.

People can also have different reasons for creating communities. Some people may create communities with a stated goal of being altruistic B. Butler et al., 2002. In some cases, people may even create smaller more niche communities that are off-shoots of larger communities with a goal of filtering some of the information and act more like a safe space (Hwang and Foote, 2021). However, others may never have an intent to create a community, but are instead motivated to create a space. For example, some owners of Fandom wikis primarily wanted to create an information resource (Foote, Gergle, and Shaw, 2017). This has potentially important implications for how we think about success of some online communities, given that the creators may not approach with a goal of creating a large or even thriving communities (Foote, 2019; Foote, Gergle, and Shaw, 2017).

Some research has also found that a person's initial motivation for participating may change over time. Someone may approach an online community first with an intent on getting information, but may become more motivated by answering questions or just being entertained (C. Lampe, Wash, et al., 2010). In some cases, they may even become develop a sense of attachment or belonging to the group, which also affects their motivation.

2.3.2 PARTICIPATING IN ONLINE COMMUNITIES

Many online communities enable multiple methods of participation. People can generally produce content in some way, such as creating a post or sharing an image. They can also passively consume content such as reading posts and comments or looking at images and watching video. Many early forms of online community such as IRC and Usenet largely stuck to these forms of participation, but over time many platforms have developed forms of interaction that take less effort and time, such as Facebook Likes/Reactions, favorites, upvoting/downvoting, and sharing. These more lightweight forms of interaction can lower the barrier to people engaging.

A common theme for many online communities is *participation inequality* where the amount of participation is heavily skewed. A small handful of people may produce the majority of content in a community, while the majority may only consume or produce very rarely. One Usenet study found 27% of people contributed only a single message (Whittaker, Terveen, et al., 1998). On the WELL, an early internet discussion forum, 50% of the content was produced by about 1% of the users (M.A. Smith, 1992), and another Usenet study found 14% of people created 75% of the content (Golder and Donath, 2004). However, many people do not create content at all and simply consume in these communities and these are generally referred to as *lurkers*.

2.3.2.1 LURKING

The number of lurkers can be difficult to estimate, especially in communities where the information is publicly available. Some may passively engage only once, while others may passively engage many times. Some unsourced early work places the number of lurkers at 90% or above (Katz, 1998; Mason, 1999). One

analysis quantifying lurkers found a wide variation in the percentage of lurkers across different email distribution, however this differed by the type of online community, with software support communities having higher variance (Nonnecke and Preece, 2000).

Lurking has been viewed as problematic. For some scholars, online communities and the knowledge they share were seen as a public resource and thus the resource must be sustained through continued discussion (Kollock and M. A. Smith, 1996, eg.). Lurkers are often the largest pool of would-be contributors. Finding ways to motivate lurkers into “good citizens” that contribute to the common good was seen as important by some by some people (Beenen et al., 2004; Rafaeli, Ravid, and Soroka, 2004).

Lurkers have a variety of reasons for lurking and many are not free-riders. Many lurkers are satisfied simply by browsing or consuming content from the community, which can happen when people come to online communities primarily to satisfy information needs (Nonnecke and Preece, 2001). Lurking is also akin to overhearing conversations, which can help people learn about topics and the environment (Kraut, Fish, et al., 1990), and this can be an important part of learning the language and norms within a community (Nonnecke and Preece, 1999). Crawford reframed lurking to be a form of *listening* to shift discussions of lurkers toward a more active and reciprocal form of engaging (Crawford, 2009). Ellison et al. (2020) also found that people who do not actively interact were not necessarily less engaged or spend less time. Some lurkers may eventually ask questions, which could stimulate discussion and ultimately be valuable to the group (Tedjamulia et al., 2005).

2.3.3 ENCOURAGING CONTRIBUTIONS AND COMMITMENT

After people are motivated to post, the challenge becomes encouraging people to continue posting. Without continued contributions and interactions, communities may die as is the fate of many studied online communities (e.g. Kraut and Fiore, 2014). One line of research has looked into factors that affect commitment as well as design decisions which may contribute to it. The longer term commitment may be important for continuity and enforcement of norms (Backstrom, R. Kumar, et al., 2008).

2 Background & Related Work

After people make their first contribution, many never make a second one (Arguello et al., 2006; C. Lampe and E. Johnston, 2005). In some cases, the person may have had a single information need and one reply was enough to satisfy that need. Another possibility is that the person's post was part of a trial period to evaluate the community (Freeman and Wohn, 2019) before deciding to look elsewhere.

Receiving a reply to a first post may be an important first step in encouraging people to continue posting. One line of research has found that new posts are less likely to receive responses than posts made by more established users (Arguello et al., 2006; Joyce and Kraut, 2006). Receiving replies may make people feel welcome (Shrestha, Barik, and Parnin, 2021), especially when that response is positive (Wohn and C. Lampe, 2018). This can also increase the commitment to the community (Yang, Kraut, and Levine, 2017).

A number of other factors relating to the post itself may affect whether a post receives a reply poster familiarity to the group, including message length, personal introductions, and containing a clear question (Arguello et al., 2006; Burke, Kraut, and Joyce, 2010). Backstrom, R. Kumar, et al. (2008) found that people who eventually became core contributors received more responses than people who would become less engaged, suggesting that other core users may recognize specific elements of the post as being more valuable.

On reddit, a significant amount of content may receive little or no interaction from other users (E. Gilbert, 2013). This may be due partially to the design of the site, which attempts to rank content based on previous interactions, which may render some content less visible. This design can create a small feedback loop for some content that receives upvotes quickly, resulting in a Matthew effect.

2.3.4 MEMBERSHIP

Membership is a common concept of many definitions of communities. Boundaries helps demarcate the who is a member. Some early scholarly on communities focused on geographical boundaries like town or city borders to help define who is a member (Hillery, 1955). However, a separate line of research started to note the fuzziness of membership in general, from rural communities (Sanderson and Polson, 1939), through collaborative groups (C. P. Lee, Dourish, and Mark, 2006), and through online communities (Chancellor, T. Mitra, and De Choudhury, 2016; Rotman and Preece, 2010).

2 Background & Related Work

However, communication changes have helped shift the way many organizations work together, enabling forms of remote work, and groups where membership is less clearly defined (Chancellor, T. Mitra, and De Choudhury, 2016; Harris et al., 2019; C. P. Lee, Dourish, and Mark, 2006). While participation is important for the social interaction that's central to groups, organizations, and communities, not everyone who participates will see themselves as a member potentially because there are no clear boundaries for membership (Baym, 2009; C. P. Lee, Dourish, and Mark, 2006). Even when asking people what defines membership, the answers are not always clear. For example, in many physical communities, people have differing perspectives on what makes someone truly a member. Members with more power, such as those who own housing, may set high-level boundaries that reflect their status, which can be reproduced over time, to separate themselves from other members and effectively cast them as an out-group (Toolis and Hammack, 2015). This can have severe effects on the out-group, such as limiting access to many resources which can require membership.

In online groups, this can be especially important as many platforms have made it easy to “join” groups and communities, but even passive engagement after joining is not guaranteed. Joining simply enables the ability to quickly return and view content from that community, but often times, as is the case with reddit, this content is selected algorithmically and can be biased towards recent content. So less active communities may require users to be actively checking in order to see content. This affects the algorithmic visibility of the information.

On reddit, we can see membership as participation or as ‘subscribing/joining’ a subreddit, but joining does not guarantee any level of participation. Though people may join or subscribe, they may never contribute or even passively read content from the subreddit. We might also see throwaway accounts (Leavitt, 2015) as a special form of membership, wherein a single person has multiple accounts, but uses the throwaway accounts to maintain distinct identities or to grant them some level of privacy separate of their main identity.

Computer recommendation algorithms have also been used to suggest groups and even “create” them, raising more questions about the meaning of membership. Social media platforms like Instagram³ and TikTok forgoes traditional groups like subreddits, but instead enables them to form through engagement and participation. These groups are less structured and more permeable, which depending on the platform may affect other people’s ability to discover the community (Hillman, Procyk, and Neustaedter, 2014; Milton et al., 2023) depending on the specific affordances. This can enable a “bubble” to form which can help people feel safer (Carrasco and Kerne, 2018). Despite the lack of clear boundaries, people still feel that these are communities (Milton et al., 2023).

2.3.5 GOVERNANCE

As a community grows, members will likely develop norms and rules around behavior that is or is not acceptable. Many online spaces are subject to unwanted behaviors such as spam, off-topic content, harassment, and bullying. The rules and policies formed by the community are important for maintaining the focus of a community (Preece, Nonnecke, and D. Andrews, 2004) and limiting anti-social behavior (Kraut and P. Resnick, 2012). These can help people feel more welcome, help create safe spaces where people feel inclined to engage and share (Carrasco and Kerne, 2018), and help develop trust which is important for some communities such as those facilitating trade and commerce (TeBlunthuis, Kiene, et al., 2022).

As Reddit has grown through the years, a number of problems have developed which they attempted to address through cooperation with the larger community of users. The co-evolution of the platform is a particularly interesting line of research – especially in how the often unpaid labor of moderation has evolved with the platform (Chandrasekharan, Pavalanathan, et al., 2017; E. Gilbert, 2013; Matias, 2016; Matias, 2019). Some platforms, like reddit, continue to rely on the labor of its unpaid subreddit moderators despite calls for better tooling to manage unwanted behaviors and prevent harassment. The lack of support can leave moderators to feel unsupported (Matias, 2016; Wohn, Jough, et al., 2019)

³Instagram has since implemented groups more formally, though the cited works here were before this feature was introduced (*New Sharing Features on Instagram: Notes, Group Profiles and More 2022*).

2 Background & Related Work

In reddit's early years, the site relied on the community and their voting to moderate content, under the belief that unwanted content would be downvoted. This eventually led to hiding of content below certain thresholds, which was likely influenced from Slashdot and their usage of distributed moderation (C. A. Lampe, E. Johnston, and P. Resnick, 2007), a concept that has been shown to work reasonably well in some cases (C. Lampe, Zube, et al., 2014), but has not worked as well on Reddit (E. Gilbert, 2013; Matias, 2016).

Reddit allows the creator of each subreddit to be a moderator and set their own rules, in addition to the site-wide rules (Chandrasekharan, Samory, et al., 2018; Fiesler, McCann, et al., 2018). However, as platforms and communities grow, the amount of moderation work needed also scales (B. Butler et al., 2002). The work that moderators do extends beyond simply removing negative content, but includes communicating with users, other moderators, and configuring chatbots. They may recruit additional moderators to help with the maintenance of each subreddit, which includes enforcing some of these rules. However, many moderators may be initially untrained and have to spend effort learning the skills and potentially technology in order to do their job (Noveck et al., 2021).

To help overcome the shortage of volunteer moderators, reddit has provided support for configurable software that functions as a bot named AutoModerator, which can be difficult to program (Jhaver, Birman, et al., 2019). Software solution, sometimes known as bots, are common for many forms of online communities and spaces. These have been employed successfully in many online spaces including Wikipedia, and are an important element of the larger ecosystem, including work that can be less visible (Geiger, 2018; Geiger and Ribes, 2010), and reduce transparency of moderation work.

The lack of transparency within moderation is an ongoing tension between moderators and users on many platforms. The moderation process on reddit has also been a black box to many users, resulting in confusion about why content was removed. (Jhaver, Appling, et al., 2019; Jhaver, Bruckman, and E. Gilbert, 2019), and suggesting that transparency and governance may be important dimensions.

Some platforms do employ some people to help enforce some of their content policies (S. T. Roberts, 2016, e.g), but the work can be mentally challenging including looking at many forms of harmful pictures. There's also a challenge of adapting policies to align with many different cultural norms, rules, and

regulations across many different countries the platform may have a presence. Additionally, the policies set by these companies may align more with their corporate values rather than helping their users (Gillespie, 2010).

2.4 PEOPLE, BOTS, ALTS, AND ACCOUNTS

When discussing behavior online, it is important that we acknowledge that every user account on reddit may not map to a singular, unique person. People may operate multiple accounts, multiple people may operate a single account, and accounts may be automated by script or other forms of code (aka bot) On reddit, norms and necessities have created an ecosystem of many different kinds of accounts which may be operated by a person or some type of script.

2.4.1 MULTIPLE ACCOUNTS

There is a norm around people operating multiple accounts on reddit, enabled through the platforms low barrier for creating new accounts. On reddit, people commonly refer to the a primary account as a *main* and the other accounts as *alt* accounts (alts). People have a variety of reasons for using multiple accounts including privacy, curation, identity separation, engaging in deviant behaviors, avoiding harassment, and manipulating content (Wohn, Yuan, and Siri Jr, 2021). One specific type of alt is a *throwaway* account, which is typically a short-lived alternate account. These provide a level of anonymity to users such that some content is not associated with another account (Leavitt, 2015). These allow people to ask sensitive questions or participate in conversations that they don't want related to their everyday identity.

2.4.2 BOTS

The platform also provides a programmable interface (API) which has enabled people to create automated accounts (bots) which automate a number of tasks. The most popular such account is AutoModerator a configurable tool that helps users moderate their subreddits. It was created by a reddit user, and later reddit employee, for their own usage (u/Deimorz, 2012a; u/Deimorz, 2012b; u/Deimorz, 2015). However, it is one of a much larger ecosystem of bots which accomplish many different tasks including posting content

from external sources, linking external content mentioned in comments (e.g. YouTube and Wikipedia links), converting units of measurement, and making jokes.

The distinction between accounts, bots, and people is important especially when studying limited trace data where we cannot reliably identify who or what is on the other side of those traces. This can complicate a number of potential claims I might make, especially given the ease with which the text generation can be automated. Even as early as 2012, bots were able to copy text from other sources and later generate natural language summaries. This has enabled bots to create human-looking text, but not actually be human. I have attempted throughout this work to remain cognizant these facts and to be careful of how I word claims to reflect that some of observed behaviors may not be human. As of 2026, bots that can generate legible human text are everywhere through the magic of large language models and a new round of “AI”.

2.5 EVOLUTION OF ONLINE COMMUNITIES

Studies of the evolution of online communities are not common. Many studies use datasets collected over time, but often these are used for descriptive or explanatory models that do not explicitly consider time (e.g. Whittaker, Terveen, et al., [1998](#)). Others look at across multiple communities, but take an individual or platform approach to explaining behaviors (Singer et al., [2014](#), e.g). Some of these have been mentioned in the previous sections.

From a theoretical perspective, some scholars have proposed models or frameworks that attempt to describe different phases of evolution in the life-cycle of an online community. D.C. Andrews ([2002](#)) proposed a 3 stage framework for design consisting of creation, encouraging early commitment, and moving towards a self-sustaining community. Iriberri and Leroy, [2009](#) synthesized a body of work on online communities and proposed a life-cycle model of online groups and communities: inception, creation, growth, maturity, and death.

A few studies have used longitudinal approach to understanding the growth, survival, and evolution of online communities. For example, Backstrom, Huttenlocher, et al. ([2006](#)) found that having friends in a

2 Background & Related Work

group and the group size were useful in predicting the amount of growth a group would experience. Other works have looked at social networking structure (e.g. Backstrom, Kleinberg, et al., [2013](#); Tsugawa and Niida, [2019](#)).

Structural evolution of online spaces has also been a productive area of research, wherein scholars model network graphs over time. In the social sphere, these model connections or interactions between entities (e.g. people). A number of large, real-world networks evolve into a scale-free networks where the number of connections of each node in a graph follows a power-law distribution (Barabási and Albert, [1999](#)). Scale-free networks occur when a network is growing continuously and new edges follow some form of preferential attachment – the idea that new connections are more likely to attach to existing well-connected nodes, with Barabasi and Albert’s originally focused on older nodes being preferred over newer. More recent studies of scale-free networks have found that the process of preferential attachment may work differently for some real-world networks like the world-wide web (e.g. Adamic and Huberman, [2000](#)) and especially some social platforms (Leskovec et al., [2008](#)) with attachment occurring through other processes like diffusion. Using simulation, Johnson, Faraj, and Kudaravalli ([2014](#)) found that no single process would result in power law distributions commonly found in online communities.

Other works have attempted to use structural characteristics from networks to predict either growth and survival of communities. Predicting growth has had some mixed results with some research finding graph characteristics may vary by age and size. Kairam, D. J. Wang, and Leskovec ([2012](#)) found that structural features were more accurate at predicting growth for smaller groups, but not groups larger. Conversely, Tsugawa and Niida ([2019](#)) found that network characteristics did not have a large effect on growth. Some structural characteristics like having a high number of members involved in the largest clique could increase survival times while high levels of density may decrease survival time (Kairam, D. J. Wang, and Leskovec, [2012](#); Tsugawa and Niida, [2019](#)).

In multi-community spaces like Usenet and reddit, some users tend to explore communities over time (Tan and L. Lee, [2015](#)). While some of this exploration may be about changing interests, people often explore similar topics – an insight which has enabled recommendation algorithms for finding similar

2 Background & Related Work

communities (e.g. Martin, 2017; McGlohon and Hurst, 2009). The overlap of users between communities can be used to understand some aspects of evolution of those communities and the platform as a whole. Multi-community platforms like reddit can often have many related communities, which can share users leading to competition and mutualism. Some previous studies have suggested that high amounts of membership overlap may be an indication of competition and limit growth or survival (e.g. Hessel, Tan, and L. Lee, 2016; X. Wang, B. S. Butler, and Ren, 2013; Zhu, J. Chen, et al., 2014), but mutualism between different communities is also prevalent on reddit where related communities benefit from new niches (TeBlunthuis and Hill, 2022). New communities can find a related but topical niche or specific purpose not covered by an existing community (Hessel, Tan, and L. Lee, 2016; TeBlunthuis, Kiene, et al., 2022). Overlap of early members of a community may also positively affect growth, if the overlap to a parent is particularly strong (Tan, 2018).

The work done early in the life of a community may be important for its survival. Having early activity has previously been found to predict the lifespan of communities (Cunha et al., 2019; Raban, Moldovan, and Jones, 2010; Solomon and Wash, 2014). Early content may provide not only informational value to visitors, but may help others in understanding what to contribute. Kraut and Fiore (2014) found that the early activity of founders could influence the growth of a community, though not if they attempt to take on too much activity. Cunha et al. (2019) found that work concentrated to a small number of people initially could improve the odds of survival.

Lastly, one additional area of study has been on the linguistic evolution of communities. The language a community uses may change over time, with either shifts in topics or terminology changes. Users participating in a community may also adopt the language and norms of collective community over time (e.g. Nguyen and Rosé, 2011). Users may also switch their pronoun usage over time from singular first person, e.g. 'I', to collective pronouns like 'we' (Cassell and Tversky, 2005). Danescu-Niculescu-Mizil et al. (2013) found patterns of language adaptation in the earliest participation of new members to more closely align with the community up until a point where they finally stop responding to community language changes.

2.6 EXISTING FRAMEWORKS (AND MODELS AND CHARACTERIZATIONS)

One contribution of this work is an analytical framework that can be used to characterize, analyze, and compare communities. To better understand the contribution of my framework, I will first review a few frameworks and works include characterizations. I consider a framework to be a form of organizing or structuring which may aid in the planning, analysis, or understanding of research.

I found several works which either explicitly propose frameworks or include methods of characterizing of online communities (e.g. De Souza and Preece, 2004; Lazar and Preece, 1998; Prinster et al., 2024; J. Zhang et al., 2017), but include one generalized framework which is similar to the one I propose (C. P. Lee and Paine, 2015). I also review some research that contains dimensions or structures which could be used to compare or analyze online communities. These typically use quantitative metrics to characterizing the communities they studied (e.g. Adamic, J. Zhang, et al., 2008; Danescu-Niculescu-Mizil et al., 2013; Tsugawa and Niida, 2019; Whittaker, Terveen, et al., 1998).

These include a range of frameworks around many aspects of online communities and groups. Some provide conceptual frameworks for thinking about the ecosystem of an online community, including the design and technological aspects of the platform and the social elements of the community itself. De Souza and Preece (2004) focuses on the social aspects (people, purpose, and policies) and the usability of a community. Kraut and P. Resnick (2012) also proposes many design claims based on social research around networks and influence, among many other topics. These frameworks may be useful for platform and community designers, who need to think about how the platform or software are designed, and how this can shape the social aspects of its members.

Some works focus on psychological aspects of community. Bishop (2007) proposed a conceptual framework focused on individual participation within a community which included motivations, decisions, and methods. Kraut and P. Resnick (2012) used thematic categories to organize psychological and behavioral research and their design claims. A line of research was derived from McMillan and Chavis (1986)'s model of the sense of community, which described the elements of the psychological feeling of a community. Their model was focused around traditional communities and included elements like

membership and shared emotional connection. Their framework was adapted by multiple scholars to apply to online communities (e.g. A. Blanchard and M. Markus, 2002; Koh, Kim, and Kim, 2003) and that typically include multiple dimensions similar to McMillan and Chavis (1986). Blanchard's line of research included a measure (A. L. Blanchard, 2007) and a tested model (A. L. Blanchard, 2008) to identify antecedents of feeling a sense of community.

2.6.1 MULTI-DIMENSIONAL FRAMEWORKS & MODELS

In this section, I focus on research that *explicitly* proposed frameworks, which are similar to the framework I propose in this work. These include frameworks for characterizing and comparing communities, and also to analyze the evolution of communities. These each propose multiple dimensions or methods for characterizing the communities.

Lazar and Preece (1998) proposed a set of attributes and dimensions with the intent of characterizing and comparing online communities which included community attributes, supporting software, and the community's relationship to off-line communities, and the social structure of the community. Their attribute list is not exhaustive and includes examples like shared resources, shared activities, and "intense interactions and strong emotional ties", which may be difficult to operationalize on trace data alone. Prinster et al. (2024) makes two contributions that could be considered frameworks. The first is a collection of four constructs from a thematic analysis with researchers including interactivity, boundaries, homogeneity, and norm enforcement. They also proposed a framework based on 5 community archetypes which describe different behaviors and norms within those communities. One novelty of their framework is that it build on the idea that many communities can serve multiple purposes and that the content and roles may vary. The five archetypes they identified were: topical Q&A, learning & broadening perspective, social support, content generation, and affiliation with an entity. A community can support multiple of these archetypes or it could focus on a single one. A community like r/AskScience may be be most like a prototypical Q&A subreddit, but may still exhibit traits from some of the other archetypes.

C. P. Lee and Paine (2015) proposed the Model of Coordinated Action which was not specifically focused on online communities, but instead designed to compare different forms of coordination. This includes

groups to large-scale crowds, which could include communities. Their framework has 7 dimensions which includes synchronicity, scale, nascence, and planned permanence which could be useful in characterizing communities across many different platforms and stages of development.

I drew inspiration from both Lazar and Preece (1998) and C. P. Lee and Paine (2015), but my framework was developed on community evolution from the point of creation. Thus my framework focuses on different characterizations like interaction and promotion, which are important for online communities.

I developed my framework such that it each perspective can be operationalized with trace data alone (though does not exclude other data sources like surveys) and can be quantified along a dimension. Lazar and Preece (1998) on the other hand has some elements that could be dimensionalized, but other attributes like 'intense interactions and strong emotional ties' may need to be measured by instruments like surveys. Further, like C. P. Lee and Paine (2015), these appear to have been inferred from a literature review.

2.6.2 CHARACTERIZING ATTRIBUTES

Each of the works I analyzed used multiple attributes or dimensions to characterize communities. Some attributes like size appear in multiple works, though many works have unique attributes. Rather than list each attribute or dimension, I grouped some of the attributes into thematic categories and list the more common themes in table 2.1 and described in the sections below. To contextualize my own framework, I've aligned some of the themes with the names I used for my framework's perspective names.

2.6.2.1 COMMUNITY TYPE, TOPIC, & CONTENT

Online communities can have varying purposes and can be focused around different forms of media. A group offering technical support will differ from those focused around producing memes, and the behavioral processes that create the content may differ. Thus the type of community or the types of content the community produces have been one way scholars have characterized communities. This was the basis for Prinster et al. (2024)'s five community archetypes which included Q&A, content generation, and support communities.

2 Background & Related Work

Theme	Subtheme	Example Papers
Type / Topic / Content		Horne, Adali, and Sikdar (2017), Prinster et al. (2024), X. Wang, B. S. Butler, and Ren (2013), and Y. Yu, Jiang, and Dhillon (2024)
	Language	Soliman, Hafer, and Lemmerich (2019) and J. Zhang et al. (2017)
Interaction		Kang and Shin (2006), McGlohon and Hurst (2009), Soliman, Hafer, and Lemmerich (2019), and Whittaker, Terveen, et al. (1998)
Participation / Membership		Kang and Shin (2006), C. P. Lee and Paine (2015), Panek et al. (2018), Schoberth, Preece, and Heinzl (2003), and J. Zhang et al. (2017)
Size, Volume, & Growth		B. S. Butler (2001), Kairam, D. J. Wang, and Leskovec (2012), Kang and Shin (2006), C. P. Lee and Paine (2015), Mensah, Xiao, and Soundarajan (2020), Panek et al. (2018), Rothaermel and Sugiyama (2001), Schoberth, Preece, and Heinzl (2003), and Whittaker, Terveen, et al. (1998)

Table 2.1: Some of the larger themes for attributes and dimensions used by other works to characterize online communities.

X. Wang, B. S. Butler, and Ren (2013) analyzed a number of communities and selected them across four broad topical areas: hobbies, health, technology, and discussion. Horne, Adali, and Sikdar (2017) selected communities based on a four subreddit dimensions including the topic's specificity, the niche-ness of the audience (targeted at a specific audience or general), and the style of the content (e.g. asking questions). These capture several aspects of a topic and the content type.

A related line of research also focused on the language used by the community in the content they generate. This has been used to characterize aspects of language including topical diversity, language change, and political bias. J. Zhang et al. (2017) focused on the language dynamicity and distinctiveness which they use as a framework to explore relationships between content and user behaviors.

Given some of the differences identified for community types, I incorporated this into my selection criteria for the subreddits I selected in the first study. Based on the differences I observed when analyzing

the communities, I also included content type as one of the five perspectives in the framework. I made my *Content Type* perspective broad enough that characterizations used in these works could also be used.

2.6.2.2 INTERACTION

Several scholars have characterized communities by the social interactions found within them. These are focused on the person-to-person activity including replies, likes, and votes. Social interaction can be the basis for developing friendships and getting social support which are two motivations for participating in online communities (Kraut and P. Resnick, 2012; Ridings and Gefen, 2004).

Conversations within online communities were a common measure for a few scholars. This included interaction measures like posts that receive replies, the proportion of posts that receive no replies, and posts that received other forms of interaction (e.g. Kang and Shin, 2006; Soliman, Hafer, and Lemmerich, 2019). These may be seen not just as social interaction, but as the community's capacity to respond (or not) to other members. For example, when a question is asked by a new participant and isn't answered, they may not return to the community.

Other works have built social networks from the interaction data, and used measures of these social networks including node degree and reciprocity (e.g. McGlohon and Hurst, 2009; Tsugawa and Niida, 2019).

I also include interaction in my framework, based on my initial analyses. For later analyses, I borrowed operationalization methods from some of these works. The importance in these works played a role in my deciding to include this as a theme in my framework.

2.6.2.3 PARTICIPATION / MEMBERSHIP

Another series of attributes and measures focused around an individual's participation within an online community. While this can overlap somewhat with interaction, it is possible to participate in a community without any interaction from other people (e.g. the person never receives a reply or other interaction). To that end, this theme captures attributes and characterizations that measure a person's overall activity within a community, regardless of whether they engage in interaction.

2 Background & Related Work

Scholars have used a number of attributes to measure forms of participation including the average of participants age in the community (e.g. Schoberth, Preece, and Heinzl, 2003; J. Zhang et al., 2017), user retention between time periods (Kang and Shin, 2006), measures of content production (Kang and Shin, 2006; Panek et al., 2018). One particularly interesting attribute are the level of equality between all participants when producing content. Many scholars have noted the relative inequality in content production such that a small number of participants may produce a majority of the content. I also observe this inequality in this work, and will describe it further in later chapters.

As with interaction, I do include participation in my framework, but from my own analyses. I did later analyses using some of the characterizations used in these studies (e.g. Gini coefficient for inequality) and many of the other measures used here could be used as an operationalization in my framework. I found both the retention and tenure important characteristics for describing community differences, and these analyses also helped me refine some of the fostering analyses.

2.6.2.4 SIZE, VOLUME & GROWTH

Lastly, many models and frameworks use measures of size or volume. These typically measure the number of active participants or the volume of content that they produce in a community. These attributes may overlap with other themes, but are so common that I want to highlight these specifically.

The general intuition is that the size or volume may influence other aspects of a community. Thus size or volume has been commonly used as a control variable in statistical models. Other models and frameworks, like resource-based models (e.g. B. S. Butler, 2001) have found that the relationships with size and volume may be complicated.

Although many other works characterize communities by size or volume, I do not include it in my own framework. I found that size was often reflected in other perspectives like interaction and participation, and thus did not include it. However, future iterations may include this if it is useful.

3

METHODOLOGY & DATA

In this chapter, I provide an overview of my methods for analysis and the data I use in this work. This overview is intended to provide additional context to the reader for the methods I used and the rationale behind my choices, such that the reader may evaluate this work as a moment in time. I will also provide an overview of the research communities where I situated this work. I follow this with an overview of my methods for this inquiry, and the data used within. Lastly, I will also include some limitations of both my methodology and the dataset which affect how this work should be used.

3.1 HCI & CSCW

To help the reader understand the research questions, the contributions, and the validity of the methods, it may help to situate this work. I generously consider this work most at home within the field of human-computer interaction (HCI) and computer-supported cooperative work (CSCW) which concern themselves with socio-technical systems and their effect on both the people who use the technology, and the technology itself. Both fields are interdisciplinary, bringing together computer scientists, engineers, statisticians, social psychologists, and sociologists among others. Both have been host to several lines of research related to online communities over the last few decades.

The field of CSCW was founded in the 1980s to focus specifically around social issues of groups and organizations around technology, especially within real-world contexts (Greif, [2019](#)). Early work often focused on supporting collaborative *work* (Schmidt and Bannon, [1992](#)) and to understand the *socio-technical gap*, or the gap between what people need to support their work and what computers can technically

3 Methodology & Data

enable (Ackerman, 2000). As technology and the people around them evolved, both fields began to expand their view of technology and work (Crabtree, Rodden, and Benford, 2005) and to explore the effects of such systems more broadly¹. Despite some debate around the focus of the field, scholars within the field continue to use social media and online communities as a topic of study.

Specific to online communities, scholars within the fields have explored a diverse set of challenges around them. These include growing online communities (e.g. B. Yu et al., 2017; Zhu, Kraut, and Kittur, 2012; Zhu, Kraut, and Kittur, 2014), motivations for participating (C. Lampe, Wash, et al., 2010), regulating unwanted and harmful behaviors (e.g. Chancellor, Pater, et al., 2016; Jhaver, Birman, et al., 2019), and issues of privacy and participation within online communities (e.g. Leavitt, 2015). These are just a sample of the topics studied within the field and related fields.

Generation of knowledge in the fields utilize a wide variety of methods from controlled experiments, natural experiments, interpretative quantitative methods, and many qualitative methods as well. Online spaces, however, can create methodological problems with many methods both in forms of data collection and analysis. Online communities can be incredibly large and generate millions of pieces of content monthly. The connections of interactions between members can appear random (Rotman and Preece, 2010). Ethnographic methods like observation, while valuable, may fail to reveal the full context and story of the interaction, especially when other users are involved. Studying large online spaces, some of which have been labeled 'big data', present challenges in terms of data collection, sampling, and analysis, which can necessitate new methods (e.g Rotman, Preece, et al., 2012).

Within online spaces like online communities, a number of previous methods like ethnographic methods have also been adapted to better explore and research these spaces. As such, the mixed-methods leveraged within this work to create the knowledge draws from existing methods within the field.

¹It should be noted here that there was, and perhaps continues to be, debate around the nature of *work* within the field and whether that should be a center of focus for CSCW. For some additional context and information see Schmidt's (2011) reply to Crabtree, Rodden, and Benford, 2005

3.2 METHODOLOGY

In this work, I leverage a grounded, interpretative approach that utilizes both qualitative and quantitative methods. By grounded, I mean that the work is generative and is grounded within the data leveraged. I draw on Grounded Theory (Charmaz, 2006; Glaser and Strauss, 1967) for many aspects of my collection, sampling, analytic process which provides a methodological flexibility, while still being systematic. This is especially true of the earliest part of this work, which is largely qualitative, with hints of quantitative to aid in theoretical sampling (Glaser and Strauss, 1967).

The work relies on a large, “found” dataset made of trace behavioral data from reddit (Baumgartner et al., 2020). With a dataset of such scale, I had to leverage a number of quantitative methods to both explore and understand the underlying data and to help focus on the most theory-productive form of sampling. The qualitative portions are largely focused on the content of the posts and contents that users provided to the site, and while I constrain the sample of data for qualitative analysis, I leverage a second method, trace ethnography (Geiger and Ribes, 2011) to better contextualize the user behaviors around these posts

In the rest of this section, I will first describe some of the methods I draw upon in more detail and provide rationale for my choice of these methods.

3.3 GROUNDED THEORY

In order to answer these problems, I used a grounded, interpretative approach. My approach in for this project has been largely informed by Grounded Theory (Glaser and Strauss, 1967), a general process for developing theory that is derived from data. Glaser and Strauss introduced the general strategy as a call-to-action for sociologists to generate more theory instead of simply using research to verify existing theories.

Glaser and Straus (1967) viewed generating theory as a process wherein the researcher collects data, codes, and analyses at the same time, and the sampling and analysis are done in service of generating theory. The analysis is done through comparative analysis, and the sampling for data collection is done so

3 Methodology & Data

as to maximize the theory (theoretical sampling). As the process is iterative, the researcher will seek out new data that helps develop specific aspects of this theory, and then analyzes that data.

Grounded Theory as a general methodology has evolved as many researchers have adapted the process for their needs and adjusted for their stances². In this work, my work stance is closely aligned with constructionism, similar to Corbin (2009) and Charmaz (Charmaz, 2006). They adopt a postmodernist way of thinking – that knowledge is constructed rather than discovered. Theories and knowledge generated from research are a reflection of both the researcher's past experiences and the world they inhabit, and thus bring to the analysis itself. This work is constructed from my experience within the world, and is a product of that experience, and this should be seen as an interpretation of that experience.

For this work, grounded theory provides me with a framework for developing theory around the formation and evolution of online communities. As noted, the general process of grounded theory is agnostic of quantitative and qualitative methods (Glaser and Strauss, 1967, p. 18), and the methods should match the specific research questions and goals (Corbin, 2009; Morse, 2009).

In the present work, I leverage both qualitative and quantitative data, and the earliest process involved a lot of comparative analysis and memos of both forms of data, where I would look at graphs of the evolution of specific subreddits and compare those against other examples. I would find points of interest, and begin an iterative approach where I would run additional mixed-methods analyses. During the qualitative analysis presented in Chapter 4, my process of coding and note-taking was iterative and adaptive. I also adopted additional methods as new questions arose, which introduced new forms of data to follow up on. For example, looking at aggregate data and post history of individuals involved or visualizations to better understand the passage of time in certain phases.

²For a brief historical perspective on how Grounded Theory has changed, see Morse (2009). As Morse notes, the process isn't "formulaic" and requires adaptation to the specific data and research question.

3.4 SAMPLING

In the first study, I have adapted the general spirit of theoretical sampling (Glaser and Strauss, 1967) to match my goals. Theoretical sampling can be seen as a form of purposeful sampling (Palinkas et al., 2015), wherein the data is collected and sampled with the purpose of developing theory rather than attempting to provide representative, unbiased, or based on a preconceived set of behaviors. I also sample for a specific purpose, but I do deviate.

My goals for the initial study is to understand a broad set of behaviors across different subreddits, and to that end I have sampled to ensure a diverse set of subreddits are analyzed, such that I can better understand the behaviors across size and type of subreddit. My research questions here are broad, but to keep the problem tractable by a single researcher, I select just 8 subreddits that meet a number of criteria. This process is not random.

I filtered out subreddits by activity to ensure that the selected subreddits would have enough data to properly analyze. I then selected 2 subreddits from each of the quartiles of activity from the eligible subreddits such that I can at least explore the possibility that size may affect behavior. I then select among these to match subreddits to a set of known “types” of subreddits – again to ensure that a wide range of subreddits and behaviors will inform the final theory. This process is more closely aligned with *criterion sampling* (Palinkas et al., 2015) in that I have selected subreddits which conform to preconceived samples which I believe may affect collection.

I do later return to the core principle of theoretical sampling, when I start to collect additional data sources to understand specific behaviors as they extend from the subreddits in question. This is largely done as part of leveraging trace ethnography to better understand how some of the accounts are behavior in other subreddits which may add to the theory.

3.5 TRACE ETHNOGRAPHY

When people interact over the internet, the internet servers generally record specific actions over time including additional data from that action – metadata. In some instances, these records are recorded in

3 Methodology & Data

text files to keep a “log” of data, such as HTTP server logs. For social platforms like reddit, these actions are generally stored within a database such that they can be recalled and shown to other users who visit the site.

In my work, I look at six specific subreddits on reddit, which can be seen as a form of multi-sited ethnography. Trace Ethnography (Geiger and Ribes, 2011) is a process for effectively establishing a co-presence with the actors who leave traces of activity on reddit. Through this data, I am able to recreate a partial image of the work individuals do across reddit on their visits, which can help me reconstruct activities and interpret their work.

Geiger and Ribes (2010) illustrate an example of how this analysis process is useful for recreating and contextualizing work by tracing a number of edits on Wikipedia within a specific context to reconstruct the work across distributed actors. In their work, they look at a series of edits occurring from a vandal on a specific page, and the edits and their metadata to help understand what occurred when this happened.

I leverage is a collection of reddit submissions and comments over the course of nearly 11 years. I leverage not simply the text of each submission or comment, but a number of the other metadata fields to help understand not only the work of the people over time, but to also see how these may or may not interact.

I predominately leverage two specific dimensions about which I ‘trace’ the behaviors: by subreddit and by user. The main analysis is done via looking at a temporal slice of all submissions within a specific subreddit to understand how the community developed. However, part of this work can occur outside the subreddit. In some instances, to better contextualize some of what I’m seeing and to uncover some work that would otherwise remain invisible I look at the timeline of actions and content of users across subreddits. For example, when a user posts in a subreddit that I analyzed, in some instances I will look at other comments and posts that user has made outside the subreddit. This is important because part of the work uncovered herein involves how users ‘advertise’ their subreddits such that others might discover it.

3.6 DATA

I draw on a number of data sources in this dissertation. The primary dataset is the pushshift reddit dataset collected by Jason Baumgartner (Baumgartner et al., 2020). The dataset was created through the reddit API by requesting submissions or posts with every single sequential id from reddit.com. All of this data is public, though is not easy to collect holistically using the API through other methods. Baumgartner made this data available to the public for download, separated into multiple files containing individual content types (submissions by month, comments by month, and subreddit metadata).

The data itself consists of the submissions and comments that existed on reddit at the time they were retrieved by Baumgartner and his collection tools³. He stored them as retrieved from the API, as JSON objects for each submission or comment and all of their metadata. While each has slightly different metadata, both do contain elements like the author, the creation date, an id, the subreddit, and its score⁴. Comments have additional information about its submission, and the parent of the comment to preserve the entire tree-like threads. These make up the core data used in the analysis, though there are other fields.

Metadata fields also evolved as the platform itself evolved. This includes fields which were added later in time. The temporal availability of these fields do limit some of the research questions I can reasonably ask given the time periods of my constrained dataset.

I used all submissions and comments from June 23rd, 2005 through December 31st, 2016, which consists of 2.8 billion comments (2,862,749,750) and 313.5 million submissions (313,493,880). This dataset, while large, is incomplete as some of the content had been deleted or removed prior to it being collected (Gaffney and Matias, 2018). I also used the subreddit metadata which was released on February 12, 2019, which contains metadata about 49.7 million subreddits (49,723,295), with nearly 13 million having been created prior to December 31st, 2016 (the end of the post & submission data).

³The dataset was collected over time, which means there's a differential

⁴Although it contains a score, the score may not be accurate. See the data limitations for an explanation.

Additionally, Baumgartner provides an additional dataset for subreddit metadata. This includes information about many of the subreddits and their metadata as reddit provides through their API. The metadata includes many additional pieces of data like the creation time of the subreddit, though it lacks user that created each subreddit – a strange exclusion given that the information was publicly visible on the site prior to 2016. To circumvent this limitation, I leveraged archive.org in the qualitative analysis to know who the subreddit creators were.

Because of the length of time of this analysis, I also constructed a brief timeline on the evolution of the reddit and its features. To do this, I built a python application to scrape the reddit blog, which includes posts from its founders from the October 17th, 2005 ⁵ through December 9th, 2016. I used this as supplemental information to better understand the evolution the platform and its features. For example, to understand when user-created subreddits was enabled by the platform and in understanding when and which subreddits were added or removed as a default subreddit.

Selectively as needed, the work was also supported with additional context via urls to other sites. Reddit began as a link sharing site, and many people continue to use the site in that way. Understanding the site requires viewing these links. Many links however, do not still exist and thus I would once again attempt to leverage archive.org as I tried to build context and depth around the qualitative analysis.

Lastly, I leveraged archive.org extensively to find some missing and deleted submissions and comments, find subreddit creators, and to find missing content from other sites. I also leveraged archive.org later in the process as reddit first moved the URL of their blog, and later removed it altogether ⁶.

3.6.1 RELATIVE MONTHS

I use *relative month* throughout this work to describe periods of 30-day periods relative to the subreddit's creation. Since subreddits are created at different times and days, I need a way to compare their evolution and so I focus on their life relative to the time they were created. I then use 30-day periods relative to that

⁵The exact time may be off because of timezone adjustment.

⁶Some of the blog content is still available on specific subreddits which were created to make announcements to the community like r/announcements. I did not verify which or how many were also available within the reddit data since the archive exists of the blog.

creation to keep comparisons between time periods as similar as possible. So the first relative month would start at day 0 and the second would start at day 30. In graphs and analyses, the relative month is 0-based so the first relative month will be 0. This does complicate some terminology such as years because $12 \times 30 \equiv 360$, which is obviously less than $365/366$, however this trade-off is necessary for accurately “monthly” comparisons. I attempted to use relative years sparingly throughout.

3.7 HISTORICAL ANALYSIS

I am finishing this dissertation in 2026, but the core dataset is from 2005-2016. I want to briefly discuss some of the considerations and issues with publishing a work 10 years after the last piece of data.

I chose to use this dataset for a few reasons. First, the dataset includes an interesting time in the platforms development where things were still changing quickly, with occasional major changes. I focus the 2nd study around the year 2012 as this allowed several years for people to adapt to and develop strategies for creating and developing subreddits. By this time, many of the core features on reddit had been implemented and appeared to be stable. Second, the dataset is large enough that I can explore evolutionary questions about communities started within the data. Lastly, by capping the dataset at the end of 2016, I was able to make the analysis tractable on lower-cost computing hardware, while providing enough data for longitudinal analyses.

The analyses in this work largely spanned from 2020 through 2025. There were some exploratory analyses from 2018 onward, but the published analyses all started in 2020. I had started a separate analysis of r/TD in 2016, but this was shelved for personal reasons in 2017. After political events in the time between 2017 and 2021, I decided to revisit that in a different capacity. However, I changed my research questions and did entirely new analyses leveraging the framework, but the banning of the subreddit made this more interesting.

However, doing this work with such a large gap presents a number of challenges and introduces some potential biases, which I describe in more depth in the limitations section below.

3.8 METHODOLOGICAL SUMMARY

To summarize, this work leverages a mixed-methods approach to understand the behaviors across subreddits. I employ elements of Grounded Theory and Trace Ethnography to answer questions to identify and understand behaviors within the subreddits. I also leverage quantitative methods to look at some behaviors and features of the subreddits across the entire dataset. I explain these in more detail in the chapters for each of the three studies.

3.9 LIMITATIONS

Like all works, this project has limitations. In the course of this work, I have made many attempts to identify and check for possible limitations and to either address them through methods or remain cognizant of them such that I can share those with the reader. I will detail a few of the limitations around my data and methods below, but this should not be considered a complete list. I will also detail a few of the checks I have attempted to perform to ensure data quality or methodological quality.

3.9.1 DATA LIMITATIONS

There are a number of limitations within the data that affected this work. In many cases, they limit types of analyses I am able to reasonably perform, and in others they may impact the reliability of the analyses. The dataset I base this work on is a “borrowed” dataset, meaning I did not perform the collection. Working with a large dataset involves many stages of processing and cleaning, and at each stage new biases or mistakes can be introduced⁷. I do make attempts to test and account for integrity throughout the process, this does not exclude the possibility of errors. I discuss the verification and mitigation strategies in a later section.

The dataset is large, but it is not complete. As Gaffney & Matias (2018) note, the dataset is missing many comments and submissions which may affect certain types of analyses and generalizations. While

⁷For a summary of such pipelines and how they may affect analyses, see (Olteanu et al., [2019](#)).

3 Methodology & Data

we can identify the ids of missing comments and submissions, in many instances we don't know what exactly is missing. For example, the missing data may not be uniformly distributed across all subreddits or users.

The dataset contains deleted and removed comments and submissions, which can affect both qualitative and quantitative analyses. This appears within the data for specific comments and submissions where the text and or the author has been replaced with deleted or removed. Additionally, the data is missing from the dataset as noted by (Gaffney and Matias, 2018), which is seen as “holes” within the data⁸. There are also many instances where usernames or text have also been replaced with the text “[deleted]” which account for different scenarios and behaviors. A similar scenario exists where the text of some content has been changed to “[removed]”. The behavior of the API appears to change depending upon the type of content: comments, text submissions, and link submissions. To further complicate matters, users may also create or edit the text of their message to be “[deleted]” or “[removed]”. This has been the source of confusion among many people using the dataset, and in some instances can make it difficult to know if a moderator removed content or if a user simply created the appearance of a removal⁹.

There is some valuable information missing or inaccurate within the data. For example, there is no way to find the user that created a subreddit from the data. This affects my ability to run analyses on subreddit creators on a large scale. In the Chapter 5, I work around this by using archive.org page captures for a few thousand subreddits, but at larger scales this is not feasible.

The score data is also not accurately captured. There are two issues affecting the score within the dataset. First, reddit ‘fuzzes’ the score to prevent spammers from checking if their bots are working. Second, the dataset was built by individually retrieving all elements, but the time between the retrieval

⁸Holes here refers to gaps within the ids for posts or comments. Database systems and reddit use sequential numbering as an ‘id’ for every new post and submission, so the first comment will be 1, the second will be 2, and the third will be 3. If 2 is deleted or removed, the dataset will contain only comments 1 and 3, thus creating a ‘hole’.

⁹To the best of my knowledge, there are no authoritative answers about how the data is represented and how to best interpret the behaviors leading to these field changes. There are a number of posts where users have attempted to reverse engineer the different states, and periodically reddit admins and developers will occasionally answer questions. One such answer suggests the API may behave differently based on the age of the content when scraped (u/ketralnis, 2022). One additional complication is that the platform may have changed at some point and so the data is neither stored nor reported consistently within the dataset, which can bias analyses using these fields.

and the creation of the submission or comment varies, such that new comments and submissions can still be voted on after the comment or submission was retrieved. Reddit eventually disables voting on older submissions and comments, but as Baumgartner's collection scripts captured more data and more quickly, the delta time between the creation time and retrieval time decreased ¹⁰.

3.9.2 METHODOLOGICAL LIMITATIONS

Collectively, the methods I utilize have a number of methodological limitations which also limit the questions I can reasonably answer with this dataset. Many of the issues have been described in previous work in more detail (e.g Olteanu et al., 2019), but I will detail a few of the more salient limitations and challenges.

My sampling method in the first study focused on a small sample of 8 subreddits for a limited period of time from their creation. The sampling here aligns with my research questions but biases the resulting framework such that it likely compromises the generalizability of the framework beyond reddit and the early formation processes. I also rely on descriptive statistics to help understand some behaviors at scale within this work, which can overlook or render some important behaviors or contextual details invisible. Lastly, like all qualitative methods (and arguably some quantitative-focused methods), the work is interpretive and is subject to my own personal experience and beliefs. I have detailed some which I think may affect my perspective in my positionality statement.

Limited data for theory generation. I leveraged grounded theory to generate a comparative framework for understanding online community development. However, the general process of developing this is time consuming and in order to make progress on the work, I conceded my initial ambitions of looking at more communities and for a longer period of time. As such, this resulting framework will be limited by the few that I chose to qualitatively code.

Filtered Data for low activity subreddits. In the first study, I excluded subreddits below a specific threshold of activity. The intuition here was that I needed enough behavior to analyze to build the theory,

¹⁰Baumgartner does include the retrieval time within each record, and in theory someone could reconcile historical data, but this would mean building a potentially large portion of the dataset over again.

and so focused on subreddits that a minimum level of activity. However, it is very possible that there may be some forms of online community which may fall below this threshold, or that some of these would eventually reach that interesting behaviors exist below that threshold. For example, failed attempts to grow or attract people may be more prevalent in those subreddits than in the ones I studied.

Interpretative Consistency. As a single researcher on the project, I did all of the coding myself. While the coding was done to generate larger themes and theories, there were points at which it can be difficult to interpret the authors intent with some submissions and comments, which can make the process of generating codes and knowledge from those. As such, I provide only my interpretation with no representation of how consistent the codes were applied by myself. I kept a code book, along with notes for every code, which I updated as I coded and would go back and recode elements in addition to doing a second round of codes. This, however, does not prevent other forms of bias from affecting the process.

Limited view of interaction. One of the ongoing digital challenges is understanding how people engage. Trace ethnography can reveal the active work someone does that involves direct interaction with a dataset, but depending upon the available data sources can fail to uncover other forms of work like reading and writing. With this dataset, I am unable to trace how people engage through voting, through messages, or other forms of interaction that remain hidden. There are instances where there appears to be some form of behind-the-scenes communication between moderators and specific members, but the limited traces I have cannot reveal that. Similarly, reading online posts can be seen as an active form of participation within such spaces (Crawford, 2009; Crawford, 2011), but this can be difficult to count as no traces are left.

Historical context & missing data. For the qualitative analysis in the first study, I rely on understanding what people are submitting to the subreddit to understand the behaviors within each. As many of the submissions are at least 10 years old, the original links shared in subreddits no longer work. In some cases, the submission or comment text or the linked content may have changed over time. This can bias the interpretation of some behaviors and norms in unforeseen ways. This is problematic for link and

news-sharing subreddits where early posts could have no text beside the title to help describe what is missing.

A similar but related limitation is around the number of deletions and removals of content by authors or moderators can affect some inferences. In some cases, I could mitigate (described in a later section), but in others I may never know what is missing. This may cause me to incorrectly interpret some behaviors where the important context is lost from these deleted or removed comments and submissions.

3.9.3 DATA VERIFICATION & MITIGATION

Prior to embarking on any of the analysis, the data was processed by a script and inserted into a database. During this process, I ensured that the database accepted only insertions of unique keys, thus triggering errors if the dataset contained duplicate submissions or post. This turned out to be the case with exactly one month's of data. I developed a separate tool to check for duplicate IDs across all files as an additional check, and found a method to mitigate the issue.

I used early testing and later spot-checking to check for and identify some issues with data integrity by comparing records to ensure that what was inserted matched what was in the initial files. At one point, this uncovered issues with the text fields (e.g. title and body) and scripts were modified, affected posts were found, and the data reinserted.

During qualitative coding, rather than coding the data within the dataset or a separate tool, I leveraged URLs and links to the specific submissions on reddit to build a more holistic view of the submissions with included comments and embedded content. This also enabled me to continue spot-checking the data within the database with the content on reddit (or in the case of a deletion archive.org).

An additional set of checks were ran using crafted data queries and visualizations to help understand the overall data integrity. Here, a problem was found with one version of a file (one month of data) and it was replaced with a newer version of the file. First, a script was developed to identify what content changed between the two files. The first file was used as the authoritative source for the available data, but in the case of missing content that was pulled from the newer file. This introduces a specific set of biases

As this dataset comprises around 10 years of data, and analyzed around 10-20 years later, a number of the submissions and comments have been deleted or removed. In some cases, I was able to leverage archival services like archive.org, to fill in some gaps, but this presents an obvious bias that can not be mitigated in an unbiased way. I make attempts to link to the archive.org versions when using it within this paper or to explicitly call out when I have leveraged this data.

3.10 ETHICAL CONSIDERATIONS

Because this data is considered public, my local Institutional Review Board considers this work exempt, and as such it has not been put through review. This work is focused at a subreddit level, and thus should have few potential ethical considerations, however, this is still not risk-free.

Some of ethical considerations for this work arise in the qualitative portion of the analysis. Here, there's a challenge with providing evidence and examples to readers to follow along with the results, while attempting to preserve the privacy of individuals who have posted on reddit. Those posts may still be publicly visible at the publication time of this work, but publishing the users' content limits their autonomy and ability to later remove their content from reddit. When possible, I attempt to limit my use of quoted content and usernames, but this is not complete and remains an issue. I also attempt to use examples which I perceive to cause potentially fewer harms. This means, I have avoided quotes from some subreddits which have specific personal disclosures or other personal information.

An additional ethical concern, is around this work potentially affecting some of the specific subreddits mentioned within. For example, it is possible that some of the behaviors I have mentioned may be later decided by the platform as undesirable and this work may constitute proof for action against the subreddit. Somewhat related, should this work become highly visible for some unknown reason, some smaller subreddits could potentially become flooded by visitors or new users, thus potentially destroying the communities for existing members. While both of these are unlikely, they remain concerns that I think about. As such, I have focused on what I perceived to be less sensitive subreddits.

4

STUDY 1 - BEHAVIORS DURING THE EARLY LIFE OF ONLINE COMMUNITIES

To better understand the creation and evolution of online communities on reddit, I begin with an exploratory study of a small, diverse sample of subreddits. I build a theoretical framework through a grounded, mixed-methods approach that identifies and compares some of the behaviors across subreddits. I aggregate these findings into a framework consisting of multiple categories used to characterize these subreddits and their evolution.

Scholars have used a variety of methods to explore many properties of online communities. Many of the earliest works were qualitative and describing behaviors and interactions between people in spaces like IRC (Reid, 1993), Usenet newsgroups (Baym, 1993), and multi-player games (MUDS and MOOs) (Bruckman and Jensen, 2002; Bruckman and M. Resnick, 1995). Another line of research relied primarily upon quantitative methods, with a long line attempting to identify behaviors or features which explained or predicted growth or failure of the communities (e.g. Backstrom, Huttenlocher, et al., 2006; Cunha et al., 2019; Ducheneaut et al., 2007; Kairam, D.J. Wang, and Leskovec, 2012; Kraut and Fiore, 2014; X. Wang, B. S. Butler, and Ren, 2013). More recently, another line of research has looked at some aspects of the evolution of and migration between communities (Chandrasekharan, Pavalanathan, et al., 2017; Davies et al., 2021; Fiesler and Dym, 2020; Hessel, Tan, and L. Lee, 2016; Newell et al., 2016; Tan and L. Lee, 2015), and some exploring linguistic evolution (Danescu-Niculescu-Mizil et al., 2013; J. Zhang et al., 2017).

The earliest period of a community is interesting because behaviors within this period may affect the ultimate trajectory of that community. Previous work has found that many online communities fail to see

4 Study 1 - Behaviors during the Early Life of Online Communities

activity after several months (Kraut and Fiore, 2014) with many participants may not return (Arguello et al., 2006). This may be an important period for developing rules and norms which help attract and maintain people interested in the goals of the community by providing a productive space for their participation, and to disallow behaviors which may inhibit or prevent that participation.

An understanding of the behaviors during the early life of the community may help us better understand the strategies employed by users to accomplish their goals for the community. Grounded methods enable me to identify behaviors and interpret their intended goal as well as their approach to developing or evolving the community they are within. This may lead to uncovering new behaviors that do not draw from existing research in social psychology or group behaviors that may be specific to reddit, online communities, or even specific communities within reddit. Online communities, especially as they grow, have to deal with undesirable behaviors like spam, harassment, and bullying and many platforms rely on users themselves to create rules and enforce them (Lo, 2018).

Some research from online communities suggests that there are different forms of online communities which may have different types and amounts of some behaviors (e.g. Welser, Gleave, D. Fisher, et al., 2007). Platforms like Wikipedia and Fandom.com are focused on peer-production, where the goal of the communities is to produce a shared informational artifact which can be used as a reference by other internet users (e.g. a Wiki). The behavior in these spaces may be more collaboration focused, though discussion can also occur on talk pages. Other communities like StackOverflow and Yahoo Answers are focused around question-answering from other members of the communities. While both of these forms facilitate those seeking information, the underlying processes which generate the “information” may be very different as one aims to create more comprehensive documents and the other is focused on answering user-submitted questions.

Services and platforms like Usenet and Facebook enable users to easily create online spaces and with flexibility to create different forms of communities. For example, they all support discussion and question-answer communities, and the newer platforms enable the use of multimedia enabling image-sharing and

4 Study 1 - Behaviors during the Early Life of Online Communities

meme communities. Facebook famously noticed the presence of Groups in local communities for buying and selling goods, which formed the foundation of a new product, Facebook Marketplace (Constine, 2016)¹.

Researchers have also identified a number of different forms of communities on Usenet and Facebook. Fisher and colleagues categorized some online Usenet communities as being either flame, discussion, social support, technical, or political (D. Fisher, M. A. Smith, and Welser, 2006). Kraut and colleagues looked at a number of group types including friends & family, identity based, task oriented, and topical (Kraut, Levine, et al., 2020). Past research has also found that behaviors can vary even within some forms of communities, for example, some yahoo answers forums contained discussions and support-giving in addition to simple question-answering (Adamic, J. Zhang, et al., 2008).

This work addresses an interesting gap of behaviors within existing research for the period between the creation of an online space and the formation of an online community. Previous work has explored more established online communities, phenomena on developed communities, and many factors which may affect future activity within the subreddit. In this work, I develop a framework that can be utilized to study and contextualize online communities, with a specific focus on early evolution. The framework is developed through a grounded analysis of the evolution of eight online communities, thus helping address the gap.

Put more specifically, this work seeks to answer the following research questions. RQ1: How do online communities evolve on reddit? RQ2: How do the behaviors differ between diverse subreddits?

4.1 METHODS

In this study, I use a mixed-methods approach based on Grounded Theory (Charmaz, 2006; Glaser and Strauss, 1967). This method was intended to generate theory using iterative analysis and a sampling approach that enabled new data to answer questions as they arose. Grounded Theory may be best known in qualitative works, but Glaser and Strauss (1967) proposed that the general method could be used with

¹As noted in the citation, this was actually the second product named Marketplace after a similar initiative launched in 2007 failed to gain traction. This second iteration began with what was known as Buy & Sell Groups, which lived alongside the second iteration of Marketplace.

quantitative methods as well, and some scholars have proposed more specific methods of leveraging quantitative methods in Grounded Theory (e.g. Berente and Seidel, 2014; Nelson, 2020).

I selected a diverse sample focused on eight subreddits to provide a range of behaviors across different subreddits. I used a qualitative approach involving a multi-pass coding scheme and memoing to aid in keeping tracks of developing themes and questions. I explored some questions during the analysis using additional methods like Trace Ethnography (Geiger and Ribes, 2011) to trace account activity beyond the subreddit and quantitative methods to understand the behaviors at a higher level. I also leveraged descriptive statistics and inductive reasoning based on graphs of statistical data from the subreddits to answer additional questions and incorporate into my analysis.

4.1.1 DATA & DATA PREPARATION

I created a sample of 8 subreddits from the larger pushshift dataset (Baumgartner et al., 2020) described in the Methodology chapter. The original dataset comprised of comments, submissions, and subreddit metadata from 2005 through the end of 2016. The larger dataset contained 986 thousand subreddits that had been created through the end of 2015 (1.3 million by 2016). However, 46% saw no activity by the end of 2016 and only 8% (77,573) subreddits had more than 100 pieces of content by the end of 2016. The sample I derived focused on subreddits created between 2010 and 2013, such that all of these subreddits had at least 400 submissions to code in the qualitative analysis². I use the data through 2016 for the 8 subreddits in most of the quantitative analyses and some of the descriptive stats.

I also leveraged archive.org's archived page data to further extend the dataset. I leverage this in a few different ways, but in this study it is used primarily to retrieve some missing data that was critical for constructing history and context of the subreddits and the behaviors within those. For example, the subreddit metadata does not contain information about which user created the subreddit, which is important knowledge that cannot correctly be inferred by comments or submissions³. Many of the earlier

²The 400 submissions minimum was more of an issue for the smaller subreddits, which produce less content, thus requiring more time to reach that threshold.

³While it is possible to assume the author of the first submission was the creator, this is problematic given the amount and patterns of deleted and removed content within the dataset. Several of the subreddits in this work have the first piece of

submissions and comments are also missing from the dataset, which can be important to understanding some of the early behaviors. The list of moderators, like creators, are available on the main page of the subreddit and sometimes an additional page for moderators. Archive.org has historical page captures for many subreddits, including some small and medium sized subreddits.

I also collected a secondary dataset of all submission data for the eight sampled subreddits for the purpose of getting updated metadata. The pushshift dataset contains submissions and comments that were collected at varying times from when the submissions were created. Some submission metadata in the Pushshift data can be inaccurate depending upon how long it was collected after the content was created. There are two primary challenges here that need working around. One is the 'fuzzing' of the score the platform uses to make it difficult for bots to detect whether they can manipulate scores well. The second is that people can vote and comment on posts months after they were created. I utilized the scores in a few of the analyses around interaction.

4.1.2 SUBREDDIT SELECTION CRITERIA AND DIMENSIONS

To ensure a range of diverse subreddits, I chose a modified form of stratified sampling across three separate dimensions. This method of sampling across dimensions is similar to the modified stratification used in prior works (Ridings and Gefen, 2004; X. Wang, B. S. Butler, and Ren, 2013), which also aim for a broad sample of communities. I selected eight subreddits that were spread across these dimensions to analyze in more depth. The dimensions were informed both through existing research in online communities and my own prior knowledge of online communities, including reddit. These include content volume, purpose and type of content, and topical category. I also employed lower bound thresholds for inclusion in the sample including number of authors and total activity to ensure there would be enough data to analyze, while also excluding some subreddits dominated by bots.

I chose a minimum inclusion criteria of 5 unique authors, 1000 pieces of total content, and 400 submissions across the subreddit's entire lifespan. This provided a minimum amount of data for analysis

content where the author is listed as 'deleted' and thus not recoverable. Additionally, as seen in the data in Chapter 5, there are a number of successful subreddits where the creator never participates.

4 Study 1 - Behaviors during the Early Life of Online Communities

Purpose / Format	Description	Example Subreddits
Discussion	Focused more on discussion topics and article sharing for discussion. Less focused on	MakeupAddiction, biology, Fitness, Parenting
Question / Answer / Advice	Focused around people asking questions and getting answers. These can be objective as in AskScience or more advice focused like AskRelationships.	AskReddit, AskScience
Media/Meme	These focus largely around sharing media or memes.	Pics, Videos, and memes
News / Updates	Subreddits which share more information. These are anchored almost entirely around links to other sites	News, WorldNews, Politics, UpliftingNews

Table 4.1: The four subreddit formats used for sampling. These are not mutually exclusive, as evidenced by some subreddits including multiple.

and some of interaction, but this also excluded the vast majority of all subreddits created through 2016. Nearly 98% of the subreddits in the dataset did not have at least 1000 pieces of content and 5 unique authors. The additional threshold of 400 submissions is necessary for a minimum amount of content for the qualitative analysis. This resulted in 22,338 eligible subreddits (1.7%).

4.1.2.1 CONTENT VOLUME

Content volume is one measure of size, and has been known to affect both the subreddit and its behaviors as well as the perceived feel of the subreddit. I measured volume based on the combined total number of submissions and comments from its creation and the end of 2016. Size has been found to contribute to perceptual differences (Hwang and Foote, 2021; TeBlunthuis, Kiene, et al., 2022) and behavioral differences (Backstrom, R. Kumar, et al., 2008; TeBlunthuis, Kiene, et al., 2022; J. Zhang et al., 2017) in online groups and communities. I chose 2 subreddits in each of the quartiles of total activity, and the two from each quartile provides some representation across the scales.

4.1.2.2 PURPOSE AND FORMAT

The second perspective focuses on the purpose of the subreddit and the type of content it supports. For example, one popular style of subreddit are the “Ask” subreddits (e.g. r/AskReddit and r/AskScience),

which exist for people to ask questions. Another type of subreddit that exists focuses on content sharing, sometimes generated by users. This includes a number of subreddits focused around art, photos, and memes. Other examples of types of observed subreddits include discussion subreddits, news sharing subreddits, and advice subreddits. Prior research has also found that group types may affect behavior (X. Wang, B. S. Butler, and Ren, 2013), be related to other structural attributes like size (Kraut, Levine, et al., 2020), different communities can provide differential benefits (TeBlunthuis, Kiene, et al., 2022), and that multimedia content has grown popular on reddit (Singer et al., 2014). However, even within a broad category like mental health, behaviors may vary (Park, Conway, and A. T. Chen, 2018).

For this perspective, I identified four high-level subreddit purposes and formats as seen in table 4.1. I included at least two of each by browsing content within each subreddit to gain a general feel for the type of content shared and the purpose. These categories are not mutually exclusive as some communities support a variety of different forms of content, and some of the sampled subreddits supported multiple.

4.1.2.3 TOPICAL CATEGORY

The last difference is one around the topical category of subreddit. Many subreddits are focused around specific topics of varying granularity. For example, r/science is focused on science related content and posting political content would potentially be removed by moderators of the subreddit. Using an inductive quantitative analysis, I identified several high-level topical categories on reddit, and used this to select a single subreddit from each of these, which allowed me to see some behaviors that may differ based on the topical category of subreddits. Some work has found some differences around some behaviors like advice seeking (Galegher, Sproull, and Kiesler, 1998), and the prevalence of bad behavior (Cheng et al., 2017). However, a

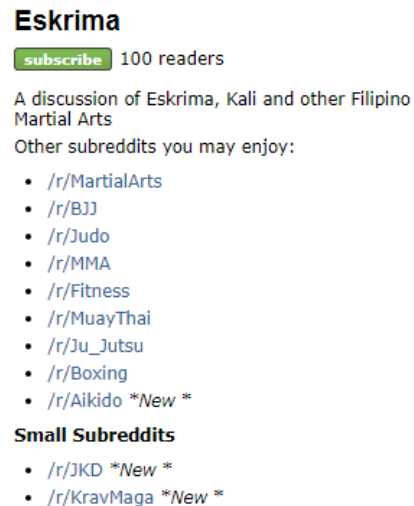


Figure 4.1: Example of subreddits linked from r/Eskrima's sidebar which are used to find similar subreddits.

4 Study 1 - Behaviors during the Early Life of Online Communities

number of works focus more on topical groups as a distinction

between other large classes of groups like hobby or identity groups (e.g. Chancellor, Hu, and De Choudhury, 2018; Kraut, Levine, et al., 2020).

I identified these categories by identifying clusters of subreddits using links in the subreddit's description. This relies on a common practice for subreddits where they include subreddit links to similar subreddits in the description as seen in figure 4.1. I first built a network graph of these subreddits, and used a combination of a community detection algorithm and inference to identify topical clusters within the graph (Figure 4.2). I excluded a few larger clusters some of which seemed mostly populated by bot content (e.g. the Imaginary* and the *Pics clusters) or were topics I chose not to manually code (e.g. the pornography cluster). I leave these additional clusters as areas for future research. The boundaries from the automated detection were often blurry, with some subreddits that bordered multiple. In some instances, I grouped these clusters together in a larger category (e.g. Health & Fitness).

Category	Seed subreddits
Health & Fitness	health, fitness, food
Science	Science, askscience
Technology	technology
Sports	Sports, nfl, nba, nhl, CFB, baseball, soccer, FIFA, MLS
Gaming	gaming, games, pcgaming, nintendoswitch, n64, 3ds, nintendo, playstation, xbox, xboxone, ps3, ps4
Relationships / Sex / Identity	relationships, relationship_advice, TwoXChromosomes, Men
News & Politics	news, politics, worldnews
Entertainment	music, aww, memes, funny, jokes, movies, books, gifs, television

Table 4.2: This table contains the identified categories along with the subreddits used to core identified subreddits. Note that while some are simple like r/Technology, some of these can be insular like the Gaming cluster which includes a large number of subreddits but often focused on specific platforms.

To focus on more representative categories, I chose multiple 'seed' subreddits from these clusters, which were typically larger subreddits like r/Fitness and r/science. Some topical categories, like gaming and sports, were more insular and so multiple seed categories were used. These are visible in table 4.3. I then induced a subgraph by taking all subreddits connected to those subreddits with a distance of one. I chose one through exploration of results with different values, and found one yielded the most focused results.

4 Study 1 - Behaviors during the Early Life of Online Communities

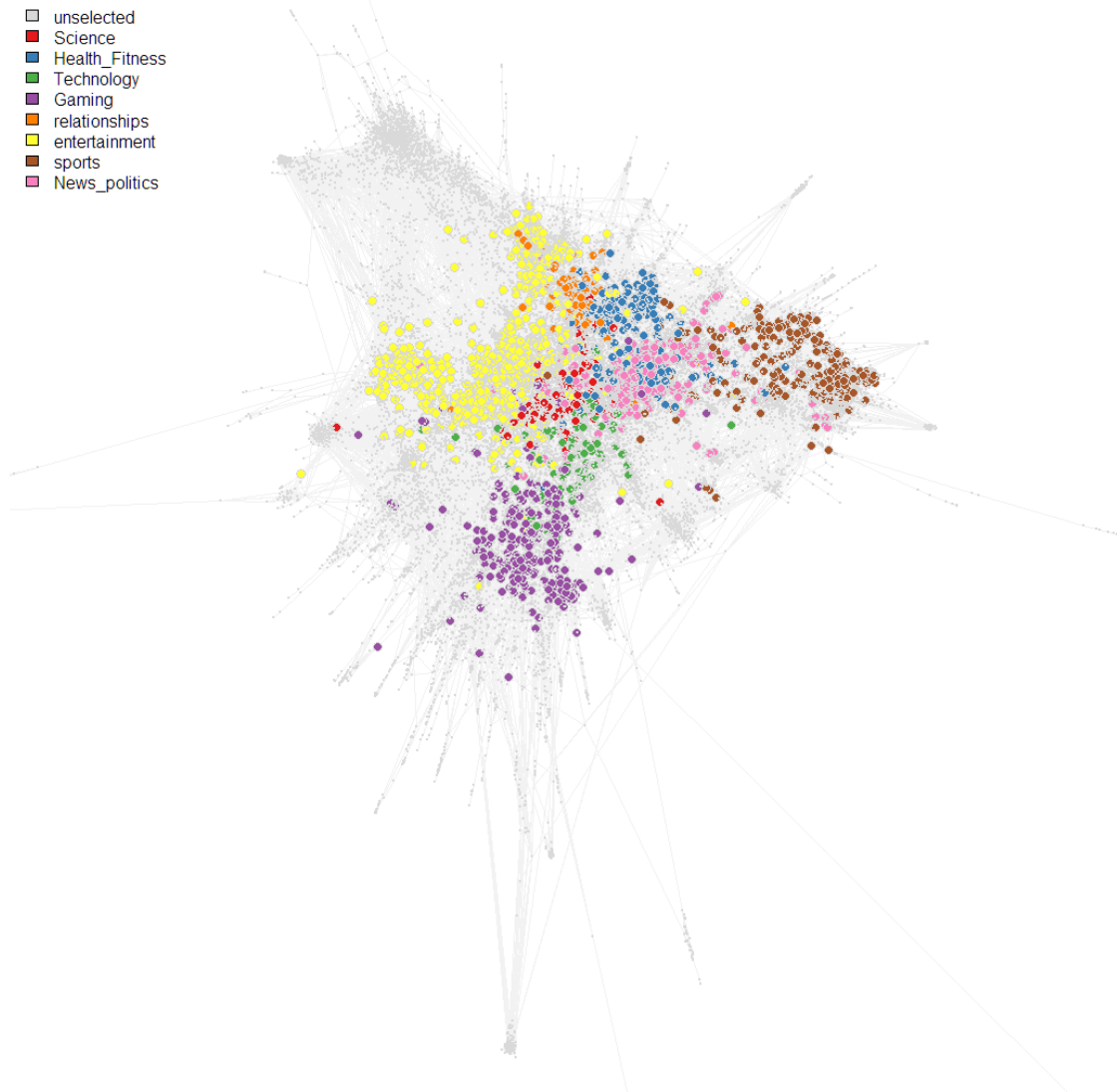


Figure 4.2: Network graph built using sidebar links from one subreddit to other subreddits. The colors indicate the topical category of those subreddits.

4 Study 1 - Behaviors during the Early Life of Online Communities

Subreddit	Category	Quantile	Format/Content Type
AnimalsFailing	Entertainment	0	Media/Meme
FZero	Gaming	0	Discussion
Eskrima	Health & Fitness	25	Question/Answer, News/Content Sharing, and Discussion
EarthScience	Science	25	News/Content Sharing and Discussion
LinusTechTips	Technology	50	Discussion, News/Content Sharing, Media/Meme
WomensSoccer	Sports	50	News/Content Sharing
AskWomenOver30	Relationships, Identity, and Sex	75	Question/Answer
UpliftingNews	News & Politics	75	News/Content Sharing

Table 4.3: Summary table of the selected subreddits.

This method is similar to other approaches for finding similar subreddits, but relies on existing metadata within a subreddit, which was not always present. Some other methods have used text or membership overlap to identify similar subreddits, though these do not attempt to infer categories from the similarity. The reliance of this method on description links means some subreddits were not considered because of their lack of these links.

4.1.3 SELECTED SUBREDDITS

Using the three dimensions, I first began with the topical category. I generated summary statistics for each subreddit within the topical category. I used a mix of sampling strategies to ensure that I selected one subreddit from each topical category, two from each size quartile, and at least one from each of the format/content type categories. As there are fewer than eight format/content type categories, several were represented multiple times.

The final list of selected subreddits can be seen below (Table 4.3) along with the matching sampling criteria for each of the eight subreddits.

4.1.4 ANALYSIS

I used an mixed-methods approach based on Grounded Theory (Charmaz, 2006; Glaser and Strauss, 1967). This offered a iterative method for identifying and structuring themes across the diverse sample of subreddits. This also enabled me to mix qualitative and quantitative methods in order to explore additional research questions and to triangulate across forms of data as new questions arose.

I began with a qualitative analysis of the first 400 submissions in each of the eight subreddits. I analyzed this using an open-coding approach for five subreddits on the first pass with memo-ing as I coded. I chose 400 as a means of limiting the total amount of work while providing a reasonable sample of the earliest behaviors in each subreddit. In this first pass, I applied 225 total codes and generated 88 unique codes across the first five subreddits. Sixty-one codes were used only a single time (though this included typing mistakes and variations on other codes).

After coding 5 of the subreddits with this open-coding approach, I stopped finding new larger themes. I regrouped the original codes into a closed-coding scheme and recoded all eight subreddits, still updating and adjusting the codebook to clarify examples. I revised the codebook as I came across new data, and revisited earlier codes to adjust as necessary. I also continued to memo throughout the process.

For both rounds of coding, I occasionally extended leveraged additional data to explore some questions. This included other datasets like historical archive data on archive.org and data outside the sample described earlier (so beyond the sample of 400 submissions from the 8 subreddits). The archive.org data was often used for content that had been deleted to better apply some codes, or to better understand an accounts activity within a subreddit.

Some of the interesting behaviors occur outside the subreddit of interest (beyond my sample). To help account for this, I employed elements of Trace Ethnography (Geiger and Ribes, 2011) to trace through other parts of the dataset and find behaviors, outside the subreddits themselves, that can help explain some of the observations made within the dataset. These were not explicitly coded, but were used in an attempt to triangulate and theory build, as these tended to be either occasional behaviors. As part of the

triangulation, I generally used these investigations as the basis for additional analysis, when they would further refine the results.

4.1.4.1 ITERATING WITH QUANTITATIVE METHODS

Some concepts from the memos yielded additional questions. I used some quantitative methods to explore the prevalence of specific content or behaviors. This generally focused on operationalizing a behavior and generating statistics and visualizations to answer these questions. For example, I noticed several forms of advertising subreddits, and this led to questions about how common this ultimately was across the subreddits in addition to outside of it.

These explorations developed the theory by providing a means of providing additional evidence to the qualitative methods. This provided a means of separating some less common behaviors from more prevalent ones, while also helping find additional examples to both stress or expand the conceptual dimensions. In these cases, some limited additional coded did occur, which used simple thematic analysis.

4.1.4.2 REPORTING AND ANONYMITY

In this chapter, I include some quotes from a few accounts and also some activity graphs of their activity. I have avoided reporting names in the final draft of this document to allow these accounts to potentially distance themselves from the content. In most cases, the names are not necessary because what matters for contextualizing this is whether they were the creator, a moderator, or a normal account. I include that data when describing the account.

4.2 OVERVIEW OF THE SUBREDDITS

It is helpful to first look at some basic descriptions and statistics about the eight different subreddits used in this study. This is intended to serve as both transparency about this analysis, but also to provide context for the characteristic themes described in the next section. These subreddits again were selected

4 Study 1 - Behaviors during the Early Life of Online Communities

across multiple dimensions, with one dimension being stratified across the volume of content across their lifetime (through 2016).

Subreddit	Creation date (Y-M-D in UTC)	Total Activity	Total Submissions	Total Comments	Unique Authors
AnimalsFailing	2013-08-24	1749	454	1295	976
Fzero	2011-06-10	2460	599	1861	722
EarthScience	2011-03-08	2757	1876	881	746
Eskrima	2010-09-09	3033	452	2581	417
LinusTechTips	2012-09-29	6135	1635	4500	2165
WomensSoccer	2011-08-17	7401	3452	3949	1080
AskWomenOver30	2013-08-22	49960	2222	47738	5005
UpliftingNews	2012-05-16	916314	98295	818019	253548

Table 4.4: Descriptive statistics for each of the 8 sampled subreddits. These are ordered by their quartile with the lowest activity ones at the top. Total activity is the sum of submissions and comments.

As visible in the above table, table 4.4, there were 2 subreddits selected from each of the four quartiles of the eligible subreddits (having at least 400 submissions in order to code and over 1000 total pieces of content (“activity”)). This includes a wide range of subreddits across many size, with a specific focus on smaller subreddits, which may be less studied by other research.

In looking at the graphs of submission and comment volumes over time (Figure 4.3), there are a few things of note. First, the two largest subreddits began large. Both had hundreds of comments within the first month. This is a bit of a puzzle. The smaller subreddits tend to have somewhat unstable levels of activity, often with months of seeing no comments or submissions. The one exception to this is r/EarthScience, which has a steady stream of submissions, largely from the subreddit's creator.

Another interesting difference is that as many subreddits grow, the number of comments tend to exceed the number of submissions by a large margin in many subreddits (note that the graph here has a y scale in logarithmic units, so each horizontal line represents a power of 10 over the previous.). Again, the notable exception here is r/EarthScience, which appears to operate in a slightly different capacity than the other subreddits, perhaps being more informational than interactional.

4 Study 1 - Behaviors during the Early Life of Online Communities

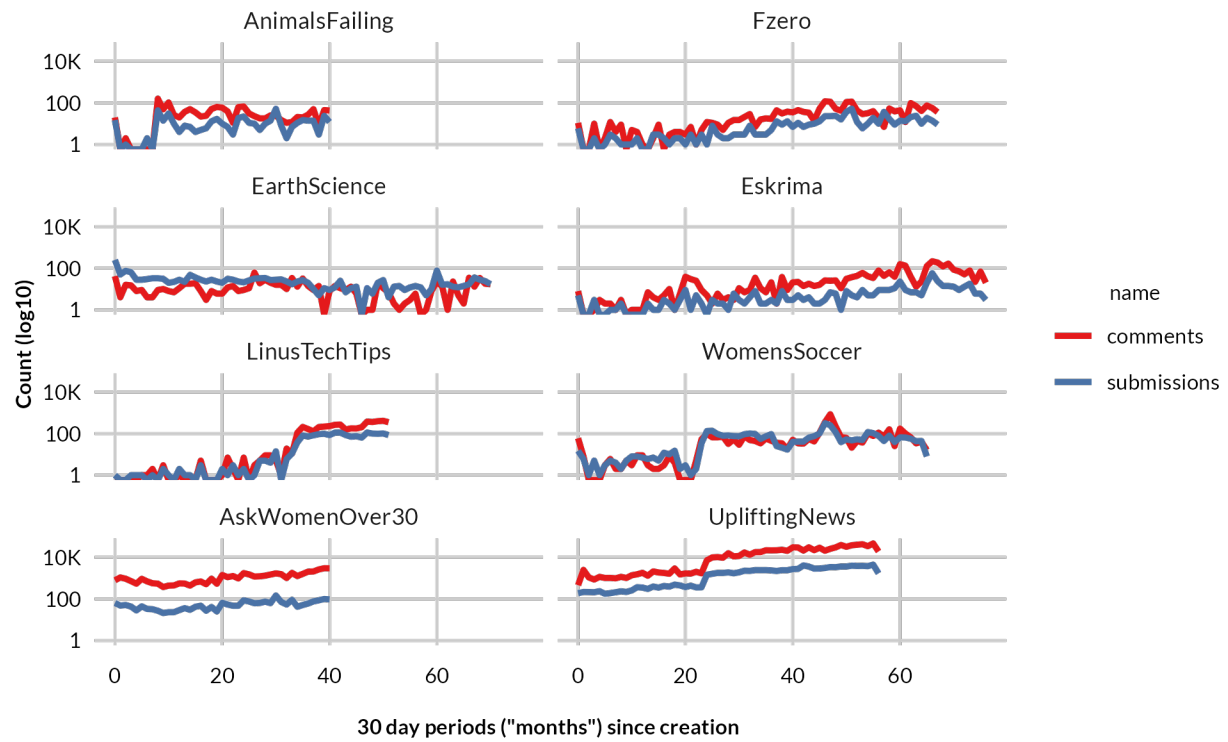


Figure 4.3: This chart shows the lifetime of each of the subreddits from creation through the volume of activity (comments and submissions). Each point is a 30-day theoretical month such that subreddits are aligned at creation despite having different creation dates across years.

4.3 RESULTS: THEMATIC CHARACTERISTICS FROM THE SAMPLED SUBREDDITS

I organized the results of the analysis into five broad themes: content composition, interaction, participation, fostering, and promotion. These describe different types of behaviors and characteristics of the subreddits. In the following sections, I describe each of these five themes and compare the subreddits along a number of different dimensions within these themes. I present these as smaller analyses designed to help illustrate how and why these themes can be useful for comparing subreddits/communities, and in understanding their evolution.

4.3.1 CONTENT COMPOSITION

During the qualitative analysis, I found that subreddits differed in the types of content submitted to them and how much of different types of content were submitted. Some of the subreddits were focused on producing specific types of content, like images and video in r/AnimalsFailing and questions in r/AskReddit. Others were focused around posting links to content on other websites. Despite selecting subreddits based partially on their purpose and content types, I found it surprising that all eight subreddits contained a mix of different content types. There were variations in the ratios of the content and variations over time. The proportion of the content may hint of the subreddit's purpose and the underlying norms. Changes in this content may signal changes in norms, audience, or even platform-level changes.

On the platform as a whole, the content itself has also evolved over time. Initially, the platform only allowed links to be shared, before eventually enabling users to create text submissions. Between 2008 and 2012, people began submitting more image and text submissions (also known as self-submissions as the link is to the submitted post). Images on reddit by the end of 2012 had received around 85% of votes, but text (self) submissions attracted 50% of all comments despite making up roughly 40% of the submissions (Singer et al., [2014](#)).

There are many ways to compare the types of content within a subreddit, but here I focus on the two I found most interesting. In the first, I look at content from a metadata level and compare the subreddits

ratio of text and links submissions. In the second, I present an analysis of the forms of questions present in text submissions.

4.3.2 SUBMISSION TYPES (LINK VS TEXT)

Every subreddit had both links and text submissions (Figure 4.4). Even image/meme-sharing subreddit like r/AnimalsFailing had a small number of text submissions (3 total). Two were requests for pictures and one from a moderator discussing a spam wave and asking people to report any spam. In looking at the percentages of links vs text submissions, some subreddits appear to have preferences either from the rules or norms set by the community, beliefs that norms exist, or by the submitters' preferences. The differing ratios are not surprising given the sampling criteria included an image/meme subreddit and a Q&A style subreddit, but the different ratios may be useful for characterizing the subreddits.

A person's decision to submit a specific form of media to a subreddit may not necessarily represent what the subreddit's members want or value. For example, an image sharing community may not want to see text submissions. There are two possible proxies for exploring what the community members may see as valuable: commenting and voting. Comments require time to type out, so one or more comments may indicate the submission has some value to members. Additionally, reddit's voting is one method for the members to enforce norms around content by downvoting undesirable or less valuable content, and conversely upvoting valued

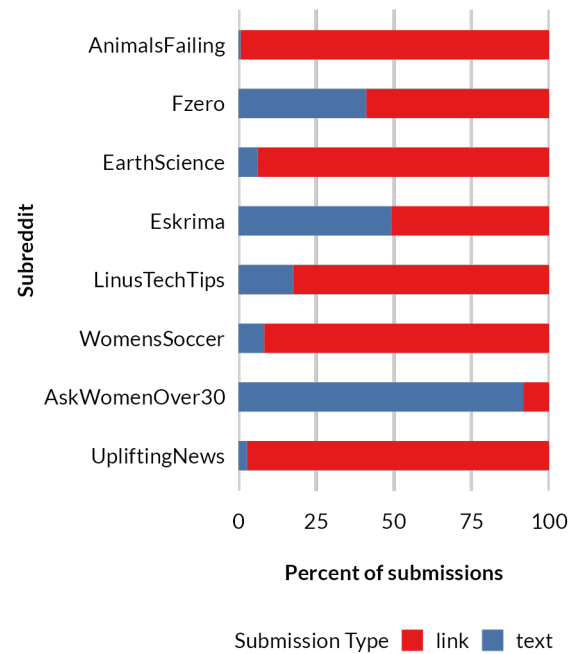


Figure 4.4: This bar chart shows the percentage of all submissions that were positively engaged during the sample window for each of the sampled subreddits. Of note is that all subreddits include some level of positive engagement for both types, but the overall ratio within each subreddit varies. Removed posts are excluded from this graph.

4 Study 1 - Behaviors during the Early Life of Online Communities

subreddit	submission type	total	engaged count	engaged percent
AnimalsFailing	text	3	3	100.00 %
AnimalsFailing	link	449	329	73.27 %
Fzero	text	241	221	91.70 %
Fzero	link	345	223	64.64 %
EarthScience	text	117	68	58.12 %
EarthScience	link	1750	1088	62.17 %
Eskrima	text	222	204	91.89 %
Eskrima	link	230	176	76.52 %
LinusTechTips	text	283	248	87.63 %
LinusTechTips	link	1329	1151	86.61 %
WomensSoccer	text	284	253	89.08 %
WomensSoccer	link	3163	2538	80.24 %
AskWomenOver30	text	1958	1674	85.50 %
AskWomenOver30	link	176	52	29.55 %
UpliftingNews	text	2676	437	16.33 %
UpliftingNews	link	95618	38180	39.93 %

Table 4.5: Each of the 8 sampled subreddits and the counts of submissions for both link and text submissions (aka. self posts). This also includes the number of each that were positively engaged by other users by votes or comments to focus on submissions valued by other users.

content. This influences what reddit's algorithms show to other members. Other works have used similar formulations to measure the reception of content as a proxy for what the (Datta and Adar, 2019; Waller and A. Anderson, 2019; Weld, Glenski, and Althoff, 2022).

By focusing on only submissions with at least one comment or that have a score greater than zero ⁴, I can estimate what types of content the subreddit may value.

While this may provide a reasonable estimate in some communities, it should also be interpreted with caution since reddit allows anyone to vote or comment and not just subscribers of the subreddit and there are a number of factors which may bias this number for some subreddits (such as increases in spam, brigading, harassment, moderators disabling specific types of submissions, etc). Additionally, there are a number of situations where submissions can receive comments which are not positive, such as comments that complain or are intended to provide feedback on a low quality or unwanted submission. The latter such case is a use case for AutoModerator.

⁴While submissions can have negative scores, reddit does not show negative submission scores in its interface or in the API.

4 Study 1 - Behaviors during the Early Life of Online Communities

Looking at table 4.5, we can see that a higher percentages of text posts are engaged across 7 subreddits with the exception being r/UpliftingNews. Some subreddits like r/AnimalsFailing had very few overall text submissions. The differing rates of engagement of text and link submissions across subreddits suggest that both types were still valuable, though one may be less valued within a particular subreddit. However, the exact content of those text or image posts is important, and potentially more so than the simple distinction between text and link. For example, a text post in an image/meme sharing subreddit that includes an update to rules about spam would likely be valuable, but a question may not. Reddit has also provides affordances for moderators to prevent or allow submissions of specific media types from being created. This appears to have been employed in r/UpliftingNews, as the last text post was on May 18th, 2015.

Links to image and video may provide more entertainment value than the longer text submissions in some subreddits. This was true in r/AnimalsFailing, which focused on more humorous videos of animals. Some image submissions, also showed up in r/LinusTechTips.

Many of the shared links were likely intended to be informational, including some news content in subs like r/UpliftingNews and r/EarthScience. In r/LinusTechTips, a large portion of the submissions were links to YouTube videos from the LinusTechTips channel, which were eventually syndicated by a bot when the videos were published.

Some subreddits like r/FZero, r/WomensSoccer, and r/Eskrima were more mixed in their content including links to recent news articles, videos, etc. In these, especially in r/Eskrima and r/FZero, sometimes articles and video links would be submitted with additional context including questions about the specific link (e.g. What do you folks think about this?) or indicating that the content is something they created and wanted to share (e.g. Here's something a video I made of me practicing eskrima).

4.3.2.1 TYPES OF QUESTIONS

A common usage of the subreddits was to ask questions of its participants. While these types of posts were most commonly text submissions, a few did occur through the usage of the submission title and a link

to some piece of content (e.g. What jacket is that person wearing?'). I distinguished between 3 specific types of question types across the subreddits. It was common for many submissions to contain multiple questions, which often included multiple types of questions. Questions leverage the collective knowledge of the members of the subreddit as a resource to get information, advice, and resources.

4.3.2.1.1 INFORMATION SEEKING The most prevalent form of question involved seeking information including about facts or people's experiences. This was sometimes used as a way to get a better perspective on particular issues and learn about things they were not familiar, where people's experiences and opinions might be valuable for thinking through a situation. I observed these types of questions in 6 of the 8 subreddits.

In r/LinusTechTips, someone created a submission asking what specific wrist watch the host was wearing in a video. The desired answer here was a make and model of a wristwatch.

Another common type of information seeking question occurred when people were using the subreddit as a survey of their collective experience. In r/Eskrima, one person asked what they should look for when trying to find and evaluate an Eskrima school. This type of question was also common in r/AskWomenOver30. For example, one person shared a personal story and asked if others had experience with marrying past a certain age, how they felt about it, and how the relationship changed over time. These questions are not simply trying to start a conversation. The author genuinely wanted an answer. This survey style question leveraged the collective participants to get a range of answers

4.3.2.1.2 ADVICE SEEKING In four of the subreddits, there were questions focused on seeking advice for the authors. This included advice about their life and recommendations. Sometimes this was as simple as a reputable store or which type of equipment they should look buy (e.g. r/Eskrima). Other times the questions revolved around larger decisions in their lives. In r/AskWomenOver30, this was often about personal situations.

Many of the advice-seeking questions include short stories where the authors share personal details about the specific situations to provide context to readers. The context varied in length from a single line

to lengthy descriptions with many details. In some cases, authors used the specific details to help find other people with more similar experiences in an attempt to get more relevant responses. For example, in one example a woman posted to r/AskWomenOver30:

Having doubts about my relationship and wanting some advice from someone my own age(F40).

...

What do I do?

-<Anonymized>

In this instance, *advice* is an important distinction because it leveraged the collective experiences and knowledge of the community. Advice in many of these instances relies heavily on the personal context of the situation, and so these sometimes developed into conversations where questions were asked back to the original submitter. There was occasionally overlap between advice seeking and information seeking, often in the form of submissions with multiple questions, often looking to survey other people's experience.

Advice seeking was less common in some subreddits like r/LinusTechTips, but people did occasionally ask them. For example, one person asked about what they should do with older computer parts, specifically if it was worth selling them or doing something else.

4.3.2.1.3 RESOURCE SEEKING People also turned to these subreddits to find resources beyond the subreddit (e.g. books). These were topically related to the subreddit and often the questions were looking for broad information, authoritative sources of information, and resources physically close to them. This form of question was present in 7 of the 8 subreddits, though in lower quantities than information-seeking and advice-seeking questions.

4 Study 1 - Behaviors during the Early Life of Online Communities

Many of the resource-seeking comments occurred in r/Eskrima, which is a Filipino martial art involving training and specific gear. One person posted looking for resources, like books, on the history of the martial art. Others would post asking for sites or places to buy gear.

Similarly, in r/AnimalsFailing, one of the few text posts asked for images beyond the subreddit. Here the expectation was that because the subreddit is often posting pictures of similar types of animals that they may know of some resources beyond the subreddit. The request isn't for a specific image, but a place to find these images. The place to find that is a resource.

While resource requests were less common than other seeking behaviors, they still leverage the knowledge of the community in hopes people have familiarity in more specialized topics that may be less appropriate for the type of information sharing in the specific subreddit. In the first example, the history of the Eskrima may be more complicated or longer than can be easily answered within a few comments. The poster may also have expected to not find an authoritative expert on the history within the community prompting them to ask for more authoritative resources on the internet.

One additional type of resources that people sought were local resources. I found examples of these in 3 different subreddits, but the majority of these questions happened within r/Eskrima. There people looked for local places to train or for resources to buy equipment. For example, one person was looking for a new group to train with in his town after having lost them. Other examples included people looking for sites or even local bars to watch soccer matches in r/WomensSoccer or people looking for local arcades with an FZero game in r/FZero.

4.3.2.1.4 QUESTION TYPES AS A CHARACTERISTIC All three of these question types were present across the subreddits, though in varying proportions, suggesting that they may offer an alternative method for characterizing the communities. In each case, the questions written to the communities cast the collective knowledge of the community members as a whole as a resource for different forms of information. This is inline with research on other forms of online communities including newsgroups (e.g. Arguello et al., 2006), question/answer platforms (Liang and Introne, 2019), and many types of forums. The perception of the communities as a place to bring questions is interesting. In a day of search engines, people still sought to

ask these questions in a community. With the growing popularity of LLM-based AI, it will be interesting to see if people will continue to ask questions of communities in the same ways and in similar volumes.

4.3.2.2 CONVERSATION STARTERS

I also observed the behavior of people trying to start conversations in six of the subreddits. These typically involved people sharing a relevant link and asking broad questions to the community. The questions were not seeking a specific answer, but were intended to engage the audience through stimulating discussion. These were most common across the more text-heavy subreddits like r/AskWomenOver30, but some also occurred in more link-sharing subreddits like r/WomensSoccer and r/UpliftingNews.

In their simplest form, many appeared as simple questions, often about someone's life and experiences, which make them easy to reply to. Take this early post in r/AskWomenOver30 as an example:

What was the title of the last book you finished and what was it about?

-r/AskWomenOver30 participant and moderator of r/AskMenOver30

The submitter is asking a general question and one that's relatively easy for people to respond. The ease of the question makes this easy for many people to simply engage through comments.

Conversation starters were common even in other subreddits like r/WomensSoccer, r/Fzero, and even r/UpliftingNews. In the lead-up to a new game release, participants in r/FZero asked what they hoped for in the new game. In r/UpliftingNews, there was a short-lived weekly series where a submission asked the members about positive things from their lives.

Events were also sometimes used to start conversations, as is somewhat common on reddit (e.g. Leavitt and Robinson, 2017a; O'Hallarn, Slavich, and Emmons, 2023; Unkel and Kümpel, 2020). In r/WomensSoccer, some members would create submissions for game matches and events. These came in several forms,

with some being simply recaps of the event (possibly updated in real-time like other reddit live threads) and others had a lot of resources including information, links to information, and links to other sites to view the match. Five of the 10 most commented submissions in the subreddit were match threads, suggesting that they are valuable within the subreddit.

The act of trying to start a conversation is distinct from the questions covered previously by way of intent. Any question may engage potential users, but the conversation starters were generally not questions the author needed an answer and they were broad enough that many people could easily respond. If someone asks about the last fast food restaurant they ate at, they probably aren't looking for recommendations.

Like all submissions, some of the examples I analyzed were not always successful at starting conversations, but some of that may be in part due to a lack of visibility of the content (e.g. E. Gilbert, 2013).

4.3.3 FOSTERING

Another theme that came from the qualitative analysis were sustained efforts intended to engage people in a community, likely with a goal of building community and placemaking. These were behaviors that were a time investment into the community. I refer to this group of behaviors as *fostering* as they attempt to establish, grow, or maintain the subreddit. Within this dataset, these were visible as a form of content curation, which has generally been seen as valuable for online communities (B. Butler et al., 2002), and may encourage people to participate. These behaviors include attempts to provide content, establish norms, and encourage participation.

I first encountered this during the qualitative analysis, where I recognized accounts that created multiple submissions. Follow-up analyses across the sample found that a small portion of accounts were highly involved in creating a lot of content. Some of fostering accounts I found were creators or moderators, but not all. As a set of behaviors, this is interesting because prior work has noted that people in roles of power, like moderators, spend time engaging with the community (Seering and Kairam, 2023; Wohn, Jough, et al., 2019). My analyses here focus on visible work that is the trace content, but invisible work is often involved in curating or writing that content. Some forms of fostering, such as moderating

4 Study 1 - Behaviors during the Early Life of Online Communities

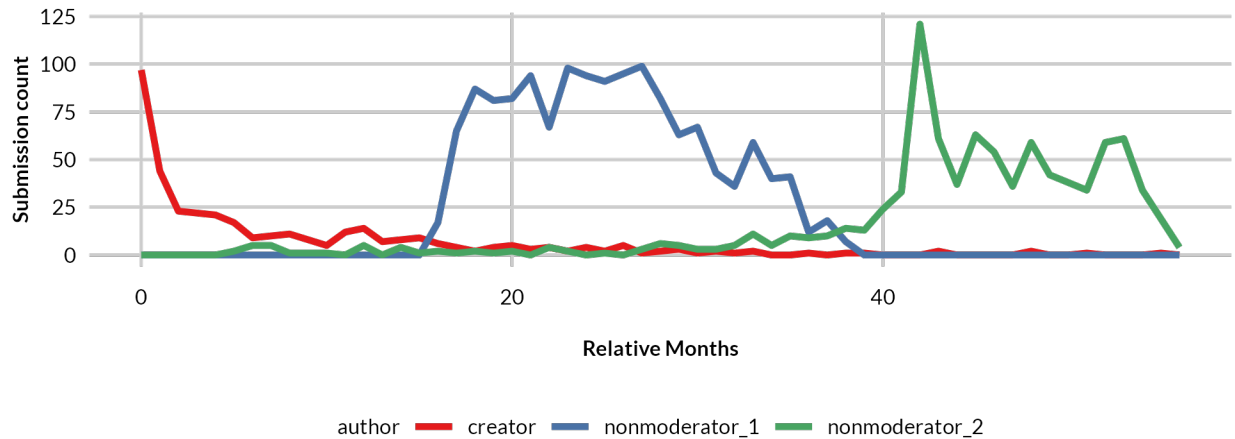


Figure 4.5: A line graph showing the number of created submissions in r/UpliftingNews by its creator and two active non-moderators across 30-day periods. All three submit a lot of content over time, though their highest periods of activity do not overlap. The creators content starts high and quickly declines after the first 30-day period.

content and subreddit configuration, would not be visible in this data, but could be considered fostering since these can also be sustained efforts to maintain or build the community (S. A. Gilbert, 2020; Lo, 2018; Seering and Kairam, 2023).

4.3.3.1 SUSTAINED CONTENT CREATION

The first visible form of fostering was the sustained creation of submissions within the subreddit. These submissions can provide a regular source of content for people to consume, but they may also act as exemplars to help people understand the value, purpose, and norms of a community. I found this behavior across five of the subreddits, including the four most active subreddits.

The largest subreddit within this sample, r/UpliftingNews, had a very active creator who continued submitting a lot of content for a very long time. In the first 28 days, the creator submitted 91 submissions which made up 51% of all submissions in the subreddit. The second top submitter in this time created only 4 submissions. Over time, the creator slowly reduced his content creation, but other authors began submitting articles in high quantity (Figure 4.5).

These submissions only hint at the amount of time people are investing in the community over time. However, each author could spend a significant amount of time finding relevant content to share within the subreddit. The creator of r/UpliftingNews noted this in a submission in that subreddit requesting the help of other users to find the content:

STATE OF THE SUBREDDIT ADDRESS:

...Sadly, searching for good, quality, uplifting stories is harder then one might expect. Simply search for 'good Samaritan' in Google news, and you'll notice half of the articles will be about somebody getting mugged, taken advantage of, or even killed for their actions. I've had some success searching news for certain key words such as 'kindness,' 'community rallied,' 'heartwarming' etc...

-creator of r/UpliftingNews

The sustained content creation can be useful for attracting new people and engaging existing ones because of reddit's front page algorithms, which aggregate recent content from across a set of subreddits (*hottest page* 2005; *reddit is so normal* 2008). For growing subreddits, the platform also has special page, r/popular, which contains popular submissions across all subreddits, even if not subscribed, which acts as a discovery mechanism for subreddits, but only those with recent content. In the case of r/UpliftingNews, it experienced rapid growth faster than other subreddits.

In r/AskWomenOver30, this work falls on people other than the creator. Two specific users were responsible for a majority of the subreddit's submissions, and these are sustained over the lifespan within the dataset. One was a moderator of the subreddit and the other was a moderator of the sister subreddit (r/AskMenOver30)⁵. In figure 4.6, these people continued submitting submissions through the end of

⁵it's possible the second account was a moderator in r/AskWomenOver30 at some point, but I could not find evidence to support this.

4 Study 1 - Behaviors during the Early Life of Online Communities

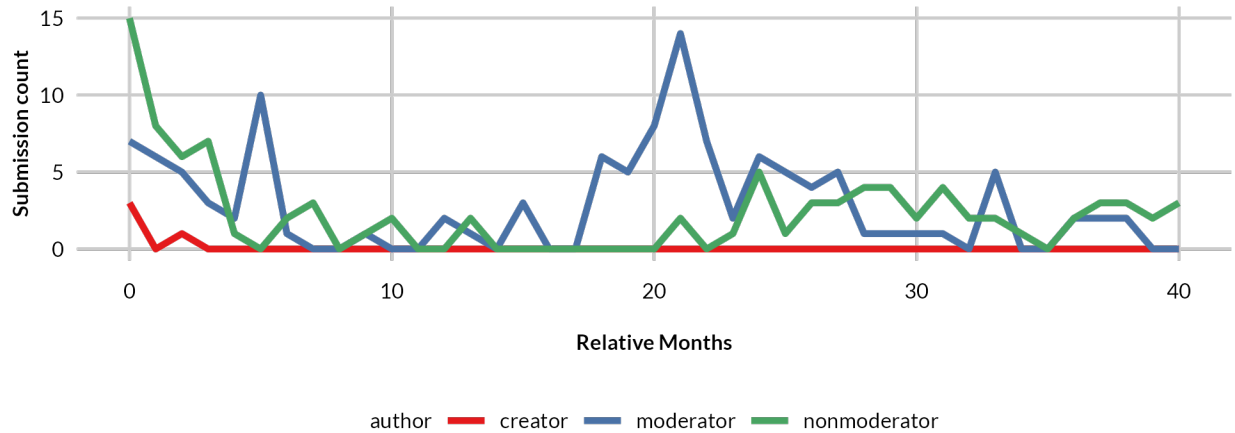


Figure 4.6: A line graph showing the number of created submissions in r/AskWomenOver30 by its creator, a moderator, and a moderator of a related subreddit. The creator creates only 3 total submissions at the beginning of the subreddit, but the other two create many submissions through the end of 2016.

2016 (the end of the dataset). Many of these submissions were conversation starting questions, with forms like 'How do you feel about X?'.

In the several subreddits, the sustained content creation may have been important for attracting new members who eventually become the source the vast majority of new submissions on the subreddit. However, the sustained creation of submissions did not always increase in active members in r/EarthScience. The creator generated the bulk of the submissions in the subreddit. In the first 28 days of the subreddit, there were a total of 238 submissions, and the creator of the subreddit created 201 of those submissions (~84% of the total, and ~7.2 submissions per day). This person continued submitting content for over 2 years until they were seemingly banned, at which point they continued under a slightly modified username. The person submits the majority of content for nearly three years (approximately 36 30-day windows, see figure) and 52% of all content in the subreddit through the end of 2016.

Additionally, the fostering behavior was observed in accounts that were not creators or moderators. Two of the most active users in r/UpliftingNews submitted a lot of content, though over slightly different periods of time. This can be seen in figure 4.5, with two non-creator and non-moderator accounts contributing over 100 submissions in their peak 30-day windows. They created 1438 and 916 submissions

respectively, both more than the creator. As far as I can reconstruct, neither were ever moderators within the subreddit.

4.3.3.1.1 SEEDING *Seeding* is a related behavior focused on creating an initial set of submissions for users, thus making it a subset of sustained creation of submissions. This typically happened near creation of the subreddit, but in one case when a subreddit was taken over. The term seeding has been around for decades (e.g. Preece, 2000), and is also used by redditors and the platform. In total, I observed seeding behavior in 5 total subreddits, and in many situations the work was done by creators or moderators.

Seeded content was often started before other users had participated in the subreddit, and this was sometimes soon after creating the subreddit. For example, in r/EarthScience, the creator created the subreddit at 2:26 (UTC) on March 8th, 2011. Five minutes later, the creator submitted the first submission. The creator continued submitting and within the first 24 hours had submitted over 27 pieces of content. (The actual count may be higher, but 5 of the 34 submissions in the first 24 hours are from deleted authors. One includes the same exact link that the creator submitted minutes later, suggesting that these may have been removed for some reason but these could not be resolved).

A similar situation occurred in r/UpliftingNews, which was created on May 16, 2012. Hours after the subreddit was created, the creator submitted a single submission link to a news article. Two days later they submit 7 additional news articles and an additional 4 articles the next day. In the first week, they created 24 (77%) of the 31 total submissions, which provided a decent amount of seeded content for anyone visiting the subreddit. Their initial activity can be viewed in figure 4.5, which shows a significant amount of content posted by the creator, before slowly declining over time. This initial burst of content, visible near this peak is the initial content which defines a channel and can potentially encourage others to participate.

I also observed two instances of delayed seeding, such that the content wasn't created at the start of the subreddit. In r/AnimalsFailing an account seems to have created some fostering content before deleting that content. The subreddit saw some activity, and then three different accounts created seeding content which helps kick the subreddit off. This highlights two key points: first the seeding is happening

4 Study 1 - Behaviors during the Early Life of Online Communities

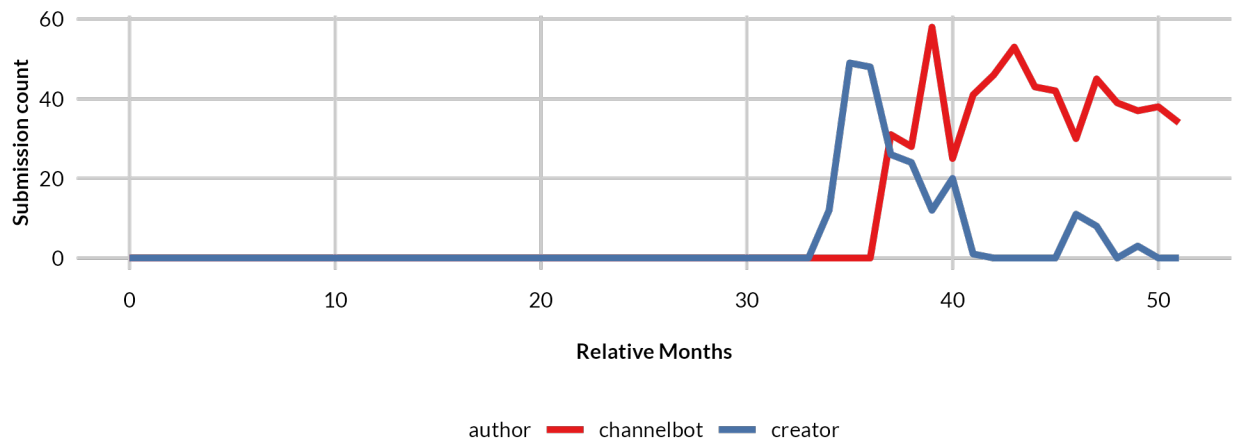


Figure 4.7: In r/LinusTechTips, the original creator never created any submissions, but in 2015, the subreddit was taken over by a new user, who began posting a lot of links to the Linus Tech Tips YouTube channel. He later leveraged a bot, u/ChannelBot, to largely automate submitting videos.

in a largely inactive subreddit that's relatively new, and second, the seeding authors do not need to be moderators or creators.

An example of both of these points occurred in r/LinusTechTips. There, the creator appears to have done nothing aside from creating the subreddit. A few people enter and make submissions, some lamenting the lack of activity, but the creator never appears to have posted or commented in the subreddit.

The subreddit sat nearly dormant for nearly 3 years until a new account took control of the subreddit and became a moderator⁶. The new moderator made a post announcing the his takeover and began posting links to video content on YouTube from the Linus Tech Tips channel. They posted at least 2 videos in the 24 hours after that announcement (3 are deleted or removed but follow a similar pattern) and an additional one the following day. They continue creating submissions almost daily for over a month. The subreddit has a low level of activity prior to his takeover. This provides a form of quasi-seeding behavior, because his submissions do help set a tone for the subreddit, they are also different from what others

⁶It is not clear how the account became a moderator. It is possible they asked the original creator to take control, or they may have asked the platform and an administrator made them one.

continue posting. Additionally, the subreddit was already with 11 submissions in the 30 days prior to that message.

An important question about seeding is when does seeding end and normal fostering begin? I have defined seeding here as providing early content before a community is fully established and few if any participants are active. The latter part of this definition provides for the case of *LinusTechTips* where the subreddit is created by one user but taken over by a new user sometime later. In other subreddits, the question remains about a cut-off point between seeding behavior and just continued fostering.

One commonality visible across several subreddits are typical bursts or spikes in activity, followed by a decline. This is visible in *r/EarthScience*, *r/UpliftingNews* (Figure 4.5), and to a lesser extent *r/AskWomenOver30* (Figure 4.6, though all three cases are slightly different. In *r/EarthScience*, the amount of initial content for the first 30 days is large, with 211 submissions created in the first 30 days is about 7 submissions a day – more than any other subreddit.

4.3.3.2 TYPES OF FOSTERING CONTENT

I also explored the submissions that I considered fostering by comparing the fostered content to the codes from the qualitative analysis. Many of these were content that was simply topical or specific to the subreddit (e.g. pictures of animals failing in *r/AnimalsFailing*). However, there were a few behaviors and forms of content that were less common, but still intended to engage and build community. These include content intended to start conversations, meta-discussions about the subreddit, and sharing resources with the community.

I previously described conversation starters as generally questions where the submitter was trying to engage other participants instead of looking for a genuine answer. This occurred in the fostering content for a few subreddits like *r/AskWomenOver30* and *r/Fzero*.

Meta discussions were a rarer form of submission, but these included discussions about the subreddit. These included announcements from creators and moderators, such as requests for help finding content

in r/UpliftingNews. They also included requests and questions from normal participants about subreddit features like header images or flair for favorite teams in r/WomensSoccer.

4.3.4 PROMOTION / ADVERTISING

The qualitative analysis found a number of ways in which people promoted subreddits. One of the challenges with starting an online community is encouraging people to visit and contribute to the space (Kraut and P. Resnick, 2012). However, people must first be made aware that a community exists before they can visit and participate. The platform offers a number of affordances that can make promotion easier, even in situations where the author does not intend to explicitly promote a subreddit. The potential importance of promotion in growing a community makes this interesting enough to include in the framework.

Reddit recommends promotion as a means of raising awareness and growing a subreddit (Reddit, 2023a), but also has strong norms against spam (Reddit, 2025). Some forms of promotion, like cross-posting, are more acceptable in multi-community spaces like reddit and can enable discovery of other subreddits in the spaces they subscribe⁷.

Promotion behaviors appeared within initial qualitative analysis and the larger dataset in three primary methods: content posted from other subreddits into the sampled subreddits (e.g. cross-posting), content linking or mentioning other subreddits, and people explicitly promoting a subreddit. Each of these involve different mechanisms for taking advantage of various affordances, while also potentially avoiding some norm-violations around spam. I found four types of promotions across the data: links, mentions, cross-posts, and explicit advertisements in submissions.

In this section, I blend together both the qualitative analysis results with additional quantitative analyses looking deeper at specific promotion behaviors within the sampled subreddits. This allows me to look at the entire timeline of subreddit data beyond the qualitatively coded range of 400 submissions to better understand the prevalence of these behaviors. I analyze the full range of each subreddits data by

⁷There is a different form of cross-posting where people can share content across platforms, such as a news article, image, or video. For example, someone might upload the same video to TikTok and YouTube to reach a broader audience.

identifying mentions and links using regular expressions across the metadata of each submission and comment.

4.3.4.1 SUBREDDIT MENTIONS AND LINKS

The majority of promotions were mentions and links found in the text fields of comments and submissions. I found only a few explicit forms of promotions within these. Many were used for sharing or to react to content, and could act as a promotion because of platform affordances and their ability to raise awareness of the subreddit.

Mentions reference only a subreddit (e.g. r/Eskrima), while links reference a specific resource of a subreddit, like a submission or comment (r/Eskrima/comments/...). I identified a total of 9787 submissions or comments containing 10734 unique mentions or links. I use unique here as some content contained multiple mentions to the same subreddit or links to the same destination. This means that some comments and submissions contained multiple links or mentions to different subreddits or specific content.

The majority of all promotions in this sample were made in comments (98%). The majority of promotions were also mentions (74%). Unsurprisingly, the larger subreddits tended to have more links and mentions. r/UpliftingNews had the majority of content referencing other subs (90%), while r/Eskrima had only 28 (less than 1%).

Not all of the links and mentions were valid links. I found 432 pieces of content that point to subreddits that did not exist at the time the promotion was made. The majority of these were mentions (99%). A total of 317 mentions pointed to subreddits that did not exist. A total of 115 of these were promotions occurred before the subreddit existed (27%). The remaining 73% were to subreddits that did not exist by 2017⁸. Some of the non-existent promotions were apparent misspellings (e.g. r/nottheoinion), but most were used as tags or exclamations as a reaction to the content. This includes 69 mentions that were invalid, having a name longer than that allowed for subreddits (eg. r/PuttingAReallyDepressingSpinOnUpliftingNews).

⁸There is one subreddit that appears to have existed, but reddit seems to have renamed the subreddit.

4 Study 1 - Behaviors during the Early Life of Online Communities

Looking at the most mentioned or linked subreddits from each of our sample subreddits, we can find that many of them appear topically related. The most referenced subreddit in r/LinusTechTips was r/buildapc, a subreddit for building PC computers. The top referenced subreddit for r/EarthScience is r/askscience. This was true across all of the subreddits, and suggests that these can be used by people to find similar subreddits. However, not all are completely related. Two of the most-referenced subreddits were from bots in r/UpliftingNews. The top reference was r/totesmessenger, which is a bot that comments in submissions that were linked to from other subreddits. The mention of r/totesmessenger provides a link to the bot's home subreddit, for people who want to visit and learn more about the bot.

To better understand the context of these links and mentions, I conducted a thematic analysis on a random sample of 500 comments containing mentions and links from these subreddits. The links were most often used as resources or pointers to other content on reddit.

Mentions had several different forms of uses. One was using it as an explicit form of promotion, by pointing someone to the subreddit. For example, one person commented on a story in r/UpliftingNews and suggested the submitter also submit it to r/aww, suggesting the content would be welcome there. Sometimes the comments only contained a mention of a similar subreddit.

Mentions were also used as a reaction or a tag to some content. One commenter mentioned r/hailcorporate on a submission in r/UpliftingNews, suggesting that original submitter was acting like an advertiser for a corporation (the topic of r/hailcorporate). This type of reacting/tagging behavior was similar to the aforementioned invalid mentions, where people used them as reactions or tags

The practice of using links and especially mentions is not rare. The presence of the reaction/tags, especially to non-existent subreddits suggests that some of these mentions are not intentional promotions of subreddits, even if they can act like it. In some cases, these can lead to the mentioned subreddits being created (Krohn and Weninger, 2022).

4.3.4.2 ADVERTISING SUBMISSIONS

Across the 8 subreddits, I found 17 instances of explicit promotion within submissions. These submissions explicitly tried to attract members using language like “Tired of depressing news? Check out r/UpliftingNews.” Though this number is low, it’s somewhat surprising given how small many of the sampled subreddits are, and advertising in larger subreddits would reach larger audiences. I found two different versions of explicit promotions. The first included submitting a link directly to the subreddit, while the second involved a text post that simply mentioned the other subreddit. In both cases, other users can simply follow hyperlinks to the other subreddit.

As an example of the submitted link style, an author submitted a link to r/NWSL to r/WomensSoccer with the title “Want to follow the new NWSL? We’re trying to start up a subreddit for it here.” The title is used to provide background about how r/NWSL is a new subreddit they’re starting, though without describing what the NWSL is. It is also worth noting that topically, the NWSL is generally a slightly more niche subreddit that covers only the US and Canada whereas r/WomensSoccer had theoretical broader coverage of all women’s soccer. Thus the author is leveraging an established subreddit that covers a broader range to advertise a related, but more focused subreddit.

The second version of explicit advertising via submissions occurred through simply mentioning the subreddit in the submission’s title or in text in the submission itself. This was more common than submitting links directly to the subreddit. For example, before the account took over r/LinusTechTips, the subreddit was largely inactive. Someone had created a subreddit for the same purpose named r/Linus_Tech_Tips and advertised it in r/LinusTechTips.

The author tried to promote the newly created subreddit by describing some of the work they’re putting into the new subreddit. This was likely to make it sound like the advertised subreddit would be actively maintained, which wasn’t the current case for r/LinusTechTips. In contrast to the previous example of advertising, where the submission was to a broader channel with existing subscribers, this submission is trying to create a more active space and is advertising to people who subscribe to the

largely dormant r/LinusTechTips. This was a unique instance because the subreddit was dormant for an extended period of time. The other examples of advertising were just for related or more niche subreddits.

4.3.4.3 CROSS-POSTING

Cross-posting is an umbrella term for a number of methods of submitting the same content to multiple subreddits. It was the most common form of promotion within the qualitative sample with a total of 79 instances across all 8 analyzed subreddits. The name is likely borrowed from Usenet, an early multi-community platform which also enabled users to post to multiple newsgroups (B. S. Butler and X. Wang, 2012; Whittaker, Terveen, et al., 1998). It's history on reddit dates back to at least early 2006 when they enabled

affordances to cross post between the few subreddits that existed then, though may have been manually performed earlier (shuffman56, 2006a). Since that time, it has also evolved with the platform. People on reddit value cross-posting for sharing related content across different audiences, which can enable different conversations about the shared content (TeBlunthuis, Kiene, et al., 2022).

For example, a news story about Seattle may be relevant to the r/news subreddit and the r/Seattle subreddit. The content can provide value to members of both subreddits in a few ways. First, not everyone interested in the story may subscribe to both subreddits or may not even be aware of both subreddits. Even for those who subscribe to both subreddits, the platforms algorithm may make only one visible to a specific user. Lastly, the discussions in each subreddit may differ in ways that provide different forms of valuable information .

There are a few different ways in which people can cross-post content, but in each way the name of the primary subreddit is less visible than the content itself. So depending on the specific placement of the

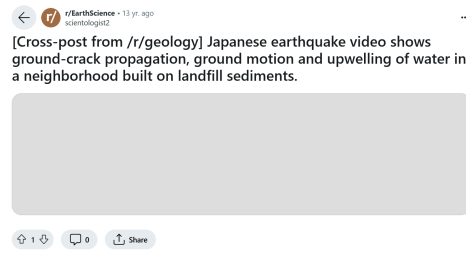


Figure 4.8: An example of a cross-post from another subreddit into r/EarthScience. This one is of a video that no longer exists, but the title still illustrates some of the manual involved in labeling some content found on other subreddits as cross-posts.

other subreddit name, some users may not notice the subreddit where the content was shared from and choose to visit.

4.3.4.3.1 COPIED LINKS FROM OTHER SUBREDDITS The first involves copying a URL that was submitted to another subreddit. In these cases, sometimes authors would amend the title of the submission to note that it's a cross-post and attribute the source subreddit. The attribution in these posts serve to raise awareness of other subreddits and potentially attract new users.

For example, in the submission below from r/EarthScience, we can see how the submitter here copied a link to video from cnn.com that was originally posted on r/geology.

In the post above, the creator the subreddit r/scientologist2, prepended the article title to identify the article as a cross-post and to point to the source of the content. Although, it is impossible to tell from the data itself at this point, this was likely a manual process which involved finding the original content on reddit and manually copying-and-pasting the URL.

Cross-post identifying text	count
xpost, x post, etc	778
other	103
from	33
crosspost, cross-post, cross post	29
in/on subreddit	29
•	19
repost, re-post	13
(subreddit) or [subreddit]	10

Table 4.6: This table shows a list of some of the most common ways people manually labeled cross-posts within the sample.

For manual cross-posts, there was no adopted standard of attributing the source subreddit or author when using this method. Typically this was done through prepended titles or appended titles, but with many different sets of words and notation. I found a total of 1014 manual cross-posts across all eight subreddits (Table 4.6), and the vast majority (77%) were attributed using some form of xpost, x-post, Xpost, etc. Other forms of attribution included “from <subreddit>”, crosspost or cross-post, repost/re-post, etc. It is also possible there were many other unattributed cross-posts, since there was no standard or norms enforced on the practice.

4.3.4.3.2 SHARING REDDIT LINKS TO OTHER SUBREDDITS Another method of cross-posting was submitting the URL to an submission in another subreddit. A similar method would later become the standard method of cross-posting when reddit built later built sharing affordances. On the platform today, these links appear as an embedded submission that showcases the original author and subreddit the content appeared in.

There are a total of 411 links to reddit submitted within the sampled subreddits through the end of 2016 (about 0.4% of submissions). Of these, 380 are links to subreddits, and 358 are links to submissions on other subreddits. People submitted a number of reddit links to other locations on reddit like user profiles, subreddits, and specific pages, none of which would be be promotion of a subreddit and thus were not analyzed further.

The source subreddits for these links are more interesting, as most appear to be topically related. This relationship can be visualized with a network graph, with the vertices being subreddits and the edges being the number of cross-posted links from the source subreddit to the sampled subreddit (see Figure 4.9). To make the graph readable, I removed any source subreddits with fewer than two cross-posted links. Several of the sampled subreddits like r/AskWomenOver30, r/Eskrima, and r/Fzero do not appear because of the cross-posted links they call came from different subreddits. In the graph, you can see that the 3 sources for r/EarthScience are mostly topically related, except perhaps r/HomeworkHelp. The same applies to r/WomensSoccer, which received cross-posts from other soccer-related subreddits. Incidentally, r/WomensSoccer received the highest number of cross-posts from a single subreddit, r/soccer (24 total cross-posts).

The majority of the cross-posted submissions were created by people who did so only once within the sample (206 of 358). Of the 358 total cross-posted links to other reddit content, 70 of those have had the author deleted (some of which may be removed from the subreddit). Only 17 authors had 2 or more cross-posts. Three of the top four cross-posting authors were moderators or creators. This suggests that many cross-posts may be more focused on content sharing than promotion, though both can still be achieved.

4 Study 1 - Behaviors during the Early Life of Online Communities

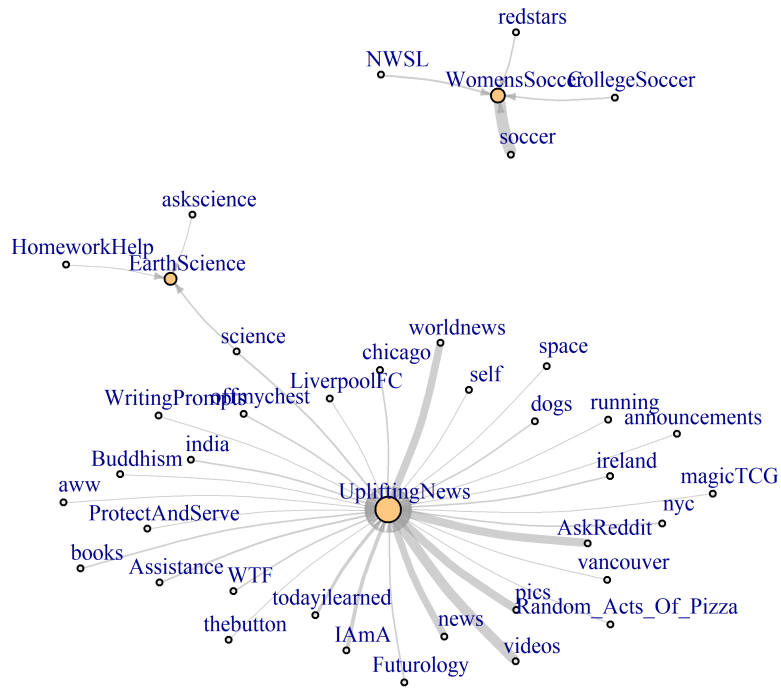


Figure 4.9: A network graph showing the sources of the cross-posted reddit links into 3 of the selected subreddits. Of particular note here is that for both r/EarthScience and r/WomensSoccer, the source subreddits are topically related. For r/Uplifting News, some of the source subreddits are related, but many tend to be more general (e.g. r/pics, r/videos, r/worldnews). The selected subreddits are shown as pale orange circles. Subreddits which were the source of these links are shown as smaller gray circles. The line/edge size shows the relative number of cross-posted articles from the source subreddit into sampled. The size of the circle is relative to the number of cross-posts they received.

4.3.4.4 SUMMARY OF PROMOTION

I have included promotion as a perspective this framework to highlight the work users invest time into, which may aid in the discovery and growth of subreddits. On the surface, promotional work can generate awareness of communities, especially within multi-community spaces. This can drive new people to visit the community. However, promotional work can do more than simply raise awareness. It can also be a form of fostering behavior, boundary shaping, and even placemaking.

The promotional behavior on reddit works through people leveraging specific affordances of the platform. Some of these affordances have been built around existing behavior (such as subreddit mentions being turned to links) as the platform co-evolved with its members. Simple text mentions alone may still be useful, the affordance of the links helps make this easy to quickly view subreddits, where people can decide to subscribe or not. Even when people use mentions as responses, and even jokes, they can still quickly direct people to subreddits. This has even led to the creation of subreddits that did not exist at the time of the mention (Krohn and Weninger, [2022](#)).

The act of promoting through advertising, cross-posting, links, and mentions can also be seen as a form placemaking. People are using these links to either delineate shared boundaries with other subreddits for similarity (or lack thereof). When a person shares content from one subreddit into another, with the exception of undirected spam, this usually has a potential target audience in mind and the person hopes the audience overlaps with the receiving subreddit. Moderators also have the power to edit some aspects of the subreddit's page, to include links directly to other similar subreddits, which is a specific form of customization.

Promotion is not welcomed across all subreddits. Reddit's own policies suggest that promotion needs to be done carefully so as not to violate the platforms own spam policies. As with everything easy on the internet, it has been misused by many people trying to grow their subreddits. This has led to the development of norms and policies within each subreddit to also disallow some promotional behavior like cross-posting. This is another form of placemaking, where these policies are developed to protect the space and integrity of the subreddit. The platform has also implemented affordances to disable the UI

4 Study 1 - Behaviors during the Early Life of Online Communities

affordances to easily cross-post content to help support this form of behavior. This can be particularly important in some subreddits, where people can share sensitive information or images, and do not want them shared more broadly.

In the case of advertising, the examples I analyzed suggest the authors often attempted to highlight how they intended the advertised subreddit to be different from the one they were submitting to. In the example from r/LinusTechTips, the submitter promises that it will be a more active subreddit (as the current was currently inactive). In some of the other examples, they attempted to carve out more specific niches around even more focused communities. This serves not only to grow awareness of a community and its intent, but also to help define this as a place with specific boundaries and norms. The boundaries may overlap and so the advertising may possibly reach some of its intended audience, but by suggesting it to people in the target subreddit the authors are extending an invite for people interested in the advertised subreddit while also declaring their similarity.

In a sense, this is not unlike community outreach programs which attempt to make their presence known through door-to-door outreach or stands at specific events. This is common with many traditional organizations and clubs, which can be regularly recruiting new members. Many forms of political action and social movements also employ similar strategies, which can attempt to communicate their goals or visions. With this metaphor, it is perhaps easy to understand the connection to placemaking. These organizations leverage other spaces, like shared spaces, to promote and further their clubs or communities. The organizations may choose the specific events or locations where they are most likely to reach a new audience, which is not unlike choosing a similar subreddit.

Promoting of online communities needs more study. Some prior research has noted its importance as a form of work for attracting new participants (B. Butler et al., [2002](#)). Some work exists on cross-posting (B. S. Butler and X. Wang, [2012](#)) from Usenet. On reddit, some recent work suggests that comment mentions can increase engagement within those subreddits and also led to the creation of new subreddits (Krohn and Weninger, [2022](#)).

4.3.5 INTERACTION

Subreddits also differed in how the people interacted with other people. There are two main forms of interaction on reddit: voting and commenting. Many people come to these online communities for some form of interaction, so the patterns across subreddits is important to understand. I look at the rates of interaction in each subreddit, their effect on first-time submitters, and the depth of conversations in each subreddit.

Comments occur on submissions and other comments, and are often a reply or reaction. Both submissions and comments can be voted on. Votes are aggregated into a score, and they affect the visibility of the submissions and comments. The platform's ranking algorithm uses the score to surface content that is the most highly-engaged by the community, and the platform can hide or reduce the visibility of poorly received content. This works as a form of collaborative filtering similar to platforms like Slashdot and Stack Overflow. These votes are also translated into a score for the user, referred to as karma, and can be leveraged by others to make judgments about other users and how the community values their contributions. In this way, the votes may also act as a form of lightweight communication or feedback from the community to the author. However, the anonymity of votes and the lack of additional information make them unlikely forms of interaction necessary for building or maintaining individual person to person relationships. Instead, they may be interpreted as a value indicator from the community-at-large (S.A. Gilbert, 2018), though this can be a biased indicator.

Whether someone is asking questions, sharing content, discussing a hobby, getting support, or playing a game, social interactions are an important motivator for people continuing to engage within the community (Burke, Kraut, and Joyce, 2010; S.A. Gilbert, 2018; Joyce and Kraut, 2006; C. Lampe and E. Johnston, 2005; Ridings and Gefen, 2004). Receiving high numbers of upvotes on reddit may be as support from the community (S.A. Gilbert, 2018) and can be viewed as a form of value to the community. Receiving interactions may be a motivation for deciding to post in some subreddits (TeBlunthuis, Kiene, et al., 2022). Repeated interaction can help people feel a sense of connection to the community (Fiore, Tiernan, and M.A. Smith, 2002), especially when those interactions include receiving support (Welbourne, A. L. Blanchard, and

4 Study 1 - Behaviors during the Early Life of Online Communities

Subreddit	Total Submissions	% with comment	% with vote	% with any interaction
AnimalsFailing	454.00	72.91%	71.15%	85.46%
Fzero	599.00	65.11%	66.78%	82.30%
EarthScience	1876.00	21.80%	67.80%	70.47%
Eskrima	452.00	75.00%	80.31%	88.05%
LinusTechTips	1635.00	60.18%	88.69%	91.56%
WomensSoccer	3452.00	31.58%	82.39%	83.40%
AskWomenOver30	2222.00	87.80%	88.48%	90.64%
UpliftingNews	98294.00	30.45%	32.38%	43.02%

Table 4.7: Descriptive statistics for the sampled subreddits showing the percentage of submissions with comments, votes, or either a comment or vote (score $\neq$ 1).

Wadsworth, 2013), and can be the basis for developing new friendships or strengthening existing friendships (Haythornthwaite, 2002; Ploderer, Howard, and Thomas, 2008).

During the qualitative analysis, a number of questions emerged about how people were coming to subreddits to interact and how the outcomes affected their participation. In r/Eskrima and r/AskWomenOver30, many people were approaching the subreddits to get answers to questions, get advice, and to start conversations. Some attempts were not successful in attracting a single reply (comment). Given the importance of responses and interaction, I explored a number of related questions around the frequency of interaction, reply rate on submissions, and the depth of conversations.

Empirically, there are three points of interest from my analyses that I want to highlight. First, subreddits differed how many submissions received interactions. Second, while some subreddits do experience signs of growth, about half of the sample had sustained periods with low numbers of interaction. Lastly, discussions tended to be shallow – meaning most reply-chains are a single comment deep. If the latter holds true across all of subreddits, it implies that in-depth conversation may be rare across subreddits. The depth, or lack thereof, may impact whether people feel a sense of community, but this is a question for future research.

4.3.6 SUBMISSION INTERACTIONS

Subreddits differed in how people interacted with submissions. Voting was generally more common than commenting, but the rates varied significantly between subreddits. Some of a few subreddits had

4 Study 1 - Behaviors during the Early Life of Online Communities

very high rates of commenting compared to others. The largest subreddit, r/UpliftingNews, had the lowest rates of interaction of all the subreddits. One possibility is that there is too much content posted, and the platform design doesn't show enough of that content to subscribers. Part of this might be the importance of interaction for ranking content, meaning the platform would show 10 minute old post with 1000 upvotes instead of a new post.

In seven of the eight subreddits, a higher percentage of submissions received votes than received comments (Table 4.7). The exception was r/AnimalsFailing, though the difference was less than a percent. Five of the subreddits had comments on more than half of the submissions. The exceptions were r/EarthScience, r/WomensSoccer, and r/UpliftingNews. The former two are surprising given that the rate of votes was higher. These may reflect subreddits' collective members being more interested in news than discussions.

The largest subreddit, r/UpliftingNews, had the lowest rate of interaction on submissions, which is due to having the lowest rate of voting. This appears to be a product of having become a default subreddit, where the percentage of interacted submissions drops significantly (Figure 4.10). Default subreddits were subreddits that new accounts were automatically subscribed to when they joined the platform, which created steady growth as the platform grew⁹. This was a feature on reddit for a number of years Prior to becoming a default subreddit the rate of interactions was much higher with 78% of submissions receiving a vote and 50% receiving comments.

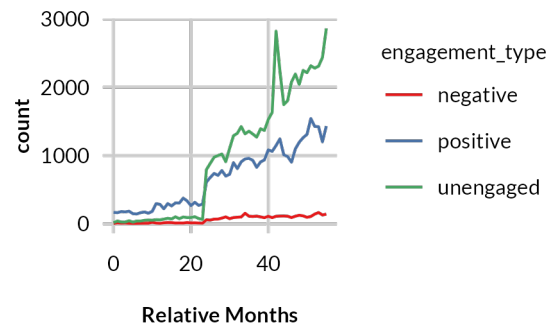


Figure 4.10: A line graph of r/UpliftingNews showing the effect of defaulting both on total submissions and engagement aggregated over 30-day periods ('months') since its creation. Here, there is a large spike in submissions with a score of 1 that have no comments, which is the default score. The sharp decline at the end of the graph is the result of the end of the dataset. That window has only 10 days and included Christmas 2016, thus having fewer submissions.

⁹This was a feature for a few years on Reddit, through about 2017 (u/simbawulf, 2017).

4 Study 1 - Behaviors during the Early Life of Online Communities

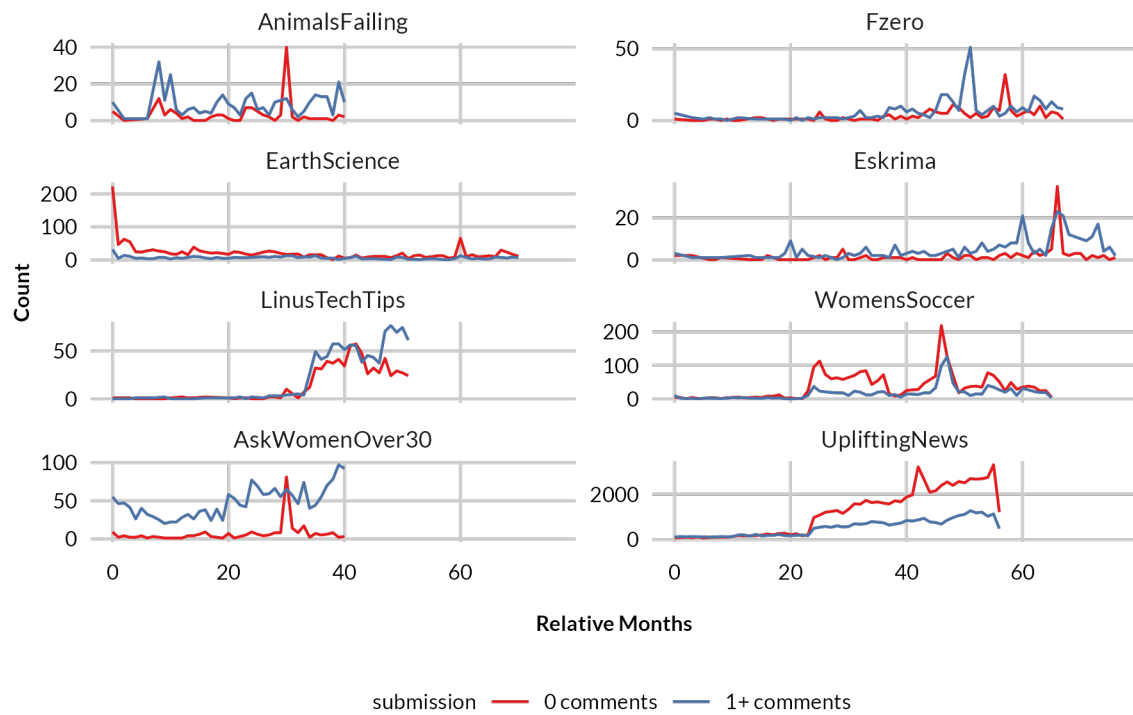


Figure 4.11: A line graph showing the submissions that received replies or did not. Of note is a large spike in submissions without replies later in the life of many of the subreddits, this is a spike of spam. However, the largest spike in r/WomensSoccer is actually the 2015 women's world cup.

4 Study 1 - Behaviors during the Early Life of Online Communities

The decrease in interactions may be due to *underprovision* (E. Gilbert, 2013), where the amount of submissions exceeds the ability of the subreddit to process and vote on this content. Some have argued this may be partly due to cognitive load (Jones, Ravid, and Rafaeli, 2004), though others have found no evidence to support this (Arguello et al., 2006). E. Gilbert (2013) had a hypothesis that the platform itself limit which and how many submissions received views or attention for other people. His design allowed him to look at content that had been submitted multiple times, where the content did not initially receive many votes, but later did.

Given that past research has found replies (comments) to be an important factor in continued participation, I explored the rate of commenting on submissions a bit deeper. Subreddits like r/AskWomenOver30 had more submissions with comments than without and this is true throughout its lifespan. This is also appeared to be true in r/AnimalsFailing, excluding an early period of no activity. Other subreddits have more mixed timelines. A number of subreddits exhibit similar numbers of submissions with and without comments for a number of 30-day windows after their creation. This is visible in r/Fzero, r/Eskrima, r/LinusTechTips, r/WomensSoccer, and r/UpliftingNews.

Some subreddits saw later increases in the number of submissions without comments. This occurred in r/WomensSoccer and r/UpliftingNews, where both saw large increases in the number of submissions received in these 30-day periods. This was especially true in r/UpliftingNews where there is a substantial increase around the 24th 30-day window, which is when it became a default subreddit. The percentage of submissions that received comments dropped from around 40% to about 30%. This is possibly a case of E. Gilbert (2013)'s underprovision, where too much content is submitted to the subreddit and the platform prioritizes recent, highly engaged content over new, unengaged content. The perception of receiving attention has been previously identified as a reason for choosing to participate in some subreddits (TeBlunthuis, Kiene, et al., 2022), though in the case of r/UpliftingNews, this does not limit it's growth.

When users come to a community looking for information or support, their continued participation in the community may depend on them getting a reply (Joyce and Kraut, 2006; C. Lampe and E. Johnston,

2005). The rate of people who continued participating in the subreddit after their first post was 47.8% (7091 of 14845) for people who received at least one comment on their submission compared to only 20.8% (5866 of 22348) for those who did not receive a comment. A chi-squared test here suggests that there is a relationship ($p < 0.001$; simulated with 2000 replications), though this relationship may not be causal.

If we look across the sampled subreddits, we can see the rates of continued participation varied across the subreddits, though receiving a reply led to a higher rate of continued participation across the subreddits. The lowest rate of continued participation past the first submission was in *r/AskWomenOver30*, where only 4 of the 120 (3.33%) first submissions continued participating beyond that post. The highest rate of continued participation for a submissions that received no replies in a subreddit was ~36% in *r/WomensSoccer*. The largest difference in percentages between continued participation was in *r/AskWomenOver30* where 3.33% continue with no comments and 83% continue when receiving comments.

4.3.6.1 COMMENT THREAD DEPTH

Comments can also be replies to other comments. These can form threads or chains that are similar to the turn-taking in a conversation. These threads or conversations can vary in length, with each reply adding a new level (these resemble trees in computer science language, thus each level would be the depth of the tree). A higher depth can resemble the back-and-forth of a deeper conversation. I found that the average depths and the distribution of depths differed across subreddits, with a possible positive relationship between subreddit activity and comment depth.

On platforms like reddit that provide “threaded” comments, users can reply to the submission or to other comments. Replies to the submission can be considered *top-level comments* and would have a depth of 1. A reply to top-level comment would have a depth of 2. As members reply to other comments, the depth would increase and be a potential measure of the back-and-forth nature of a conversation.

4 Study 1 - Behaviors during the Early Life of Online Communities

The top-level comments are effectively the start of a “thread” or branch where each reply chain descends deeper. Each submission and comment can have multiple replies, thus forming possible branches.

Previous research on Usenet has found a positive relationship between the retention of users and thread depth (Whittaker, Terveen, et al., 1998). They also found that that longer initial messages had lower thread depths, and increased cross-posting was also associated with increased the number of threads. Other Usenet research found variation in thread depth by the type of group, with Q&A style groups having shallower (shorter) threads and conversation groups having deeper (longer) (D. Fisher, M.A. Smith, and Welser, 2006). Moderation may also limit the thread depth in some communities (Whittaker, 1996).

In this analysis, I look at two ways of measuring comment depth. First, is the maximum depth of a thread which attempts to provide a look at depths of comments. This has been used in past research, but does not accurately convey the branching nature of threads where a single comment can lead to multiple separate conversations. There were 293225 total threads and a total of 880,819 comments across the eight sampled subreddits. This means that the majority of comments occur at a depth greater than 1. The second measure I used was average comment depth, which captures the issue with branches and breadth, but may not capture the depth of larger threads.

Top-level comments made up the majority of comments in 5 of the 8 subreddits (Figure 4.12). Note that these distributions are right-skewed, and are binned into the 5+ bin. The exceptions are r/Eskrima,

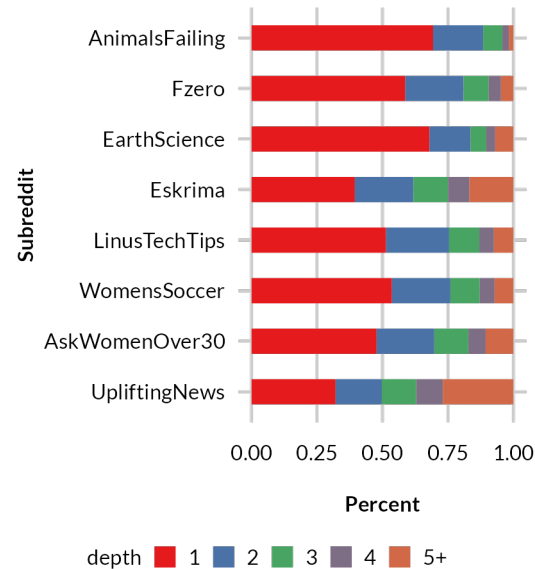


Figure 4.12: A graph comparing the percentage of comments at each depth, with 5+ containing all depths greater than or equal to 5. Five of the subreddits have more than 50% of top-level comments (depth = 1).

4 Study 1 - Behaviors during the Early Life of Online Communities

r/AskWomenOver30, and r/UpliftingNews. The distribution in r/Eskrima was interesting given it had a greater percentage of deep comments than other similarly sized subreddits. Both r/AskWomenOver30 and r/UpliftingNews were the largest of the subreddits, which may suggest that larger subreddits have more active users and thus more likely to have deeper conversations.

		Thread Depth %			
	subreddit	1	2	4+	5+
1	AnimalsFailing	82.12%	14.71%	3.18%	0.94%
2	Fzero	73.45%	19.27%	7.28%	3.24%
3	EarthScience	81.31%	14.01%	4.67%	1.73%
4	Eskrima	58.78%	23.77%	17.45%	11.01%
5	LinusTechTips	69.01%	21.29%	9.71%	5.43%
6	WomensSoccer	71.24%	19.41%	9.34%	4.54%
7	AskWomenOver30	71.14%	17.82%	11.04%	6.29%
8	UpliftingNews	79.79%	11.73%	8.48%	5.71%

Table 4.8: Percentages of thread depths across all threads in the subreddits. These numbers exclude submissions with no comments. Most of the subreddits have a depth of 1, meaning they are a top-level comment that no one replied to. The last two columns show threads with depths of 4 or greater and 5 or greater respectively. r/Eskrima has the highest percentage of threads with of 4 or greater.

The distribution of maximum thread depths was also skewed, and I present this in table 4.8. All of the subreddits see a majority of their threads are top-level only threads. The lowest rate was r/Eskrima, which also had the greatest percentages of long threads. The other subreddits ranged between 69% and 82%, which seems superficially high and suggests that most threads, but the average comment depth from figure 4.12 suggests that a small percentage of top-level comments may disproportionately attract a large number of comments.

There were 11057 total comment threads with depths greater than 5, including a counting thread which had a depth of 1165. Around 91% of these (10040) of these were in r/UpliftingNews and across 2,719 total submissions. Twenty-five submissions from r/UpliftingNews had depths of at least 20. The second largest subreddit, r/AskWomenOver30, had only 798 total deep threads. These were across 506 total submissions. Many of the smaller subreddits had roughly one of these deep threads per submission. Audience size likely plays a role here, but there may be additional factors. This provides evidence that some top-level comments may disproportionately attract more comments and the existence of a preferential

attachment process. Most subreddits had a higher percentage of threads with a depth of 1 with mostly percentages ~70% and 82%. However, r/Eskrima is again the outlier here with only 58%.

These results differ somewhat from those in D. Fisher, M. A. Smith, and Welser (2006), which found their technical Q&A groups had higher number of shorter threads relative to other types of newsgroups, and that support groups had the highest percentage of long threads¹⁰. My sample had one explicit Q&A style group, r/AskWomenOver30, and one partial Q&A group, r/Eskrima. Although both had a high percentage of top-level only threads, they had the highest percentage of threads with a depth of 4 or more (see Table 4.8).

There are a number of possible explanations for the different results here relative to D. Fisher, M. A. Smith, and Welser (2006). First, the reddit and Usenet are inherently different, with many different features that may affect the difficulty of replying. We may also have calculated the length and depth differently¹¹. I could not find information about how the authors calculated their *length* values, especially in the case where there may have been multiple replies to the same message. Both of our studies also only included two question-answer subreddits/communities, though around very different topics. The topical difference may also be the source for some of this difference, as both r/Eskrima and r/AskWomenOver30 may differ in the *types* of questions being asked, with technical questions support questions being conducive to shorter, more factual answers. In my qualitative analysis, I found both of my subreddits to be more mixed in their content than may have been present in the technical support style question-answer group from their study.

¹⁰There is a difference in terminology between the papers and related to how the structure of their platforms. Their work focused on Usenet which works similar to an email thread, and thus they refer to “length” instead of depth. They also use Q&A to focus on a technical support group, and they labeled their another group as a support group, which might also have elements of Q&A, though different style of questions.

¹¹Their reporting focuses on lengths of 2 and 5+ in terms of total messages within the thread. On reddit, the comments and submissions are separated. A starting message on Usenet is analogous to a submission on reddit. On reddit, a top-level comment is a reply to a submission and thus would be equivalent to Fisher’s length of 2. Their value of ‘5+’ would also be equivalent to ‘4+’ as I have calculated it here.

4.3.7 PARTICIPATION

The participating accounts differed in how active they were in the subreddit – some participated only once and others regularly active over long periods. Other studies have explored similar topics, as issues like commitment, loyalty, and retention of contributors and these can be important to the survival of online communities. In this section, I explore several different measures of participation within the subreddit and how this differed between the subreddits.

This theme aims to characterize how accounts participated in the subreddit over time and how equal that participation is. This differs from the interaction theme by not focusing on person-to-person interactions, but instead on how people participate more generally within the space over time. Subreddits, like many online communities, are sustained by having some amount of new content over time. Submissions and comments provide opportunities to participate and connect with other members and the community as a whole.

I focus on 3 specific facets that were useful for characterizing some of the differences between the subreddits. Two of these have a temporal focus: continued participation over time (retention) and the duration of participation within the subreddit (tenure). The final area I explore relates to an individual's contributions relative to the rest of the community, with a focus on the equality (or lack thereof) of participation. However, it should also be noted that there are other forms of participation and some areas of reddit encourage specific types of participation like r/Counting, r/PhotoshopBattles, r/WritingPrompts, and even the many r/Ask* subreddits.

4.3.7.1 RETENTION

Some of the accounts within my qualitative analysis appeared fairly often over the sample. In some cases, these were moderators engaging in fostering behaviors. Other accounts posted only once. This raised several questions. How frequently are people participating in the subreddit? How consistently is that participation over time? Without a certain level of continued usage, it would be difficult for norms to

4 Study 1 - Behaviors during the Early Life of Online Communities

develop and to be enforced, and potentially impossible for any trust to form between individual members and the larger groups.

Two concepts that have been used to understand these questions: retention and commitment. I define *retention* to be the number of people who were active in a given time period (t_n) who were also active in the following time period (t_{n+1}). Over time, the retentive users had likely found enough value from their participation to continue revisiting and contributing to the community. This can be thought of as a form of commitment to the community that kept them active over some period of time. On reddit, we can measure this by looking at people who contributed submissions or comments in one time period and again in the following time period. In this analysis, I included both comments and submissions and excluded the content where the author was deleted or disassociated with the content¹². More formally, I define the mean retention using the following formula:

$$MeanRetention = \frac{1}{12} \sum_{i=1}^{12} \frac{|A_{i \rightarrow i+1}|}{|A_i|}$$

Many first-time participants in online communities do not post again (Diakopoulos and Naaman, 2011; C. A. Lampe, E. Johnston, and P. Resnick, 2007; Singh, Johri, and R. Mitra, 2011; Whittaker, Terveen, et al., 1998). User retention has been used as a measure of success in online communities (Cunha et al., 2019; Ellis et al., 2016; B. Johnston, 2007). The continued contributions can help lay the groundwork for developing norms and culture within the community, as well as upholding and enforcing those norms (Backstrom, R. Kumar, et al., 2008). They can be a sign of fostering behavior, which is a sustained level of contributions over time. However, lower levels of retention may be a product of how some communities work. In health and support communities, it may be a sign of success as people receive support and move on (Massimi et al., 2014).

¹²The platform sets the author metadata in these instances to '[deleted]' and thus it is impossible to resolve any specific account

4 Study 1 - Behaviors during the Early Life of Online Communities

Subreddit	Mean % Retention	Mean Retentive Accounts
AnimalsFailing	5.80%	2.05
Fzero	13.65%	3.40
EarthScience	14.86%	1.86
Eskrima	22.01%	3.62
LinusTechTips	6.05%	9.49
WomensSoccer	29.74%	9.29
AskWomenOver30	36.62%	120.38
UpliftingNews	15.94%	1080.98

Table 4.9: Table showing the average retention rate for each of the sampled subreddits. Of note here is that r/AskWomenOver30, the question-answer/discussion subreddit, had the highest retention rate.

Averaged across the entire lifespan of the sampled communities, we can see a spread of retention rates for the subreddits (Table 4.9). The highest retention rate was for r/AskWomenOver30, which was the second largest subreddit. The lowest retention rates were in r/AnimalsFailing and r/LinusTechTips, the latter of which saw very little activity prior to the new moderator taking over and ultimately a 0% retention rate over that time. The retention rate since the new moderator was around 11%, though the last 5 30-day periods had a rate over 17%.

Looking at the retention rate over the lifespan of the subreddits (Figure 4.13), we can see some trends in the retention rates. Six of the subreddits experienced general increases over their lifespan: r/AnimalsFailing, r/Fzero, r/Eskrima, r/LinusTechTips, r/WomensSoccer, and r/AskWomenOver30. Several of these subreddits were largely inactive early in their life, and small numbers of retentive accounts later in their life. This is particularly clear in r/LinusTechTips, where the new owner of the mod took control around period 34. At this point, he began providing regular submissions which was effectively fostering some level of interaction in the community. The regular submissions he provided in that subreddit provided an opportunity for people to engage with it and ultimately be retentive.

Two subreddits have slight downward retention trends: r/EarthScience and r/UpliftingNews. The r/EarthScience subreddit has several 30-day windows with zero retention. This same period also saw a decline in active accounts and content. This included a period where there was retention. The decline happens a few periods after the creator's account appears to have been banned, and they move to a new account but submit less content on the new account. This new account stops posting content entirely in

4 Study 1 - Behaviors during the Early Life of Online Communities

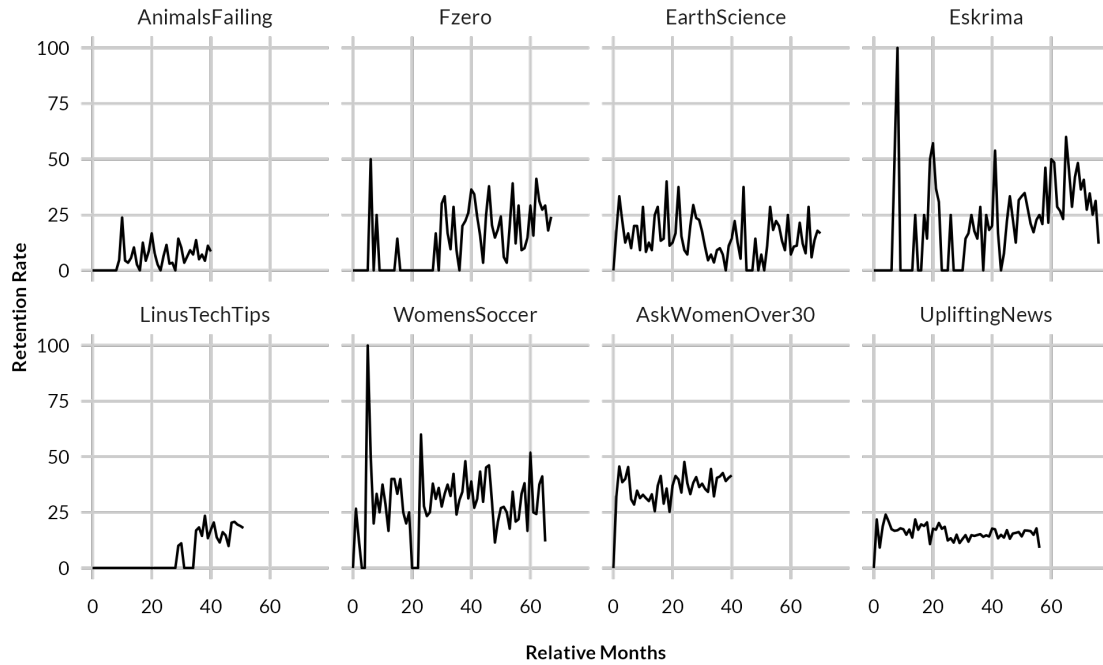


Figure 4.13: Line graphs showing the percentage of accounts that are retained from the one window to the next window. The windows are relative to the subreddits creation. This includes zeros for months with no activity. Four of the subreddits show increases in retention, with several showing increases after the subreddits began being more active.

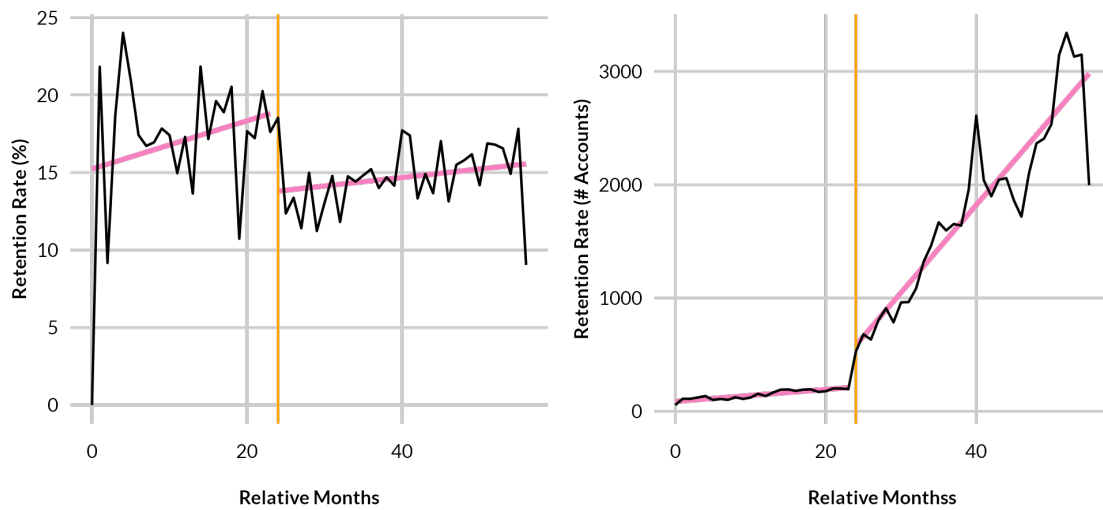


Figure 4.14: These graphs show r/UpliftingNews showing the effect of becoming a default subreddit (orange vertical line). On the left is the retention rate percentage, which declines when the subreddit was defaulted, and on the right is the number of retentive accounts. The latter is visible by looking at the raw number of retentive accounts after the event, which actually increases by ~270%, but the actual rate declines since it is a percentage of total accounts participating.

4 Study 1 - Behaviors during the Early Life of Online Communities

February 2015, but as of 2024 is not listed as deleted or suspended. Instead, we see a slow increase in total active accounts.

In r/UpliftingNews, the slight downward trend was due to slight drop around the point it became a default subreddit (Figure 4.14). This drop lasted for multiple 30-day periods before it began a gradual upward trend. The decline was driven by an increase in non-retentive users. This can be seen by looking at the total number of retentive accounts over time. This is most evident on r/UpliftingNews, which became a default subreddit around time period 24, which brought in many new users. There was a 270% increase in retentive accounts, but this increase was offset by the 400% increase in new total authors. In this particular case, it is unclear how valuable retention is as a metric as the subreddit is clearly growing, and the number of period-to-period retentive users is also increasing.

What value do retentive users have? To answer this question, I explored the amount of content that retentive authors create in every period compared to those that do not. This varied by the particular subreddit. Half of the subreddits had over 54% of their content created by retentive users. In r/AskWomenOver30, retentive users contributed nearly 65% of all content in the subreddit. In contrast, in the other four subreddits, retentive users contributed below 40% of content over its life. The retentive users in r/AnimalsFailing were responsible for 19.1% of the content in the subreddit.

In looking across the changes in evolution in these graph, there appears to be a larger trend. As subreddits, appear to become increasingly active, the retention increases – up to a point. The increased activity from others may encourage others to participate, through offering an increased amount of content. In this study, there is not sufficient evidence to back this, but there are two possible hypotheses about why this may occur. The first being reddit's algorithms have relied heavily on recency, thus more generated content may make the community more visible on a person's front page. The second is that people may have more focused interests, and the increases in content also increases the probability that the user may find something interesting enough to participate. However, additional research is needed to verify if this trend exists and then to verify what contributes to that.

4 Study 1 - Behaviors during the Early Life of Online Communities

The contributions of these retentive users provide artifacts around which other people can engage, and this can be useful for encouraging existing users and new ones to participate (Dabbish et al., 2012). This can be particularly important in subreddits like r/Eskrima and r/AskWomenOver30, where a large portion of the content are questions to the subreddit at large. Keeping people engaged can ensure that the expert knowledge people hoped for when they asked a question is still available. Retentive users can also help guide the subreddit and set norms around the content both through the content they submit and their engagement with this content.

High levels of retention may not always be necessary for a healthy community. When r/UpliftingNews became a default subreddit, the subreddit experienced extreme growth in both retentive and non-retentive accounts. The amount of content created in that subreddit by non-retentive users grew much larger than that of retentive users. This may have been enough content to still providing engaging content to keep people interesting and interacting through the subreddit. Even through the duration of r/AnimalsFailing, the amount of content produced by non-retentive was much greater at nearly every period. New members can bring new perspectives and help change norms in positive ways. This aligns with prior research, which found that turnover (the inverse of retention) can be important for the survival of a group or community (Ransbotham and Kane, 2011) and the negative effects of people leaving can be outweighed by some new people (B. Yu et al., 2017).

4.3.7.2 TENURE

Many accounts participated less than a day within all of the subreddits. This is common behavior on many social platforms and has been noted in a number of previous studies (e.g. Diakopoulos and Naaman, 2011; Singh, Johri, and R. Mitra, 2011; Whittaker, Terveen, et al., 1998). The ratio of longer-tenured accounts to these one-day accounts varied, partly due to people who were active longer than a year. These accounts may have been lurked for a while or simply missed engaging content on the subreddit to be active in two separate time periods. In this section, I explore some of the characteristics of the subreddits based on an account's tenure within the subreddit.

4 Study 1 - Behaviors during the Early Life of Online Communities

In subreddits like r/WomensSoccer, topical content may be largely seasonal like the games and people may only be interested in specific leagues or teams. Similarly, in subreddits like r/Eskrima and r/AskWomenOver30, not everyone may have the knowledge or experience to answer all of the questions that are asked within a given month. These people may still vote on content when it's submitted, and comment or create submissions less often. People may also not find content on their front page from smaller subreddits, because they did not check soon enough after a submissions was created. There are a number of reasons people may not show retentive behaviors as measured in the previous section, but still participate.

To explore this, I calculated the tenure the account were visibly active between their first and last piece of content. I excluded accounts where their first activity was after December 31st, 2015 in order to focus on accounts that had at least a year to participate again.

Subreddit	Total Accounts	Single Activity Accounts	Percent
AnimalsFailing	691	501	72.50%
Fzero	534	231	43.26%
EarthScience	531	392	73.82%
Eskrima	289	97	33.56%
LinusTechTips	598	332	55.52%
WomensSoccer	800	386	48.25%
AskWomenOver30	2851	957	33.57%
UpliftingNews	139528	75238	53.92%

Table 4.10: Table showing the counts and percentage of accounts which created a single comment or submission in the sampled subreddits. The question-answer subreddits have the lowest percent within the group, and the meme and news sharing subreddits have the highest.

Across all subreddits, the majority of accounts are active less than 24 hours, including many who produce only a single comment or submission. There were 145822 total accounts, and 78134 accounts (53.6%) created a single comment or submission. This number is slightly skewed by r/UpliftingNews which has a large number of single activity accounts. Removing r/UpliftingNews, the count of single content accounts is 2896 (46%).

Another way to look at this is to look at the length of time between a person's first and last piece of content within the subreddit (Graph 4.15). This can give us an indication of overall engagement within the

4 Study 1 - Behaviors during the Early Life of Online Communities

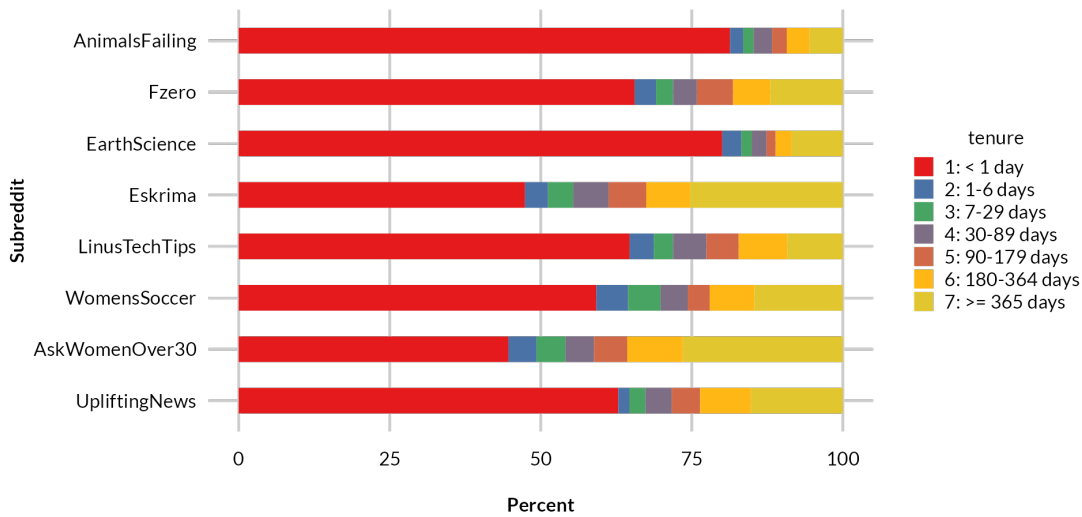


Figure 4.15: A stacked bar chart showing the proportion of users based on their tenure duration. This graph excludes people who first participated after 2016 to minimize the tenure skew towards lower tenures. Every subreddit except r/Eskrima and r/AskWomenOver30 have more than 50% of their content generated by people who are active in the subreddit less than 24 hours. Both r/Eskrima and r/AskWomenOver30 also have the highest percentage of users who were active more than a year.

subreddit. This does not show us overall activity during that period, but it does help us understand people who are less active in creating content but still monitor the community or vote. Looking at the duration the accounts were active, we can see that most accounts within each subreddit were active less than day. However, the distribution of longer participation (e.g. ≥ 6 months) varies between the subreddits. Even with a year+ of activity, there is a ceiling effect in limiting this in r/LinusTechTips, as it did not see a growth in activity until after the new moderator took control of the subreddit in 2015.

In some subreddits like r/AskWomenOver30 and r/Eskrima, we can see that the second largest group of people were active over a year. For r/AskWomenOver30, this is somewhat surprising given that the subreddit is one of the youngest subreddits in the sample at just over 40 months (comparatively, r/Eskrima is the oldest subreddit in the sample at nearly 76 months). Both of these subreddits were had more of a question-answer and discussion style of content, and one explanation is that people tend to be active longer in those types of communities.

Subreddits like r/AnimalsFailing and r/EarthScience which are more focused around information and sharing have few accounts active for over a year. r/UpliftingNews was potentially affected by becoming

a default subreddit. In one of the prior analyses, I saw that the level of interaction on submissions in the subreddit was low after becoming a default, suggesting that many people posting the first time may not receive an upvote or a comment. It is possible that the low level of engagement with these submissions negatively affects participants' willingness to return.

So while the source of this difference is unclear, by looking at the tenure we can see that some subreddits appear to have more long-term participation than others. Although all subreddits have a very high number of people who were active for a single day or less. The relative proportion of these single-day activity accounts varied substantially between the different subreddits as did the proportion of those active for a year or more.

Prior work has noted similar forms of participation in other online communities (C.A. Lampe, E. Johnston, and P. Resnick, 2007). People who visit a single day only may have no need to revisit the subreddit again (Singh, Johri, and R. Mitra, 2011), or some may simply have not received any interaction which deterred them from submitting or commenting again as has been found in the past (Joyce and Kraut, 2006). This may also just be part of a trial period to see whether they like the subreddit or not, as other scholars have noted (Freeman and Wohn, 2019). It is possible that some subreddits people, like r/UpliftingNews and r/EarthScience may be more informational and less discussion focused. Simply, seeing the content be enough to satisfy many people without the need to contribute. Similar behavior has been noted by research into lurkers (Nonnecke and Preece, 2001). Some past work has considered this to be a form of the free-rider problem (Kollock and M.A. Smith, 1996, e.g.), but the lack of long-term contribution does not appear to have affected r/UpliftingNews, but perhaps this matters less for well-established communities.

4.3.7.3 PARTICIPATION INEQUALITY

Within each subreddit, the participation was not equal. Some of the accounts were responsible for generating a large amount of the content over time, though the extent to which this occurred varied from subreddit to subreddit. One way to measure this is to use the Gini coefficient (Gini, 1912), which is a measure of inequality used in many fields to measure many different forms of inequality.

4 Study 1 - Behaviors during the Early Life of Online Communities

The coefficient produces a range of values from 0 to 1, where 0 would be complete equality and 1 would be complete inequality. For example, if every account generated only 1 piece of content then the Gini coefficient would be 0 because everyone produced an equal amount of content. I calculated this across both submissions and comments excluding both AutoModerator and deleted accounts.

Prior work has found that some communities have unequal participation early in their lifetimes (Whittaker, Terveen, et al., 1998), and that this may be a predictor of a number of forms of growth (Cunha et al., 2019). This would be true for communities which had a decent amount of seeding and fostering work, as was observed within several of these communities.

Other work has noted that there appears to be a relationship between the size of the community and inequality, suggesting that the inequality may actually increase as the community grows (Panek et al., 2018). While this may be true across many more subreddits, this does not appear to be true for all communities within this sample.

As seen in table 4.11, there is a fairly wide range of Gini values suggesting different levels of production equality. The smallest subreddit, r/AnimalsFailing, had the lowest value, while every other subreddit is over 0.5. This is related to the fact that many people posted or commented only once in r/AnimalsFailing. Subreddits like r/AskWomenOver30, r/WomensSoccer, and r/Eskrima have relatively high levels of inequality with coefficients above .7. This suggests that there are likely a few accounts creating a lot of content and many that create only one.

I also calculated the Gini coefficient overtime, by calculating it for each 30-day period relative to the subreddit creation (relative “month”). This helps understand how some of the participation within

Subreddit	Total Activity	# Authors	Gini
AnimalsFailing	1558	975	0.33
Fzero	2263	721	0.55
EarthScience	2477	745	0.67
Eskrima	2893	416	0.74
LinusTechTips	5820	2164	0.53
WomensSoccer	6955	1079	0.75
AskWomenOver30	44308	5004	0.75
UpliftingNews	731839	253547	0.54

Table 4.11: Basic activity data and the Gini coefficients

4 Study 1 - Behaviors during the Early Life of Online Communities

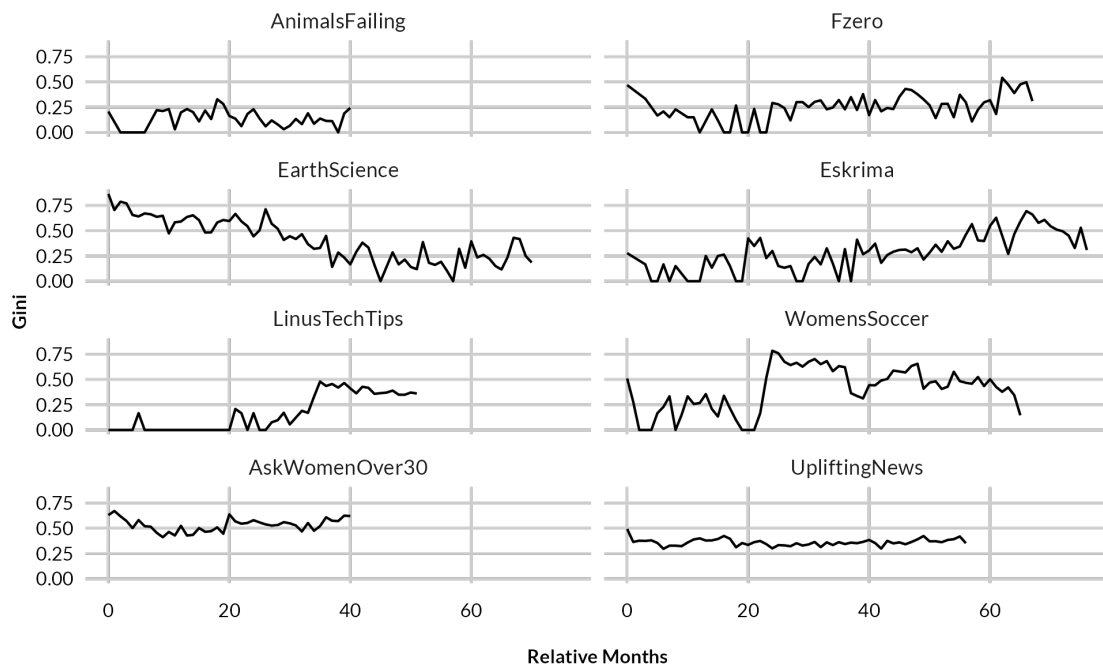


Figure 4.16: Plot showing the Gini coefficient for all activity over the 30-day months relative to the subreddit creation. Though several subreddits see a 0 near the start, this is generally a product of low activity and occasionally a single person posting.

the subreddit changed over time. Graphs of this data can be found in figure 4.16.¹³ There are two main observations worth pointing out.

Some subreddits appear to have trends over time, while others are somewhat stable, showing only some variation. For example, the two largest subreddits, r/AskWomenOver30 and r/UpliftingNews, appear mostly stable after the first few months showing only some variation over time. Others like r/EarthScience and r/Eskrima have clear trends, where the former is declining over time and the latter increasing. In the case of r/EarthScience, this was due in part to the creator gradually posting less over time while other accounts started posting more. The creator was posting a lot of the content for the first few years (e.g., fostering), while few others created content, thus creating some inequality. In r/Eskrima, we see that some authors become more active, thus creating inequality over time. A handful of authors become very active, creating portion of the content in later months.

Several of the plots show a high Gini value for the relative month 0 before falling. One contributor to this was the seeding and fostering work, typically put in by creators. For example, this is visible in r/UpliftingNews (also see Figure 4.5). Typically, their effort was largest in the first month (typically seeding) and slowly declined as others began contributing, thus reducing the inequality. The clearest example of this is r/EarthScience, where the creator is the most active account for months, posting a significant amount of content though this slowly declined over time until the creator's second account appears to have been banned.

The difference in values between subreddits suggest that they subreddits may allow or even encourage different forms of participation and that this may change over time. While some subreddits, like r/AnimalsFailing, appear to have less inequality in content production, this may not be a positive outcome, but instead could be the result of fewer people participating.

¹³Two quick methodological notes about the graphs. The Gini coefficient and many of its calculations, like many, are problematic with little data. This is visible in some of the smaller subreddits. For example, some values may be misleading due to a lack of overall content and participants. For example, r/LinusTechTips has a number of months with a 0 coefficient, suggesting it may have perfect equality. This is true, though it is because it may have had one or two authors create a single piece of content. Additionally, the low levels of activity can lead to some large fluctuations in the coefficient. This is visible in several of the smaller subreddits like r/Eskrima where it appears to alternate between 0 and about .25 for a few months.

It is also worth noting that these results do not appear to fully support the findings of Panek et al. (2018), who suggested that as communities grew the level of inequality also grew. I observe only a single subreddit with a clear growth trend for inequality, r/Eskrima, though some others do see jumps as activity increases before seeing a declining trend (e.g. r/LinusTechTips and r/WomensSoccer). For the most active months of subreddits, the Gini coefficient appears to be between 0.25 and 0.75, suggesting that some level of inequality may be necessary within these online communities. However, this dataset is too limited in its size to make such claims, it may be worth future research.

4.3.8 SUMMARY OF RESULTS

I completed a number of analyses across the perspectives to help develop the theory, and I want to briefly summarize some of those findings here. At a high level, all of the themes and analyses found differences between subreddits or over time, though the sample size is limited and thus more work is needed to validate and further the development of this framework.

First the content submitted to the subreddits different between the subreddits, though largely with the ratios in which they appeared. I selected subreddits to include a broad sample of multimedia and content. However, I found that questions and meta discussions appeared in all of the subreddits, and was generally well-received.

I also found accounts that engaged in *fostering* behavior, which I define as a form of community-building through sustained activity within a community. I found many accounts across all subreddits which created content over longer periods of time. Because my dataset was limited to content, I focused on the sustained content production, though I theorize this extends to the less visible work like that of moderators. I also found a special subset of fostering, seeding, as typically an effort in the early life of a community to bootstrap content within it. This work was often done by creators or moderators in the subreddits.

Promotion was the third theme of the framework and it was a method of raising awareness and attracting people to the subreddit. While I found explicit forms of promotion, I also found many which were implicit and may not have been intentional forms of promotion. However, the platform renders

these as hyperlinks, thus allowing them to operate as promotions nonetheless. These came in the form of links, mentions, and cross-posts.

I also found evidence of the accounts interacting differently across the subreddits. I found different rates of voting and commenting on submissions between the subreddits. This may have been partially a function of the size or capacity of a community to respond to content, as the largest subreddit also had the lowest levels of each. However, I found some limited evidence to suggest this may vary by community type, with the Q&A style communities appearing to have higher rates of interaction when compared to others. The conversation styles also varied between subreddits, with some having deeper conversations (e.g. more back-and-forth replies). This may also have varied between subreddits.

Lastly, I found that the subreddits different in the percentage of people who remained active from month to month (retention) and the amount of content they produced (participation inequality). Most of the subreddits showed signs of a positive trend in retention or at least stable retention, suggesting that more accounts are remaining active between months. For the one default subreddit, I found a sudden decrease in retention after becoming a default subreddit, and a slight upward trend after. The subreddits also had different levels of participation inequality. The subreddits generally fell within the range of moderate levels of inequality, which I theorize may be necessary as the subreddits may rely on more active and fostering authors to provide content.

4.4 DISCUSSION

In this study, I have explored the early life time of eight different subreddits to understand the behaviors within them which can be used to characterize the subreddits. My exploratory analyses have yielded a number of findings which extend existing research in CHI and CSCW around the formation of online communities and the behaviors that occur within them. This includes a more thorough classification of the types of questions asked within subreddits, description of a fostering behaviors likely intended to build community, and the promotional work that happens outside a community.

4 Study 1 - Behaviors during the Early Life of Online Communities

At its core, the central theme of this work is the emergence of community. In some of the communities I have analyzed, there was an intentional effort to build and grow the subreddit. This includes creators and moderators inviting members to help with some of the work, avid members continuously posting content, and members working together to get a subreddit made a default subreddit. Other subreddits evolved gradually with less intent, with new members slowly finding the space and asking questions or sharing content sporadically. One of the larger differences between them was the time it took until the subreddits had a stable amount of activity. The largest were thriving by the end of their first 30 days, whereas the smaller took much longer.

4.4.1 A FRAMEWORK FOR ONLINE COMMUNITIES

I grouped the themes of my analyses into five perspectives, which I believe may be useful as a framework to understand the formation and early life of online communities. These are the result of a grounded approach, such that the framework is developed from the analysis of this data, rather than from existing theory. The five perspectives I included were: content composition, fostering, promotion, interaction, and participation with the latter four capturing behaviors, and the content composition being largely descriptive of the common forms of content produced within the subreddit.

There are two key things that distinguish my framework from other frameworks I reviewed. First, my framework is grounded on the data I observed in this study. As I iterated on my coding scheme and did further analyses, I definitely incorporated some analyses based on methods and metrics used in other works, but the initial behaviors are what triggered these. Second, my work focuses specifically on the earliest part of a community as it grows (or not) from the point of creation where it is likely not considered a community. Two of the perspectives specifically address the early behaviors I observed in these communities: fostering and promotion. My framework could be used throughout the life of the community. Similarly, the other frameworks (e.g. Lazar and Preece, 1998; C. P. Lee and Paine, 2015) could be used during the early life of communities, though neither captures the placemaking behaviors.

Four of the perspectives in my framework were developed specifically from my initial analyses, with *content type* being the exception. I included content in my initial selection criteria, based on

4 Study 1 - Behaviors during the Early Life of Online Communities

my reading of existing literature like (X. Wang, B. S. Butler, and Ren, 2013). I found enough differences between the community types, especially with respect to questions, that I included it as a perspective of its own. Other perspectives like interaction and participation are common characterizations found in some other frameworks, models, and operationalizations discussed in the background section. Being a community, some measure of interaction is important for many analyses. It is especially important for young communities, especially from the point of creation until they are self-sustaining as it can act as a measure of interactivity and perceived value (Kraut and P. Resnick, 2012).

This framework has some overlap with existing frameworks. For example, the participation theme in this framework is similar to the turnover and to a lesser extent the nascence dimensions of the Model of Coordinated Action (C. P. Lee and Paine, 2015). Similarly, some of the ideas from content composition theme are similar to the Community Archetypes (Prinster et al., 2024). The latter work also contains four thematic constructs for how researchers assessed the antecedents of feeling a sense of virtual community (A. Blanchard and M. Markus, 2002). These four constructs overlap along interactivity and membership boundaries, the latter of which they include *prolonged participation* as an assessment strategy which overlaps with participation and fostering from my framework.

There are elements that are notably missing from this framework. For example, size or volume is absent. The size of a community may affect many aspects of a community, but the biggest difference I found for size was how it affected the amount of interaction. For example, when r/UpliftingNews became a default subreddit, it grew very quickly in both active accounts and content, but so did the number of submissions that received no interaction (See section 4.3.6). Some evidence suggests that size may change the perception of a community (TeBlunthuis, Kiene, et al., 2022) and how people approach them, but it is unclear whether this difference is captured in other framework perspectives. A number of past work, including several frameworks like C. P. Lee and Paine (2015) do include size as a dimension, as it can affect how many aspects of the work within them change. This may notably change how fostering works within these subreddits or how much is needed.

4.4.2 WHEN IS COMMUNITY?

This study addresses time indirectly, but time is still an important factor. Does a community become one simply after checking off a list of criteria? This work does not focus on the psychological aspects or sense of a community (e.g. A. Blanchard and M. Markus, 2002; Doolittle and MacDonald, 1978; McMillan and Chavis, 1986), but one question from this work is whether some behaviors like fostering can lead to people feeling a sense of community? In other words, are there behaviors which may lead to or otherwise predict when people might feel a sense of community?

At creation, none of these subreddits would likely be considered a community as nothing happens within them. Early fostering behaviors, like seeding, transform the empty space into one that is “habitable” in a sense. Each additional piece of content adds activity and some potential value to the space. The activity (and content) in the subreddit can be similar to decorating a house in that it can transform the place to something that feels lived in. Each contribution can contribute to decorating the space while helping convey some of the potential value of the space. In a Q&A style subreddit, questions and answers provide both evidence of activity, provide some value as a potential resource on a topic, and also conveying the purpose and expectations of the community. However, this can take varying amounts of time to amass. Some of the subreddits had a slow journey to becoming a place where people interact, which is possibly a precursor to creating the “sense of (virtual) community” that some people feel in these spaces (A. Blanchard and M. Markus, 2002; Prinster et al., 2024).

From a theoretical perspective, it can be useful to think of these spaces prior to the development of the sense of community and to focus on this “in between” space where no participants may feel a sense of community or even think of it as one. In several of the subreddits, activity was very sporadic early on. Most of these did become stable over time, though r/EarthScience does appear to start struggling as the mod and creator disappears towards the end of the subreddit and does not reappear, but the subreddit continues to see submissions and comments.

Communities are not static entities. In an early work on community studies, Homans (1951) mused that “Hilltown” had ceased being a community. Many of the properties people attribute to communities

had disappeared from Hilltown¹⁴. Works on online communities have often set theoretical thresholds for when a community 'dies' or ceases to become a community (e.g. Kairam, D. J. Wang, and Leskovec, 2012; Tsugawa and Niida, 2019). Some prior work has considered a community to die if they have a single month of inactivity. By that definition half of my sample of subreddits would have been considered dead, but ultimately every one was active by the end of 2016. From a theoretical perspective, the rocky start of some of these subreddits may be simply part of the process to becoming a community, despite failing to meet requirements set out in some research.

This raises theoretical concerns about how we view communities from simple metrics. As other scholars have noted, common metrics and a focus on growth or size as a measure of success may not align with how participants within those communities view them (Cunha et al., 2019; Massimi et al., 2014; Prinster et al., 2024, e.g.). Smaller, less active communities can still have value for participants (Hwang and Foote, 2021; Leimeister, Sidiras, and Krcmar, 2006), though those values may be different from larger ones (TeBlunthuis, Kiene, et al., 2022). This implies that in some past research, and even this research, some spaces which may be thought of as communities by its participants, are not considered communities through some of these metrics.

Further, some have argued that with continued growth, some communities begin to act more like crowds rather than communities as they scale (Panek et al., 2018), which may change how its members relate to the community and other participants. Similar concerns have been raised by early community scholars when comparing large cities and metropolitan areas to more rural towns. This may change how people think of these spaces and how they participate in such spaces. These types of growth may also help explain the perception that spaces have changed after periods of large growth (Kiene, Monroy-Hernández, and Hill, 2016; Lin et al., 2017).

These smaller spaces may provide interesting opportunities for researchers and platform maintainers to better understand the challenges of creators and moderators in these early phases. This may yield more

¹⁴Hilltown is not the actual town name. Homans chose the name when analyzing prior data where it was previously called "Indecisive Hamlet" (Homans, 1951, pg. 339).

tailored and informative success metrics and design interventions to potentially help the community. This could include structures for helping find content and growing awareness. However, the specific structures need to align both with the goals of the founders (Foote, Gergle, and Shaw, 2017) and early members.

4.4.3 REDDIT COMMUNITIES AS RESOURCES FOR QUESTIONS

I found that every subreddit I explored had at least a few questions submitted to them. While some had only a few questions, many appear to have been well-received by the community. This was surprising as some communities were more focused on sharing content generation or sharing news, but still had some questions. One possibility might be that people view some online communities similar to a resource than has been previously discussed in prior literature. Prior work has specifically called out Q&A and support subreddits as being resources for people, and members have found them to be valuable. Seeing other forms of communities as places to ask questions is an interesting behavior, especially those focused around content generation or meme-sharing. This may explain why some subreddits create additional communities specifically for asking questions as a means of keeping those separate from the main content.

I noted a number of different forms of questions people asked in the subreddits including information seeking and advice seeking, both of which have been noted in previous research. Asking questions of a community is one of the key value for many forms of online communities and a motivation for participating in the community (Galegher, Sproull, and Kiesler, 1998; Shrestha, Barik, and Parnin, 2021; Tausczik and Pennebaker, 2012). Some online communities are focused around this behavior (T. L. Roberts, 1998). For example, information and support seeking have been found to be important in health-related communities (Kingod et al., 2017; Ridings and Gefen, 2004; Welbourne, A. L. Blanchard, and Wadsworth, 2013, e.g.). Question-asking has been noted as something particularly important to newcomers in online (Golder and Donath, 2004) and this behavior may decrease over time as tenure may cause some participants to leave or switch to answering questions (Yang, Kraut, and Levine, 2017).

I also noted two additional forms of questions, which are novel among the existing literature. The first was resource seeking, where the community is for information or content that is located *outside* community. In some cases, this was pointing to books or videos about Eskrima or websites or bars to

4 Study 1 - Behaviors during the Early Life of Online Communities

watch a soccer match. The act of seeking local resources within a area was also common among some accounts. People may ask these questions of communities if they feel the information or resource is not available within the community, if they have exhausted their ability to use search engines, if have place more trust in personal recommendations over search engines, or if someone may not have the expertise to know what to language to find relevant resources. In some cases this was closely related to information seeking, but the results may be more recommendations.

The second form were questions intended to start discussions, rather than seeking any specific answer. These questions were intended to engage the community rather than seek information, and were typically framed in a way where answers were more opinion and experience based. This particular form of question is particularly important because of low barrier of entry, which makes it easier to answer. Given the perceived importance of interaction as the essence of community (Prinster et al., 2024, e.g.), identifying submissions like these could be relatively important for helping engage visitors and new members for discussion-focused subreddits.

There may be value in having conversational spaces as side channels for communities focused on content production. Some subreddits often have evolved to create secondary spaces to push certain types of related content into, in order to preserve the original focus of the first. Notable examples of this include the several r/Ask<Community> subreddits such as r/AskScience and r/AskThe_Donald, which served to focus specific questions of the broader community in. This may preserve the value and focus of the original community, while also giving people a space for other forms of content like questions of the community or even for off-topic sorts of discussions.

In a similar capacity, many communities have evolved to have “Off-topic” forums in order to support some of the more socializing behaviors for their members. These types of discussions have been observed on some communities, and Gazan (2009) has argued that the development of such behaviors is a form of self-awareness and may be part of the transition between space to being a community. These additional channels for social content can provide areas where people get to know each other beyond the standard

interactions enabled in the main communities, thus forming the bonds for community among their members.

4.4.4 FOSTERING

Another contribution of this work is the concept of *fostering* behavior as a sustained amount of community-building work. Online communities can be seen as multi-faceted resources for people seeking information, advice, resources, and even support. However, these communities and the resources they provide require some a level of sustained participation. The work for sustaining a community can include many different behaviors including curation, interaction, norm enforcement, and content moderation.

Fostering activity attempts to build or maintain the place of a community. In the early time of the subreddit, this can be searching for and finding and contributing submissions similar to the creator of r/UpliftingNews did in the early history of that subreddit. I focused on the production of content because that is what is visible within my dataset, but prior research suggests there are many tasks done by people within online communities which may be considered fostering (e.g. B. Butler et al., 2002; Li, Hecht, and Chancellor, 2022; Lo, 2018; Seering, T. Wang, et al., 2019). A prime example of this is moderation work, which often remains invisible to most participants, where moderators debate rule violations and make decisions about new rules (S. A. Gilbert, 2020; Li, Hecht, and Chancellor, 2022; Seering, T. Wang, et al., 2019).

From a theoretical perspective, fostering can be seen as placemaking (Dovey, 1983; Harrison and Dourish, 1996)¹⁵. This work is like the necessary chores and upkeep of any form of housing which is needed to keep the space habitable and ensuring that the space best meets the needs of its members and visitors. It is this work that can help transform this from an empty space, devoid of social interaction, to one that fosters social interaction.

A key word in my description of fostering is *sustained*. Some level of engagement over time may be necessary to develop and maintain norms and rules. New visitors may look to content to understand

¹⁵Many of the concepts behind placemaking and the differentiation of space and place were introduced to CSCW by Harrison and Dourish (1996), but the concepts of place as social meaning was a topic prior to their paper. The authors note some of this in that paper.

4 Study 1 - Behaviors during the Early Life of Online Communities

norms (Danescu-Niculescu-Mizil et al., 2013; Welser, Gleave, Barash, et al., 2009), but those norms and rules may not be immediately obvious to all newcomers. This can require different strategies for attempting to enforce those norms, as not all methods may work well in all situations (C. B. Smith, McLaughlin, and Osborne, 2006). The enforcement of these norms can take time, sometimes requiring significant discussion about how to handle them (S.A. Gilbert, 2020; Lo, 2018). The extended engagement also applies to those producing content as these norms also have to be learned, and the platforms algorithms require a steady stream of content to show to keep others engaged.

While many platforms and designers have sought to ease some of the burden of this work, there still remains much that can be done. Moderation tools continue to be an issue on reddit (Matias, 2016; Matias, 2019). Much of this is focused around controlling content and limiting the effects of bad actors, but there are potential design interventions for helping other aspects of fostering behaviors, like that of seeding and content curation. Additional platform changes could be made to help people with this content curation, such as recommending content from other subreddits or from popular news and media sites like YouTube.

This is particularly important for new social platforms, which are seeking to attract new members. New social media platforms are aiming to capture some of the audience that left Twitter after it was purchased, but many suffer from reaching the critical mass necessary. One aspect of this is the content curation. People come to social platforms like Twitter and TikTok to solve many needs, especially about understanding what is currently happening in the world. If someone opens an app to quickly catch up on news, get a quick laugh, or catch up with their friends they need recent content to satisfy many of those needs. Thus, there are arguably many more opportunities to help with the initial phases of platform development. Minimizing the effort to find and create meaningful content is a massive challenge, and once which seems largely unapproached by many platforms. At their best, some social media platforms offer a few prompts for eliciting content from users.

Preece (2000) saw the recruitment of early adopters or 'seeders' as a critical part of this of developing online communities. While this strategy may work, it may be difficult to do for all people. Two of reddit's founders seeded content on the platform themselves by curating and posting content under fake users

they created (C. Johnston, 2012). This may have led to the perception of the platform having more active users than it actually did. Despite the founders' personal experience with fostering work, the platform lacks features to help reduce the effort necessary to find high-quality content and share it with a new community.

4.5 LIMITATIONS

One of the greatest assets and challenges of qualitative research is the interpretive aspect. I try to limit inferring intent, but there were instances where I could triangulate through multiple data sources (e.g. author comments). This work still relies heavily on my interpretation without having spoken with many of the authors. I have spent a significant portion of time trying to find similar examples that help corroborate the work and to fortify the central points of this work. In many cases, I have attempted to trace the authors activity beyond the subreddit to better understand and contextualize their traces.

This work is also based on a limited sample. My goal was to find a diverse sample of subreddits across dimensions, but this inherently comes with a trade-off that many interesting behaviors were not seen in other subreddits because they vary in type or size. The size of the sample alone, also means that much of the findings here may not apply beyond these eight subreddits.

4.6 CONCLUSION

In this chapter, I analyzed subreddits to understand behaviors and characteristics that occur between them. I then developed a framework of 5 perspectives of community, which was grounded within the themes and behaviors I observed in these subreddits. In the next chapters, I will explore how well some aspects of the framework can be used to understand different forms of success (Chapter 5) and to analyze and understand the evolution of a large political subreddit (Chapter 6). Together, these chapters serve to illustrate both how this framework might be used, while also showing that it has some potential value for other empirical work.

5 FOSTERING AND PROMOTION DURING EVOLUTION

In Chapter 4, I described several place-making behaviors people performed in the early life of communities. Some of these behaviors were intended to establish an identity for the subreddit, attract subscribers, and encourage participation. I grouped these behaviors into four perspectives: interaction, participation, fostering, and promotion. A body of research has explored aspects of interaction and participation (Hamilton et al., [2017](#); Kang and Shin, [2006](#); Panek et al., [2018](#); Whittaker, Terveen, et al., [1998](#); J. Zhang et al., [2017](#)), but fostering behaviors and promotional behaviors remain understudied. Seeding and promotion have been named as best practices for starting online communities, but often without empirical evidence. In this chapter, I explore how, if at all, these behaviors are related to three different measures of success of online communities. I also explore who does some of this “work” of fostering and promotion within a larger dataset than the previous chapter.

Building and sustaining communities is challenging. Many communities may fail to attract and maintain enough users to sustain the communities (e.g. Cunha et al., [2019](#); Kraut and Fiore, [2014](#)). Some prior work has focused on the role of creators (aka founders) and their goals and affects on the community (Foote, Gergle, and Shaw, [2017](#); Kairam and Foote, [2024](#); Kraut and Fiore, [2014](#)). Many community participants contribute infrequently and membership can be highly transitional (B.S. Butler, [1999](#)). Some people, often called lurkers, are content being consumers and may never produce any content (Nonnecke, D. Andrews, and Preece, [2006](#); Nonnecke and Preece, [1999](#)). When people begin participating in a community, the response they receive can impact whether they participate again (Arguello et al., [2006](#); Joyce and Kraut, [2006](#)). Additionally, as members stay within a community longer, their behaviors may change as their relationship to the community changes (Ganley and C. Lampe, [2009](#)).

5 Fostering and Promotion during Evolution

When people visit an online community for the first time, they make decisions about whether to return based on the perceived value of the community and its content at the time. After a community is created, *seeding* may create some initial value for the community through the creation of content new users can engage with when visiting (Kraut and Fiore, 2014). Seeding has been recommended by some scholars (e.g. Preece, 2000) and reddit's own help documentation (Reddit, 2023c) as a means to create value for early communities. reddit's founders have also claimed they seeded content in the early days of the platform by using new accounts in order to make the platform feel active (C. Johnston, 2012). The early content and work founders put into groups may improve their survival or growth (Kairam and Foote, 2024; Kraut and Fiore, 2014). Cunha et al. (2019) found some evidence that high levels of participation inequality early in a community's life was predictive of multiple different measures of success on reddit.

In decision models of adoption, awareness is a crucial step that precedes all other steps like purchasing or joining (Bateman, 2008; De Bruyn and Lilien, 2004). People cannot join most communities without being aware of them¹. Promotional behaviors can provide awareness of the community and attract new participants. In multi-community spaces, promotion can provide a bridge between similar communities and eventually enable communities to specialize their topics. Promotions can include invites, referrals, shared content, search, advertising, and user interface elements highlighting members you might know (Bateman, 2008; Kraut and P. Resnick, 2012).

Reddit implements a number of affordances for recruiting new members including sharing content, invites, search, cross-posting, and links. Cross-posting is a behavior where a specific piece of content is shared to multiple communities. This may be done for many reasons including creating a bigger audience for the content (TeBlunthuis, Kiene, et al., 2022), sharing relevant content with other communities (e.g. possibly smaller, more niche communities), or to explicitly promote of another subreddit. A related

¹I emphasize the word *most*, as there may be communities where a person is granted membership in advance of knowing about it. On some platforms and systems, people can be added to hidden groups or communities without the use of an explicit invite. An example might be a group chat created to organize an event, where other people get added over time. This depends on the specific definition of membership and "joining", but I caveat this to avoid making blanket statements about communities I am unaware of. reddit does have some secret communities with specific criteria for joining, but these utilize invites.

concept are links and mentions, which can appear as hyperlinks in text content on reddit, which can link to a specific subreddit (e.g. r/AskReddit) or to specific content on a reddit (e.g. a submission, post, or wiki page). Prior research has found that these links can increase activity in the linked subreddits (Krohn and Weninger, 2022), though it is unclear if this increase in activity leads to longer-term participants. Reddit's own help documentation mentions a several forms of promoting communities including cross-posting and mentions (Reddit, 2023a; Reddit, 2023b). While these forms of promotion can drive new activity in other subreddits, the resulting activity may not be positive for the community. This could motivate new people to join who harass members or generate low-quality content. Reddit's help documentation warns their usage could be considered spam, and links to other communities can also encourage inter-community conflicts and harassment (Datta and Adar, 2019; S. Kumar et al., 2018).

Thus the central question of this chapter is: *Can fostering and promotion predict some forms of success in online communities?* I explore this with two different prediction tasks for three different measures of success: growth, retention, and survival. The two analyses in this chapter explore these questions, and more generally look at the possible value of the framework perspectives in the context of multiple measures of success. I also explore an additional question of 'who does the work?' with a focus on creators, moderators, and normal contributors.

5.0.1 SUCCESS IN ONLINE COMMUNITIES

It is important to note that success can take many forms for different types of communities. Founders and participants can have different goals for creating and participating in a community, and these can evolve over time. A body of early research focused on growth and survival as measures of success for online communities (e.g. Ducheneaut et al., 2007; Kairam, D. J. Wang, and Leskovec, 2012; Kraut and Fiore, 2014; Raban, Moldovan, and Jones, 2010; Y.-C. Wang, Kraut, and Levine, 2012). These were important measures of success in the early days of USENET and the Internet, as many attempts to create communities failed. Long-term growth may not be a goal for all communities (Bruckman and Jensen, 2002; Kairam and Foote, 2024; Ridings and Gefen, 2004). Many other motivations and goals for communities exist including helping others, learning, and ability to network (e.g. Kairam and Foote, 2024; Matthews et al., 2014; Ridings and Gefen,

2004; Shrestha, Barik, and Parnin, 2021; TeBlunthuis, Kiene, et al., 2022). The topic and type of a community may also influence the goals of the creator or members (Ridings and Gefen, 2004). People may participate in health-related communities to get information and support while dealing with longer term health issues, but may eventually not need to participate (Wohn and C. Lampe, 2018).

5.1 DATA & METHODS

This study relies on a subset of the Pushshift dataset (Baumgartner et al., 2020) that is described in Chapter 3. I focused on subreddits created in 2012 in order to enable longer-term analyses. The year 2012 is 4 years after the introduction of user-created subreddits, which gave time for the users and platform to adapt to the feature. There were a total of 122.6 thousand subreddits created in 2012, but only 86.5 thousand had any activity.

This study focuses on temporal behaviors. To help address the problem of unequal month sizes, I use 30-day windows for all calculations which I refer to as *relative months*. These also allow me to align all of the communities relative to their creation date, such that the first month after creation is relative month 0 and onward.

Fostering and seeding behaviors require activity (content) to measure and thus I chose to exclude communities with little activity. For the first analysis, I borrow a paradigm from (Cunha et al., 2019), which measures predictors at the relative month when the community gains its Nth new participant. When the 10th new account participates in the community, I take a snapshot of the data in that month (N) and try to predict N+12 relative months later. The activity thresholds range from 10 active accounts to 100 accounts per relative month, and thus in each set, communities which do not reach these thresholds are excluded. This attempts to control for the differing growth rates, and thus attempts to compare subreddits at similar sizes. I also excluded communities that did not reach the threshold before December, 2015, as the success metrics are year-long metrics.

In the second analysis, I set a minimum threshold of 100 comments, 100 submissions, and 10 unique accounts in the subreddit's lifetime. I needed a minimum level of overall activity to be able to

calculate some of the features while still including communities that do not survive. I also excluded communities that did not have activity in at least 3 of the 6 months after creation as the analysis needs initial activity to calculate the monthly growth rate. Since the dataset contains data through the end of 2016, the communities had at least 3 full years to reach the minimum activity requirements thus including less successful subreddits. In total, 4033 subreddits met these requirements.

For the promotion data, I used a regular expression identify subreddit mentions and links in comments, submissions titles, and the submission selftext. The regular expression identifies a number of string variations that both mention and link forms (...r/<subredditname>...). I searched through all comments and submissions across the entire dataset to create a subset of just submissions and comments and the mentioned subreddits. These were stored in a database and checked against subreddits that existed at the time. The regular expression does not properly work on multi-reddit links², but these were rare.

5.2 PREDICTING GROWTH AND RETENTION

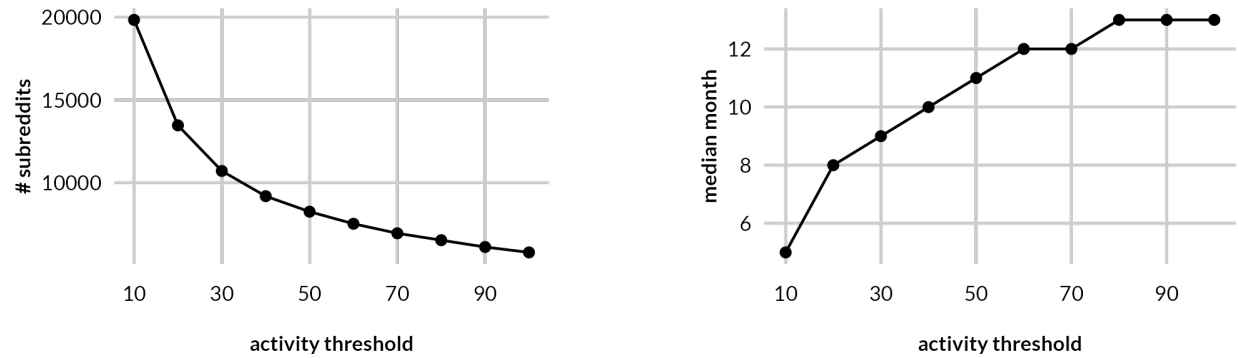
In this first analysis, I explored the relationship between the framework perspectives and two measures of success: new account growth and retention. Account growth is important for all communities early in their life and for those that value large audiences. However, some communities may prefer smaller more focused audiences where month-to-month retention of members may be a more valuable success metric.

5.2.1 METHODS

To explore the relationship between success and the framework perspectives, I employed a prediction paradigm used in Cunha et al. (2019). This casts the relationship problem as one where groups of features can be used to predict success metrics later in the life of the subreddit. If a model can predict success better than random it suggests that these variables may be valuable for explaining a particular metric. Prediction has value over methods, such as regression, when the underlying data violate assumptions of regression models. In this case, the data was zero-inflated count data that is non-normal.

²Multi-reddits is a special URL notation that includes the plus sign between multiple subreddit names like r/AskReddit+science+news which effectively specifies all three

5 Fostering and Promotion during Evolution



(a) Graph showing the number of subreddits which crossed specific activity thresholds.

(b) The median month when subreddits crossed the specified activity threshold.

Figure 5.1: Two plots showing a summary of the subreddits created in 2012 that crossed thresholds of a number of unique active accounts in a given month (relative month).

The prediction task was done as a binary classifier using a random forest (Ho, 1995). I used a median-split of both metrics to dichotomize the variable. This has the advantage of yielding a balanced classification task. I split the data into a training (80%) and test (20%) set. The data was centered and scaled to adjust for differing scales between variables. The training was done with 10-fold cross validation in R using the caret and ranger packages (Kuhn, 2008; Wright and Ziegler, 2017). I trained models for a group of general subreddit features, four groups of features from the framework perspectives, and a combined model with all features. This setup was repeated for both measures of success.

Since communities can take a while to establish themselves, I followed Cunha et al. (2019) and snapshots of the subreddit when they reached certain growth thresholds of active users. I focused on the thresholds (k) of 10 through 100 in increments of 10 (e.g. 10, 20, 30, ...). I then attempted to predict success 12 months past the threshold month. This also provides a much higher number of eligible communities at lower thresholds ($k=10$), where success may be less likely than those that reach 100. The number of eligible subreddits at each threshold is visible in figure 5.1 and the median number of months it took eligible communities to reach that. Communities which did not reach the threshold before December, 2015 were excluded from the training and test sets for that threshold and later ones.

5 Fostering and Promotion during Evolution

Group	Feature
General	Number of unique accounts
	Number of unique accounts in the month the creation month (n=0)
	Gini coefficient of total account activity in the month of creation
Promotion	Number of comments in other subreddits that mention or link to the subreddit
	Number of submissions in other subreddits that mention or link to the subreddit
	Number of crossposts in other subs which are links to the subreddit
Fostering	Number of seeded submissions in first month
	Number of fostering submissions
	Number of fostering comments
Interaction	Number of submissions with positive engagement
	Fraction of submissions with comments
Participation	Gini coefficient of all commenting accounts
	Gini coefficient of all submitting accounts
	Fraction of accounts that are retentive in month

Table 5.1: A list of the groups of features used in the prediction task.

5.2.2 MEASURES OF SUCCESS

Both new account growth and retention are focused on the number of unique accounts during a specific period. For account growth, I focus on the number of number of new accounts that joined within the time frame (J). For retention, this is focused on all accounts active within that timeframe (A).

Account Growth. This measure looks at the cumulative increase in new accounts in the 12 months after the subreddit reached the specific activity threshold. This is calculated using the union of all new accounts which first become active in that time period. This would be $Growth = \sum_{i=1}^{12} |J_i|$, where J_i would be the newly active accounts in month i .

Retention. The second measure is the average retention rate in the 12 months following the month the activity threshold was reached. This is the fraction of users who are active in month i who are also in the following month ($i+1$) relative to the threshold month. More formally, it is $Retention = \frac{1}{12} \sum_{i=1}^{12} \frac{|A_{i \rightarrow i+1}|}{|A_i|}$.

5.2.3 PREDICTIVE FEATURES

I calculated a number of different features which I split into the 5 groups of features which can be seen in table 5.1. Initial exploration included more features, but many of these had very high levels of correlation. Although random forests should prune highly correlated features, this can lead to challenges interpreting how informative features are. I opted to reduce the feature groups using my intuition about the platform and the outcome variable. For example, for promotional content many authors created only one promotion, while others created a lot. I theorized that the number of promotions may be more important than the number of authors creating promotions and removed it from the feature set.

Three features, including seeded content, were calculated on the creation month only. This is similar to the metric used in Kraut and Fiore (2014), which used the first week of founder behavior. All other features are calculated on the month the subreddit reached the specific activity threshold.

General. The general features contain the basic subreddit activity. This includes the number of accounts active in the threshold month and two features relating to the creation month: the number of unique accounts and the Gini coefficient of activity.

Promotion. The second group of features focuses on promotional content in other subreddits that mention or link to the subreddit. This includes the number of external comments with mentions or links, the number of submissions with mentions or links, and the number of cross-posts of content from the subreddit to others.

Fostering. Fostering focuses on longer-term place-making features, and is operationalized here as sustained content production. Seeding is a subset of content production in the first month. An author is considered a seeding author if they create more than 5 submissions in the first month, and this feature is the sum of the submissions from those seeding authors. I consider authors to be fostering authors if they created an average of 10 pieces of content per month for 3 or more consecutive months. I use average to capture some variation across their tenure, but they must create at least a single piece of content in each of 3 consecutive months. For these authors, I include the number of submissions and comments they created in the threshold month.

5 Fostering and Promotion during Evolution

Interaction. This group includes features about the social interaction within the subreddit. This includes the number of submissions that have at least a single comment or a score of greater than 1. I focused on positive interaction here because prior research has found responses and positive feedback was more likely to encourage future participation (Burke, Kraut, and Joyce, 2010; Joyce and Kraut, 2006). The second feature is the fraction of submissions that have comments in the threshold month.

Participation. The participation group includes features about the content creation equality and the amount of retention. The first is the equality of the amount of comments created by accounts in the subreddit. I also included the same equality feature for the number of submissions. Both of these features are calculated using the Gini coefficient (Gini, 1912) for the accounts creating the specific type of content in the threshold month. This was done using the following formula $\frac{2 \sum_{i=1}^n i y_i}{n \sum_{i=1}^n y_i} - \frac{n+1}{n}$ where n is the number of active accounts for that content, i is the rank of the that account relative to others, and y_i is the amount of content created by that account.

I also include a combined model which includes the features from all of the previously described features, which can be compared against the performance of each individual set of features.

5.2.4 RESULTS

For the results, I report area under the curve (AUC) – a single number measure of binary classifier performance commonly. The curve is the ROC curve (receiving operating characteristic) which measures the rate of true positives to the rate of false positives across a range of thresholds. AUC may correlate with accuracy in some conditions, but it is not a measure of accuracy.

The range of results is from 0 to 1, where 1 would be a perfect classifier and 0.5 means it performs at the same rate as a randomly chosen value³. Thresholds for acceptable AUC values vary depending on the context. In the case of this task, there are many potential influences on the outcomes so lower values may be acceptable in some circumstances.

³The effective range of the AUC is 0.5 to 1. Ranges below 0.5 are possible, but they represent situations where the classifier is mis-classifying more than expected by a purely random classifier. An AUC of 0 means the classifier perfectly misclassified all samples. For binary classifiers, the classifier could theoretically be negated such that the AUC would be in the range of 0.5 to 1. So an AUC of 0 could effectively be 1.0 by negating the chosen class.

5 Fostering and Promotion during Evolution

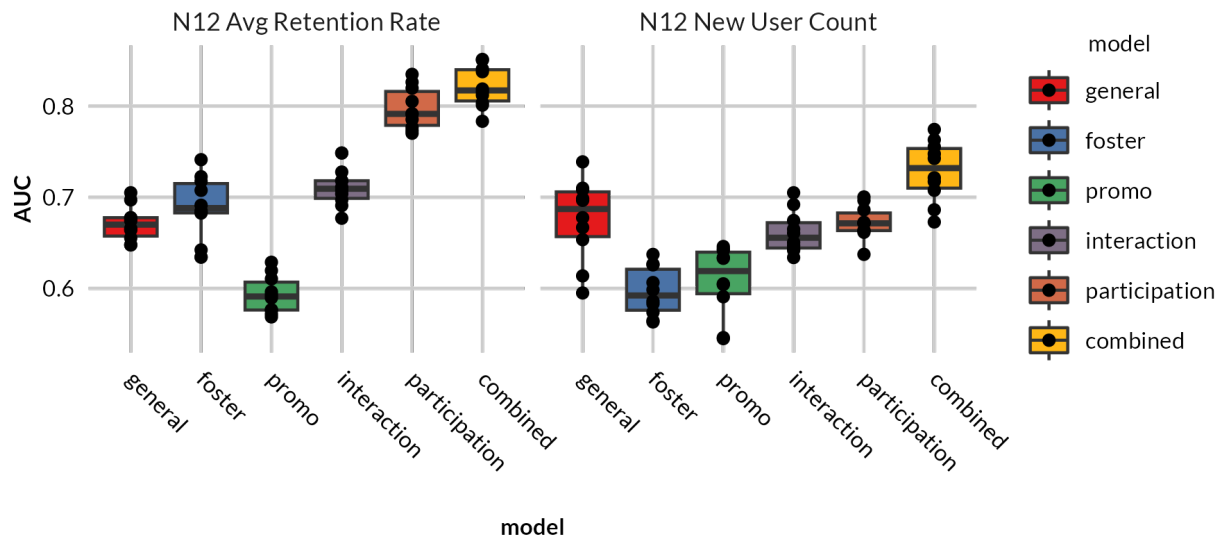


Figure 5.2: Graphs showing the AUC values for each feature group across the 10 activity thresholds for each of the success thresholds. On the left is the 12-month average retention rate and on the right is the number of new users in the 12 months following the activity threshold.

The AUC values for all activity thresholds are reported for each success metric and these can be seen in figure 5.2. In both cases, the combined models performed above 0.7 suggesting acceptable performance. The standard deviation of the AUC across all activity thresholds is low for all models, with little deviation between activity thresholds. The highest standard deviation is for the general features predicting the 12 month new accounts measure with 0.048. This suggests that the activity threshold at which these predictions are made may have little effect.

5.2.4.1 AVERAGE RETENTION RATE

On the left side of figure 5.2 we can see the combined classifier performed well at predicting the average retention rate (0.82). The worst performing classifier group was the promotion features with a median value of .59, which is likely uninformative. The highest performing feature group was participation (median=.79) which is likely due to the inclusion of retention in the feature group. It is possible that some subreddits may have relatively stable rates of participation and thus are correlated at longer time differences. Interaction had a median AUC of 0.72, which is also acceptable.

5 *Fostering and Promotion during Evolution*

Fostering here potentially informative results, with a median AUC of 0.67. The core fostering behavior (excluding seeding) is defined similar to retention, but tracks longer-term retention and more active users. Thus having a classification performance lower than participation is not surprising, especially given the features focus on the content produced by those users rather than the number of authors. These featuresets may have some marginal performance gains which are useful when added to the combined model.

5.2.4.2 NEW USER GROWTH

On the right of figure 5.2 we can see the performance of models for predicting new user growth for the 12 months following the activity threshold. The combined classifier performs reasonably well with a median AUC of 0.73. The group with general features is the second highest performing group with a median AUC of 0.69. The worst performing group is the fostering features with a median AUC of 0.58, but the remaining classifiers perform over 0.6. These other features may also have some marginal value for the combined classifier, but on their own are not good predictors of future author growth.

5.2.5 DISCUSSION

The data here provide some encouraging evidence that these featuresets may be useful in predicting retention and growth. While some individual featuresets have performance below 0.7 AUC, the combined models for both generally perform at acceptable levels (retention: 0.82, growth: 0.73) and better than the individual featuresets.

Growth is generally difficult to predict as many unknown events can happen between time the of prediction and the actual time of measurement. For online communities, some can see exponential growth when content reaches the front page or they suddenly die in a short period of time. In addition to events and behaviors within a community, there are also many external factors that may influence like events, platform changes, and behaviors outside the communities (e.g. promotion). Despite the difficulty of the task, the combined model generally had performance of $AUC \geq 0.7$ suggesting the combined

5 Fostering and Promotion during Evolution

featuresets may value in prediction. The best feature set for predicting this is the general set, while fostering performed the worst.

Somewhat surprisingly, the group of promotion features performed poorly in both tasks. The range of AUC scores across both tasks included some values below 0.6, suggesting promotion has less predictive value for these metrics. The highest AUC values the promotion features reached were for predicting the new users, but these were still low relative to other features. This is surprising given that promotion is intended to drive growth to a subreddit and prior work has found these can drive activity in the mentioned subreddit (Krohn and Weninger, 2022). One possible explanation might be that promotion (and possibly content fostering) are important early in the communities lifespan, but become less important later in its life. The paradigm I used here focuses on reaching growth thresholds, which could happen months later. This form of promotion also may attract few longer-term users. One final possibility regards the formulation of the feature. Perhaps, it is not the count of promotional content, but rather the size of the audience where the subreddits are promoted. For example, if people promote the subreddit only in smaller less active subreddits this may be less effective in attracting new users than promoting the subreddit in a larger, more active subreddit.

Both participation and interaction performed well for retention and somewhat well for predicting new users. The participation includes the percent of retentive authors, which may have little variation over time for successful communities. So failing communities likely have low retention, and that will continue to be low 12 months later. The same maybe be true for successful communities. Prior research has also found social interactions, especially positive ones, to be encouraging to members while creating the appearance of activity that can encourage further participation. However, both feature sets may not necessarily attract new members, thus yielding lower performance for new account growth.

The formulations of some of these features, especially fostering and seeding behavior could benefit from refined operationalization. For platform-holders with more data, there are other forms of fostering behavior that may also provide useful insights and possible predictive power such as actively moderating

content and communicating with people from the community as other studies have noted is common of moderators and creators.

I used the same activity thresholds that Cunha et al. (2019) used in their paper, but the performance may improve with changes to the paradigm. First using a higher maximum threshold and more space between thresholds may increase performance. In investigating the AUC scores over time, some of the results showed upward scoring trends as the communities grew (e.g. fostering for average retention rate). This may be a problem because a community might gain their 100th new participant in the first month, all ten of the snapshots will be the first month, providing little difference between the tasks. This could be a problem for promotion, where promotion behaviors may increase participation but without other behaviors like interaction and fostering may not yield longer-term retention. Additionally, the paradigm focuses on activity thresholds of new participant joins without consideration of the quality or tenure of that participation. An inherent assumption of the paradigm is that this threshold is related to the size of the community, but all of the new participants may have a tenure of a single day, meaning they may not actually be “members”.

5.3 PREDICTING REMAINING LIFETIME

Although promotion and fostering performed above random, they generally performed worse than the other two framework perspectives. One possibility may be that these matter differently at different points of time in a community, but the previous paradigm looked at lifespan as a function of member growth instead of time. This analysis addresses this with a different design and by looking at the community survival.

5.3.1 DATA & SURVIVAL ON REDDIT

Many subreddits, like other online communities, fail to survive over long periods of time. This is visible in the histogram of the 2012 subreddits shown in figure 5.1, where the death is considered 2 months of no comments. There were 86470 subreddits created in 2012 that had activity, though many had very little activity. A total of 16714 (19.3%) subreddits reach a minimum threshold of at least 25 comments and 10

unique users over the entire dataset (roughly 4-5 years). Only 3362 (3.9%) total subreddits “survived” more than 12 months. Fewer survived through the end of the dataset – 1923 of subreddits, which is 2.2% of the subreddits created in 2012. This study uses the predictors from the previous study to predict whether a community survives 24 relative months.

5.3.2 METHODS

Similar to the last analysis, I cast this as a prediction problem where I use features for promotion and fostering to predict whether the subreddit was still active 2 years later. The features for both groups are operationalized the same. I created a binary classifier using a random forest. This was implemented using the ranger package (Wright and Ziegler, 2017) in R. I used 10-fold cross-validation as implemented in the caret package (Kuhn, 2008).

I use features from the 2nd, and 6th month to predict the survival of the subreddit. Using two separate snapshots to predict survival allows me to explore the possibility that some features may vary with time. I also excluded subreddits which did not have any activity in the first month of their life to ensure a similar starting point. Similar snapshot prediction cases have been used in other works (Kairam, D. J. Wang, and Leskovec, 2012; Tsugawa and Niida, 2019). There were 8438 communities in the 2nd month snapshot and 4832 in the 6th month snapshot.

I operationalized the *death* condition to be 2 consecutive 30-day periods where no comments are created in the subreddit. Since short-lived communities could still be considered ‘successful’, I chose 24-‘months’ as the success threshold with those dying less than 24 months after the given month. This created an unbalanced dataset, and I used down-sampling to balance the classes leaving a total of 4972 and 4664 total subreddits in each⁴.

5.3.3 RESULTS

⁴The number of down-sampled communities for the 6th month may seem high relative to the first month, but 1923 subreddits which did not experience death making up half of that sample.

5 Fostering and Promotion during Evolution

The results of the three models for both time periods, excluding fostering in the 6th month, were generally acceptable as can be seen in table 5.2. This provides additional evidence of the value of both promotion and fostering in predicting the remaining lifetime of the subreddits. The highest AUC score was from the combined model which included all of the features.

In the 6th month, the features for fostering saw a decline in the performance from 0.69 to 0.65. The combined model saw a slight increase in performance. This may provide some weak evidence that fostering may have less predictive value of the remaining lifetime of a subreddit. However, the difference between these is small and may be due to randomness of the method employed. One of the strongest performing models was the control model, which included the number of unique authors and the number of authors in the first month.

Features	Month	
	2	6
control	0.77	0.81
interaction	0.78	0.80
participation	0.77	0.78
promotion	0.70	0.71
fostering	0.69	0.65
combined	0.81	0.83

Table 5.2: Results in the form of Area Under the ROC curve (AUC) for each group of features.

5.3.4 DISCUSSION

This findings provide additional evidence of the potential predictive value of both fostering and promotion in predicting the remaining lifetime of subreddits. The evidence here suggests that promotion may be important both early and later in the life of the subreddit, while fostering may be more predictive earlier in the lifespan of the subreddit. The high performance of the control group, which includes features like the number of unique authors, suggests that survival may best be predicted from features related to earlier activity. This may be a reiteration of the classic 'activity begets activity' finding from other research. It is important to note that this relationship may not be causal, and so these behaviors may simply be the products of successful communities rather than causing the subreddit to be successful.

5.4 WHO DOES THE WORK?

Another important question is: *who does the fostering and promotion for a subreddit?* I previously found that the work being done was a mix of normal users, moderators, and creators. However, this work wasn't

evenly spread across them. In several of the subreddits, the creator and moderators created a substantial amount of content at some point in the early lifetime. In this analysis, I explore the question of who does the work by looking at the different account types over the lives of thousands of subreddits.

5.4.1 DATA

I collected moderator and creator data from a total of 3575 total subreddits using archive.org's historical pages. I used their API to collect this data and I focused on subreddits created in 2012 with 400 or more submissions in their lifetime. I chose this threshold to reduce API requests and focus on communities with a some activity. I collected the main subreddit pages, which contain information about the creator and some moderators. I excluded subreddits where subreddit did not have at least one valid creator and moderator in the archives, because the one focus of this analysis is in attributing their work. This left a total of 3422 subreddits. This occurred on some subreddits where the creator or moderator name as '[deleted]', which is the text reddit uses when users delete their accounts⁵.

I collected data for each subreddit through 2016 and at most one archive per month to further limit API requests. This was mostly an issue for large subreddits which could have hundreds or thousands of archived pages per month.

Less active subreddits generally had fewer archived pages on archive.org, so the more active subreddits may be over-represented. Additionally, the tools archive.org uses for crawling and archiving pages may bias towards subreddits with more links that point to them, which may also skew which subreddits

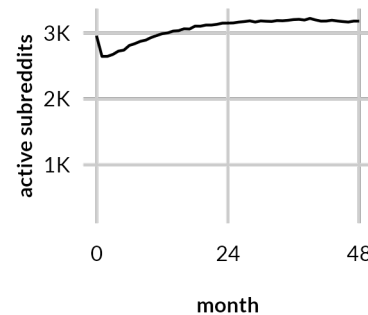


Figure 5.3: Number of subreddits with activity in each month after creation.

⁵And potentially for other actions that the platform could take. This is difficult to trace the inner workings of the platform over time and verify when and if this is true. There was, however, some hints suggesting this may also have occurred if reddit gave dead subreddits to new accounts and possibly removing the creator metadata.

are collected. The times between archived pages can be sporadic for less active subreddits, and thus while this data may be accurate, it may miss some accounts which were given moderator privileges and later removed between archival snapshots, thus under-counting the number of moderator accounts.

5.4.1.1 ACCOUNT TYPES

In this analysis, I look at three specific account types that vary by their level of potential power (Table 5.3). While the level of power and control they have in a subreddit may matter, these levels may also suggest a level of commitment to do some work for the subreddit. A creator was the account credited with creating the subreddit, and also has full moderator abilities at the point of creation. Other works refer to these types of accounts as founders (Foote, Gergle, and Shaw, 2017; Kairam and Foote, 2024; Kraut and Fiore, 2014). Moderators on reddit can have varying permissions to do maintenance tasks on reddit, but may be less likely to have the full power of the creator. I use the term *normal* accounts to refer to subreddit participants for which I have no evidence they were a moderator⁶.

Typically, I report summary data as two groups: power and normal. *Power* users includes both moderators and creators, while *normal* only contains normal accounts. I do this because both creators and moderators likely have some vested interest in maintaining or growing a subreddit. In some places, I report all three account types separately. I do this when I have found an interesting difference between moderators and creators, or when it enables me to place this work which have generally focused on founders and creators.

I considered an account to be a *power account* if they were the creator or a moderator at any point in the life of the subreddit. These analyses do not consider *when* an account became a moderator or was removed as a moderator. Though rare, there are a few cases where a creator may not be a moderator, including the person stepping down, being removed, and possibly deleting their accounts⁷. I consider

⁶I exclude communities where the creator could not be determined, so the only option is that an account may have been a short-term moderator.

⁷The latter case is uncertain, but a number of subreddits have the creator listed as '[deleted]' suggesting the account may be deleted, but I could not verify this. There are a number of alternate possibilities for this happening including bugs and the platform giving an unused or dead subreddit to a new person and setting the subreddit metadata to '[deleted]'.

5 Fostering and Promotion during Evolution

Account Type	# accounts	total activity (%)	submissions (%)	comments (%)
Normal	208,096 (97.5%)	80,383,839 (89%)	3,934,790 (67%)	76,449,049 (90.6%)
Power	5,411 (2.5%)	9,904,696 (11%)	1,940,177 (33%)	7,964,519 (9.5%)

Table 5.4: Descriptive statistics for fostering activity of power accounts (creator or moderator) versus other accounts.

accounts for which I could not verify their power status as a 'normal' account, and collectively those who were moderators or creators as 'power' users.

I use the same fostering and promotion operationalization and data from the earlier analyses. For promotion data, I excluded self-links and self mentions as these do not act as promotions in a way that may raise awareness of the subreddit for new users since only existing subscribers and visitors would see them.

I identified bots on reddit through the use an existing list (Fire and Guestrin, 2020), which was based on a list from the subreddit r/BotWatchman. I added an additional 48 accounts which brought the list to a total of 938 accounts. I found these additional bots from previous analyses. I considered accounts bots if they

posted a high number of submissions or comments and the content was highly similar. A number of these bots contained the word *bot*, like 'BitcoinAllBot' and 'TweetsInCommentsBot', but many others like 'Mentioned_Videos' and 'tweet_poster' do not. The 938 total accounts is almost certainly an undercount of bots on the platform.

Account Type	Definition
Creator	Created the subreddit. Creators have the same power as moderators, but like moderators they can be removed.
Moderator	This account was a moderator at some point in the subreddit's life, though the duration is uncertain. They can have varying levels of power as assigned by the creator or other moderators with specific permissions.
Normal	No evidence of being a moderator or creator of that subreddit.

Table 5.3: The three account types I consider from 3 different power levels.

5 Fostering and Promotion during Evolution

Account Type	months	activity	submissions	comments
Normal	4	137	4	128
Power	10	381	35	270.5

Table 5.5: Median statistics for fostering accounts.

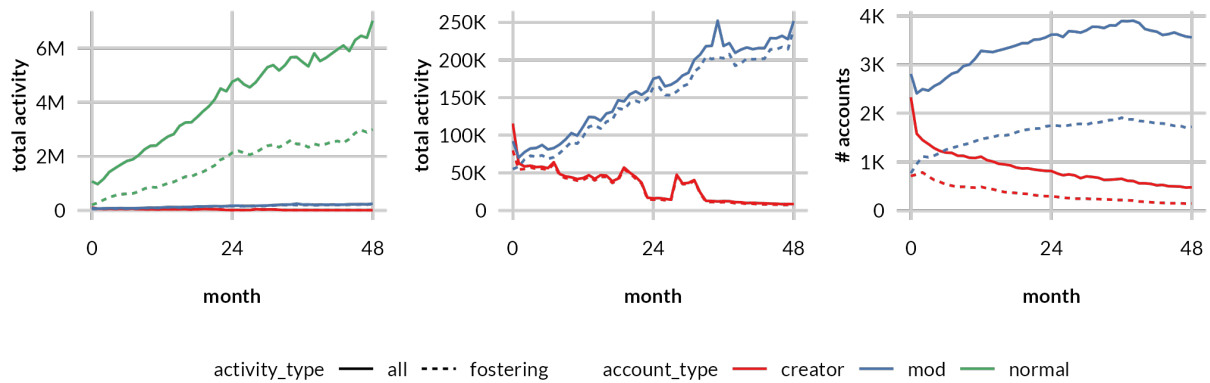


Figure 5.4: The left graph shows the total activity of normal accounts, moderators, and creators. The middle graph shows the same data, but excludes the normal accounts to illustrate the difference between creator and moderators. The right graph shows the number of active accounts over time for the account type groups. Creators became less active over time as can be seen in the middle and right graphs. Moderators and normal accounts increase in numbers over time and their activity increases as well.

5.4.2 FOSTERING

I found 212,122 total accounts that participated in fostering behavior in 2643 of the 3422 subreddits (77%). These accounts created a substantial portion of overall content ~37% across these subreddits though they were a relatively small number of total accounts across the subreddits.

The majority of fostered content was created by normal accounts. These non-moderator and non-creator accounts created 89% of the all fostered content. This work was split across 208,096 accounts (97.5%) versus the 5,411 power accounts as can be seen in table 5.4. This fostered activity is heavily skewed because the majority of fostered content created by normal accounts was comments (~77.8 million). The power accounts created a much larger share of the submissions (33%) compared to normal accounts (67%), despite the power accounts only making up 2.5% of all fostered accounts. This difference is also reflected in the median number of submissions created (Table 5.5): 35 submissions for power users versus 4 for normal. This is interesting because comments can only be placed on submissions, so the power accounts are disproportionately responsible for a significant amount of submissions and the comments on those submissions.

The median number of fostering months was higher for power accounts (15) than for normal accounts (6) as is visible in table 5.5. Power accounts were either the subreddit creator or a moderator at some point. The relationship between time and the amount of content is an artifact of how I operationalized fostering, but we can still see that power accounts were generally more involved in these subreddits.

A small portion of active fostering accounts were likely automated (bots). I identified 135 bots (0.06% of the accounts) which engaged in fostering. These bots contributed nearly 3.6 million pieces of content – 3.3% of all fostered content. This overall percent is low because of the large number of comments created by all fostering authors (84 million comments). However, bots accounted for 21% of all fostered submissions (1.2 million of 5.9 million total submissions). This is somewhat surprising given the relatively low percentage of fostering bots. However, a few bots that generated many of these submissions were reposting content from other subreddits, sometimes as a log (e.g. r/ModerationLog).

5 Fostering and Promotion during Evolution

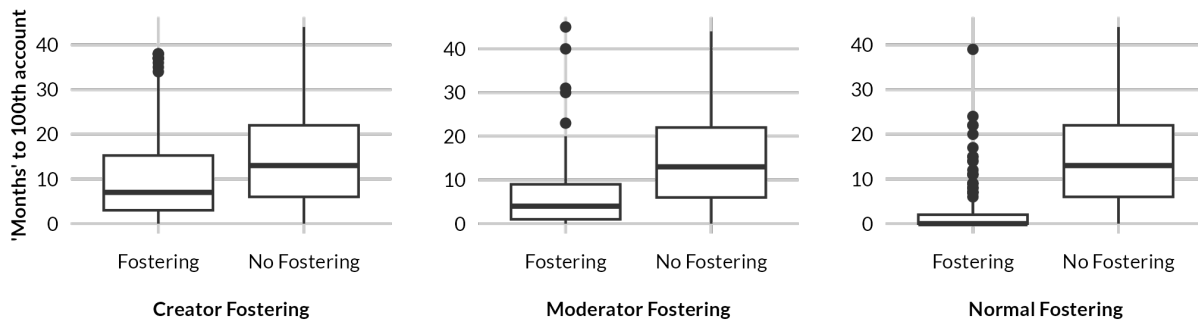


Figure 5.5: These graphs compare subreddits in which fostering occurred for only that specific group in the first month vs a group of subreddits where no fostering occurred by any group. From left to right these are creators, moderators, and normal accounts. There is a significant difference in each, suggesting fostering may be related to a faster period of new user growth.

The majority (108) of these bot accounts were normal accounts – non-moderator and non-creators in their respective subreddits. Only 35 accounts were power accounts with five being the creator of a subreddit including ModerationLog, a prolific account which attempted to record instances of removed posts across many popular subreddits (Some of its activity is visible as a spike after month 24 in the middle graph of figure 5.4). The most prolific bot account was AutoModerator, the programmable bot provided by reddit and employed in many of the subreddits.

Over the life of the subreddits, normal accounts consistently created more content (Figure 5.4), in part due to growth of new members. Creators accounted for a larger amount of fostered content than moderators in the first month, before being surpassed in the second and gradually becoming less active. By the 24th month, there were only 857 (24%) active creators and only 201 (6%) fostering creators. The decline, while interesting, is not comparable to moderators and normal accounts in these charts as the number of moderators and normal accounts is unbounded, but subreddits can only have a single creator.

Non-creators are doing a lot of content creation work throughout the lifetime of the subreddits. Creators were primarily active and creating content in the first month. However, in 28% of subreddits the creator was not active in the first month, and in 85% the creator wasn't engaged in fostering in the first

month. Even through the entire life of the subreddit, only 39% subreddits had creators that contributed fewer than 20 pieces of content.

Looking at the time it takes a subreddit to reach the 100th active account, we can see a different distribution between subreddits which had fostering in the first month and those that didn't. The nth unique account threshold was the same paradigm used in the first prediction analysis of this chapter (5.2). To compare groups, I have focused on subreddits that reached 100 unique accounts and which had fostering done by a specific account type in the first month and compared against subreddits where no fostering occurred in the first month.

Boxplots can be seen in figure 5.5. These plots isolate subreddits to compare where fostering occurred versus those where no fostering occurred. The median number of months to reach 100 accounts for the no-fostering group is 12 (n=2414). The median for creators was 6 (n=283), moderators was 2 (n=97), and normal accounts was 0 (n=98). A Wilcoxon-Mann-Whitney test⁸ test found a significant difference in all three comparisons (Creators: $U=250419$, $p<0.001$; Moderators: $U=41621$, $p<0.001$; Normal: $U=26710$, $p<0.001$). This suggests that communities with early fostering may be more likely to experience early growth compared to those without it. This should be interpreted with caution as this relationship may not be causal. It is possible other factors may confound this relationship and be affecting both. For example, fostering may be the product of more experienced moderators, who may use fostering in combination with other behaviors to accelerate early growth.

5.4.3 PROMOTION

The majority of subreddits in this sample were promoted in other subreddits. I found 3409 of the 3422 subreddits had at least one link, mention, or cross-post in another subreddit. There was a total of 893,675 total accounts that contributed to promotion of these eligible subreddits. The counts used in this analysis

⁸This test is also called the Wilcoxon rank sum test or the Mann-Whitney test. The test was first proposed by Deuchler in 1914, later by Wilcoxon in 1945, and later by Mann Whitney (1947). The test is often proposed as a non-parametric version of the t-test, though it does not compare a measure of central tendency. It does not assume equal sample sizes, though power may be lost as the imbalance increases.

5 Fostering and Promotion during Evolution

Account Type	# accounts	# comments (%)	submissions (%)
Normal	891033	3,824,683 (49.3%)	710,964 (76.5%)
Moderator	7897	3,805,452 (49.0%)	206,452 (22.2%)
Creator	2255	134,626 (1.7%)	11,558 (1.2%)

Table 5.6: Descriptive statistics for promotion behavior across the subreddits. Power-holding accounts were a small portion of all promoting accounts, but created a substantial portion of comment and submission promotional content.

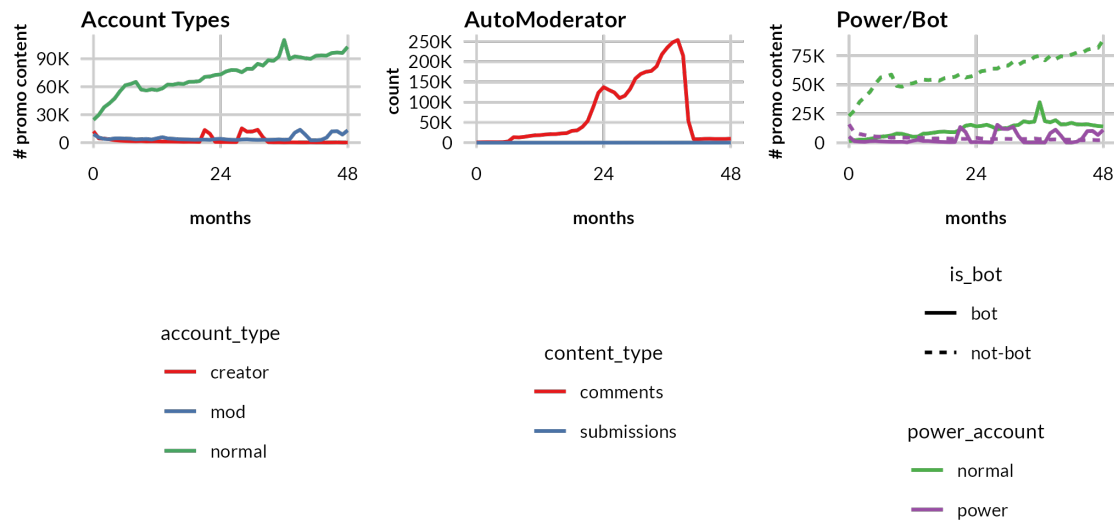


Figure 5.6: Three graphs showing the promotional content over time. The left graph shows by account types of different power levels and excluding all AutoModerator content. The majority of the content is created by normal accounts. The several spikes for moderators and creators are bots. The middle graph shows all AutoModerator content that was excluded from the left-most graph, the majority of which are comments. The gradual increase in promotions by AutoModerator is due to its early adoption after it was created in January 2012, though a majority of these promotions include a link to the AutoModerator subreddit for feedback. The sudden decline appears to be due to the removal of a link to its subreddit (r/AutoModerator) from its default comment text. The right-most graph shows promotional content broken down by power and by bot status.

are focused on unique comments or submissions (including cross-posts) as text fields enable multiple mentions and links to occur for the same piece of content.

Power accounts created a large percentage of promotional content (Table 5.6), despite being a small number of accounts. They made up nearly $\sim 10k$ or 1% of promoting accounts, but accounted for 50% of comments and $\sim 24\%$ of submissions with promotions. The vast majority of that work was done by moderators. The difference in the number of submissions and comments was partly due the number of bots in this sample, and how those bots work. Many bots on reddit create a lot of content as comments, like u/autotldr’s summaries of articles submitted to different subreddits or u/AutoModerator’s feedback when a submission was removed.

Bot accounts created a majority of overall promotional content. In total, 253 (0.03%) of these accounts were identified as bots and they produced 53% of all of the promotional content for these subreddits. Unsurprisingly, the most active bot account here was u/AutoModerator, which is a moderator tool created in 2012 automate many common or repetitive tasks for moderators (u/Deimorz, 2012a; u/Deimorz, 2012b) and made an official tool in 2015 (Basu, 2020). This increase in AutoModerator promotional content visible in figure 5.6 is due in part to it’s early adoption. It created 3.6 million pieces of promotional content (41% of all promotional content), and 99% of that content was comments. The next most prevalent bot was u/autotldr, which is an article summarization bot, and it created 356k comments and submissions. Here, u/AutoModerator skewed these results because it is used across subreddits. Although it is configured to operate differently in each subreddit, it appears as a singular account. Thus it makes promotional content created in subreddit J look as if it was posted by the same moderator account from subreddit K , despite being configured and operated very differently. I exclude AutoModerator in the left-most graph in figure 5.6, and include a separate graph of their overall promotional content (middle).

Some of the active bots had subreddits dedicated to their operation, a trend that has continued despite the growing increase of AutoModerator (Jhaver, Birman, et al., 2019). Some of this promotional work was in the from of mentions and links back to their subreddit in order to learn more or provide feedback. Both u/AutoModerator and u/autotldr engaged in this behavior to provide information about the bot or

solicit feedback. Some bots like u/TotesMessenger, left a comment in a thread when someone linked to that thread in a different subreddit, thus bridging two subreddits while also linking to its own subreddit, r/TotesMessenger.

Creators became less active promoting their subreddit over time (see Figure 5.6). Excluding some bot activity like u/AutoModerator, moderators typically did more of the promotional work than creators after the first month, with a few exceptions for very active, but short-lived bots that were also subreddit creators (e.g. u/ModerationLog in r/ModerationLog). As the subreddits age, normal accounts do an increasing amount of this work, too.

In the middle graph of figure 5.6, you can see a gradual adoption of u/AutoModerator, which included a link and mention to its own subreddit for the first few years of its life. Excluding AutoModerator, bots created less overall content than non-bots, excluding a few spikes of heavy bot activity among power accounts (Figure 5.6; right graph). These spikes were the result of a few creator or moderator bots being very active in a particular time period, with r/ModerationLog being the largest contributor for each spike.

Power accounts (blue line in right-most graph) generally decreased promotional content over time, while normal accounts increased their promotional content. However, even in the first month, normal accounts still created more overall promotions than power accounts.

As the right-most graph in figure 5.6, illustrates, not all bots were power accounts for the subreddits they promoted. In fact, for non-bot, normal accounts created a large portion of promotional content for subreddits (though the right-most graph excludes u/AutoModerator).

5.5 DISCUSSION

These analyses provide evidence that framework could be useful for explaining and predicting different forms of online subreddit or community success. Each of the perspectives differed in their performance of predicting different measures of success, which aligns with earlier work (Cunha et al., 2019). Communities can serve many different purposes (TeBlunthuis, Kiene, et al., 2022) and can have different measures of success (Aumayr and Hayes, 2017; Foote, Gergle, and Shaw, 2017; Matthews et al., 2014). The framework

perspectives may be useful for measuring and providing feedback to the community for achieving different goals they have. The predictive value of the framework may be useful for platforms to design and build features for improving the chance of communities reaching their goals.

The fostering features generally had lower AUC scores than the other framework perspectives. This is surprising given that it is a commonly suggested behavior both on reddit and in scholarly literature. Prior work suggests that seeding and early creator activity has been useful for predicting different forms of community success (Cunha et al., 2019; Kraut and Fiore, 2014), and these align with my definition of fostering. One possible explanation may be around the quality of the fostered or seeded content. Spam and other off-topic content may negatively affect the community and its perceived value (Nonnecke and Preece, 2001; TeBlunthuis, Kiene, et al., 2022; Weld, A. X. Zhang, and Althoff, 2022). Some existing work has also found that very active individuals may not encourage interaction (Schoberth, Preece, and Heinzl, 2003) and even creators doing a lot of work could reduce the odds of survival (Kraut and Fiore, 2014).

Fostering could also benefit from a different operationalization. Two possible adjustments to the operationalization of fostering may yield better results. First, including additional work that moderators and creators do beyond simple content creation. This has been the topic of prior work, suggesting moderators and creators do a significant amount of work beyond content creation (Li, Hecht, and Chancellor, 2022; Lo, 2018; Seering and Kairam, 2023; Wohn, 2019). Second, the definitions for seeding and fostering have lower bounds of activities but not higher bounds.

The promotional perspective is important because a large portion of prior work focuses on behaviors that occur within the community, rather than the work that occurs outside of it, which is important for raising awareness of the community and attracting new members. On reddit, these forms of promotion can drive activity to the subreddit (Krohn and Weninger, 2022). Founders can leverage their network of friends on some platforms to grow a network, but this can lead to smaller groups (Kairam, D. J. Wang, and Leskovec, 2012) or negatively affect survival if the founder takes on all responsibilities (Kraut and Fiore, 2014). Platforms like reddit are less focused on building explicit social networks, with the common connections being through interactions like voting and replying. Reddit acknowledges this challenge and provides

a few suggestions in its help, but also cautions against the overuse of those tactics. By focusing more on the work of promotion, especially outside the community and designing to support this work, this may further help many communities achieve the success metrics they target.

In this work, I also did not consider the subreddit's content type or purpose. It is possible that the specific community content/purpose may moderate the predictive performance of the feature groups I have leveraged. However, categorizing subreddits into a clean typology may be difficult since the previous chapter and other work suggest that many communities may support multiple types/purposes (e.g. Prinster et al., [2024](#)).

The last analysis found that a significant portion of the work was done by normal accounts. A sizable portion of subreddits that were successful had little to no creator activity at all. Some prior work has focused on the creator as being a key element of a subreddit, which is a reasonable starting point for building a community. Creators were very active in these results, especially in the first month before declining. This is a minor mystery that may be worth exploring in future work. Do the creators burn out and stop doing work in those subreddits? Do they simply move onto other types of work for those subreddits? Prior work on moderators and creators suggest the workload of other tasks may increase as the community grows, and they may spend more time doing those tasks once a community is established.

However, these findings also suggest that anyone may be able to build a community within a space, regardless of their power status. What may be most important is just that someone does this fostering work. I found that 28% of creators were not active in their subreddit in the first month, and a substantial portion contributed very little content during the life of the subreddit. This opens new opportunities for tools to be designed to support community-building for all types of accounts. For example, a platform might build tools to recommend content to share or cross-post. With appropriate safeguards, They could also make suggestions to help anyone promote a subreddit. By making tools available to any account, regardless of their power status, it is possible once “dead” communities can be revived, and to prevent others from dying when the creators stop doing work.

Future work here is important to better understand the relationships especially for fostering and seeding, especially with different forms of operationalization and looking at the behaviors beyond content creation. These behaviors have been suggested behavior by some scholars (Preece, 2000) and reddit's own documentation (Reddit, 2023c). If this behavior is not valuable or can negatively affect communities, then documentation should be updated to reflect this. Additionally, these systems' design can also be updated to warn creators and moderators if this is in fact a problem.

5.6 LIMITATIONS

An important caveat with these results is that they are not casual. Though this work has looked at predictive performance of these features, it is possible that many of these behaviors may not cause or even contribute to the success metric. They may instead be a product of other behaviors that drive success. For example, promotional behaviors may be a product of a thriving or growing community. This is also true for fostering behaviors, which is more conservative form of retention. I will leave the exploration of the causal questions open for future work.

This study has focused on exploring the predictive value of four of the framework perspectives on different success measures. Prior work has noted that that the Pushshift dataset is missing a high number of data (Gaffney and Matias, 2018), which may skew some of the results in unknown ways. In many cases, the missing data cannot be retrieved and thus it is not possible to tell which subreddits or authors they came from, which are critical aspects of this research. A related issue is the content where the author has been deleted, and where it is often impossible to resolve to the original author. The missing and deleted data may not be evenly distributed, thus possibly affecting periods of time more than others for which could heavily skew the results of this data. It may also affect some subreddits more than others. Some calculations used in these analyses needed to ignore the deleted author content, as it creates artificially skewed data where it appears a single account is responsible for all of the content. For example, this would often make '[deleted]' appear as a fostering author in many subreddits, despite being potentially many different accounts.

Another limitation is that this dataset has a lot of spamming and bot activity. I chose to leave this data in these analyses because in Chapter 4, we saw some subreddits use bots to automate submitting content to their subreddits (eg. r/LinusTechTips and the YouTube videos), and there was an increase in activity after this point. It is also equally possible that bots or spammers can negatively affect these analyses, possibly in ways which change all of the relationships. Future work should attempt to control for this.

5.7 CONCLUSION

In this chapter, I explored the ability of the framework to predict three measures of success. The four perspectives had different levels of performance for different success metrics. This suggests that the framework has some value in predicting at least three different success metrics, and potentially others.

The interaction and participation features had acceptable performance for predicting 12-month retention rate. For 12-month user growth, all framework features generally had AUC scores between 0.6 and 0.7, with fostering having the lowest median AUC. Fostering and promotion had lower scores generally, though fostering had a median AUC of near 0.7 for retention rate.

Fostering behaviors had the lowest median AUC scores for two of the success metrics. This may be in part due to a poor operationalization of these behaviors. The formula used here focuses only on content at arbitrary thresholds. However, other factors may be important such as the quality of the content, which has been known to be important to many members and visitors. It is also possible that other forms of fostering may be important such as moderator work or upvoting/replying to new participants. Both of these hinge on people actually becoming aware of the subreddit in some way.

Promotion had acceptable performance for predicting the survival of a subreddit, but had the lowest AUC scores for predicting retention and second lowest for user growth. While the performance of these features is low, it may still be useful in combination with other features in predicting the measures of success. This does provide some evidence that work done outside the community may be useful for

5 Fostering and Promotion during Evolution

predicting measures of success, but this does not mean the relationship is causal. It is possible that promotions are the product of activity and success, rather than useful in attracting engaged users.

I also looked at who performed some of the work of fostering and promotion. The core finding here was that normal accounts contributed a significant portion of both fostered content and promotional content through the life of these communities. Creators on the other hand were most active for both fostering and promotion in the first month, before becoming less active as the communities aged. Prior research has recognized the importance of encouraging participation, but some, including reddit's own help documents, assume that some behaviors would only be done by power users. This finding forefronts some of the work that everyday members can do, and more research here is needed to better understand which tasks could be shared among common members, especially when creators and moderators do not do this work.

The ecosystem of content creation within these subreddits also included a number of bots. I found very few bots that were engaged in fostering, though they did produce a small, but out-sized portion of fostered content within those subreddits. For promotion, AutoModerator produced a significant amount of promotional content. Even with AutoModerator excluded from analyses, bots still produced a substantial amount of promotional content, and a portion of this was to their subreddit for feedback.

6 PROMOTIONS AND FOSTERING DURING A POLITICAL CAMPAIGN

6.1 INTRODUCTION

In this chapter, I leverage the framework to explore fostering and promotions of a subreddit during the 2016 US presidential election. Political subreddits can focus on political campaigns and evolving real-world events. These communities can be viewed as forms of political activism through their advocacy work both on and off the platform (Gaudiano, 2015; Mills, 2018), enabling them to function like a collective action organization and a community. Thus, political subreddits potentially differ from the previously studied subreddits. I focus on the temporal growth of new participants within the subreddit, and explore questions about recruitment and growth from the perspective of collective action theory.

My site of inquiry for this analysis is the subreddit r/The_Donald (r/TD), which grew quickly during the 2016 US Presidential election to eventually become one of the most active subreddits in 2016. It became infamous for its trolling, memes, and hate-filled content, which eventually got the subreddit banned from the platform (Peck, 2020). The subreddit was seemingly composed of many different types of actors including a number of fans as well as trolls, which made moderation difficult. The subreddit had manipulated the front page by flooding it with many pro Trump stories (Lagorio-Chafkin, 2016). It was also a source of inter-subreddit conflicts (Datta and Adar, 2019), often referred to as *brigading*. The subreddit was also connected to other online communities like 4chan and 8chan as well as being allegedly

connected to the Trump campaign¹. Some of the memes, slogans, and trolling started on 4chan or 8chan and were mainstreamed in r/The_Donald (Marwick and Lewis, 2017; *World War Meme* 2017). Trump himself had retweeted some of the content that had previously circulated on the subreddit (Herrman, 2016; Koebler, 2016). Despite only by about 4% of Americans (Barthel et al., 2016) using the reddit at the time, r/TD may have played some small role in the larger 2016 election, as the subreddit was the topic of a number of news stories.

The subreddit displayed many of the properties of an online community, including a culture and language it developed. However, I do not talk to participants in the space and ask them if they feel like it's a community or if they feel a sense of attachment to the space. Thus, I will use the word subreddit to refer to the collective group of people who eventually participated. When I use community, it is in a more generalized sense.

I define political action as a form of collective action where the goal involves a governmental, organizational, or institutional change, such as legislative changes or electing candidates to bring change. It can also include shifting a party or groups political stance on a topic. In campaign-related subreddits, one goal is promoting their candidate on reddit and beyond (Mills, 2018). Like social movements, the members of the subreddit worked together to advocate for their candidate. They spread pro-Trump content through sharing and voting, which potentially contributed to it's quick growth of participants.

Theories of collective action broadly focus on mobilization of people and resources for a set of common goals (Klandermans and Oegema, 1987; McAdam, 1986; Snow, Zurcher, and Ekland-Olson, 1980; Tilly, 1978). Some early lines of research focused on organizations, such as labor unions, where members were connected in a network through the organization. These works focused on participation as the potential participants were typically existing organization members and were aware of the grievances and the movement (Tilly, 1978). However, when movements are not closely connected to existing organizations some form of recruitment needs to happen. Recruitment requires awareness of the action, and then

¹ Hope Hicks, a spokesperson for the campaign denies the campaign was involved with reddit (Herrman, 2016), though other staffers claim otherwise (*World War Meme* 2017). It was connected enough to score an AMA with Trump in July 2016.

having an interest and motivation to participate. Efforts to raise awareness of the movement, even when unintentional, can be seen as recruitment.

Once aware of the action and goals, interested people can then become participants, though the number of activated participants can be low. Even when people are aligned with the movement's goals, they may lack motivation to actually participate. Participation may also be a spectrum where some people participate more than others. For example, in a study of one protest, though 74% of people were aligned with the movement only 4% actually attended the protest (Klandermans and Oegema, 1987). In the Internet Age, people may be easily motivated to share content online but not show up to a protest. Having a personal relationship with an existing participant may encourage participation. Many religious organizations have recruited through their personal networks (Snow, Zurcher, and Ekland-Olson, 1980). In addition to personal connections, a prior track record of action may also predict participation in a movement, especially in high-cost and high-risk forms of action (McAdam, 1986).

A common form of recruitment in many early leverages personal and organizational networks. Other methods of recruitment, like mass-media was rarely reported as a successful means of recruiting (Klandermans and Oegema, 1987; Snow, Zurcher, and Ekland-Olson, 1980). A few, such as the Hare Krishna, utilized non-network recruiting, but still involved face-to-face recruitment (Snow, Zurcher, and Ekland-Olson, 1980). These can be difficult because of the time and effort involved to share information and eventually encourage a person to become an active participant.

Technological advancements have potentially reduced the cost for some forms of participation and recruiting, which have changed some elements of social movements. There has also been a fundamental shift in the nature of some traditional political organizations from a in-person volunteering and direct action to a more fundraising and advocacy (Putnam, 2001). Social media has enabled people to broadcast content to large, world-wide audiences. This has enabled protests occurring entirely online as well as hybrid online-offline protests. Recruiting and mobilization attempts through social media can be influenced by information diffusion practices in addition to network effects (González-Bailón et al., 2011). Local movements can suddenly work with people who are not physically present (e.g. Bennett, Segerberg,

and Walker, 2014; Starbird and Palen, 2011). Many forms of collective action now extend beyond local areas to potentially bring in new resources from around the world in the form of work and service contributions, resources, and financial aid.

Viewing r/TD as a form of collective action, I explore the evolution of the subreddit in terms of the account growth, with a special focus fostering accounts. The fast growth of r/TD may have been due to efficient recruitment strategies like bloc recruitment through existing communities (Oberschall, 1973). If this were true we should see evidence that promotions in politically sympathetic communities would result in large numbers of people joining r/TD. Thus, I seek to answer the following research questions:

1. How did the subreddit grow engaged accounts?
2. How are promotions related to the growth of new accounts?
3. Where did the engaged accounts come from?

Together this works to advance both our understanding of the growth of online communities and collective action, in the context of political campaign communities. This additionally leverages a part of the framework in a new form of community and illustrates the value of using a framework for a focused inquiry.

I found some evidence that promotions may precede an increase in new members who joined the subreddit, but this may not be causal. Instead, many promotions and joins appear bursty following real-world news events connected to the campaign. I also found evidence that many fostering authors that joined had previously participated in conservative subreddits, suggesting that the growth may be similar to bloc recruitment present in some forms of collective action.

6.2 DATA

I focus on a subset of the pushshift dataset (Baumgartner et al., 2020) consisting of all submissions and comments made in r/The_Donald between its creation and December 31st, 2016. It was the largest of the 3 political subreddits dedicated to major candidates from the 2016 election, and one of the most active on

6 Promotions and Fostering During a Political Campaign

Subreddit	Creation date (UTC)	Total Activity	# Submissions	# Comments	Unique Authors
hillaryclinton	2012-05-04	1428026	84131	1343895	46656
SandersForPresident	2013-12-05	4611216	234035	4377181	209017
The_Donald	2015-06-27	13077869	1425592	11652277	333563

Table 6.1: Summary of the data for the three political subreddits.

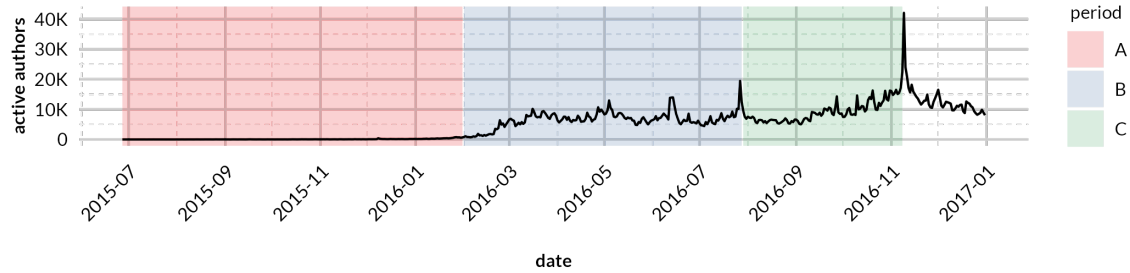


Figure 6.1: Overview of active authors in r/The_Donald in terms of unique accounts active on a daily basis between June 2015 and December 2016. The graph is divided into 3 time periods: (A) the creation and early life, (B) primary elections, and (C) run-up to the election. The subreddit sees an early growth in daily active users starting around the primary elections, before receding slightly until just before Trump won the election.

all of reddit at the time. Descriptive statistics for the content and users of these three subreddits can be seen in table ???. To answer questions about growth, I focus on accounts that become newly active. The dataset does not contain subscription metadata so I cannot know if or when an account is subscribed to a subreddit, so instead I rely on participation within the subreddit. I consider an account to have 'joined' when they make their first submission or comment.

I divided the timeline of r/The_Donald into 3 general time periods, which I bound by specific events (Figure 6.1). The first period (A) is from the r/TD's creation to the beginning of the primary elections (January 31st). The second period (B) focuses on the primaries and ranges from February 2nd to July, 31st a few days past the end of the Democratic National Convection. The third period was the run-up to the election (November 8th). There is an additional period after the election, but it is not considered in this analysis.

6 Promotions and Fostering During a Political Campaign

Subreddit	Accounts (%)	Activity (%)	Submissions (%)	Comments (%)
hillaryclinton	2,404 (5.15%)	915,693 (64.12%)	49,108 (58.37%)	866,585 (64.48%)
SandersForPresident	11,260 (5.38%)	2,622,184 (56.86%)	94,717 (40.47%)	2,527,467 (57.74%)
The_Donald	24,135 (7.23%)	8,351,760 (63.86%)	770,246 (54.02%)	7,581,514 (65.06%)

Table 6.2: Descriptive statistics for fostering activity in the three political subreddits.

This study is focused on two dimensions from the framework: *fostering* and *promotional*. I define a fostering account as one that is active in a subreddit for 3 or more consecutive months, and create at least 5 pieces of content in every consecutive month. This is a higher threshold than conventional retention metrics, which generally require any activity in two or more consecutive months. In the two previous chapters, fostering accounts made up a small percentage of total participants but they collectively generated a significant amount of content in many subreddits. While more casual participants could be part of a collective capacity, I focus on these core users because of their continued participation.

I considered promotional content including mentions, links, and cross-posts that reference or point to the subreddit. Links and mentions are URLs found in the body of submissions and comments in other subreddits that form hyperlinks to r/TD. Cross-posts are submission links that are shared directly in another subreddit, which appear similar to the native submission (So a cross-posted image submission appears as an image.). In the metadata, this occurs when a submission in another subreddit is not a text submission but is a link to a submission in the destination subreddit. I also included reddit shortened urls (redd.it/<id>) for the submissions. These were all identified using regular expressions in a database query.

6.3 CHARACTERIZING R/THE_DONALD

I begin by providing context for r/TD during the 2016 election and its lead-up by leveraging the framework to compare the subreddit to other subreddits. During the 2016 election cycle, there were three notable political subreddits that followed the three major candidates: r/SandersForPresident (r/SFP), r/hillaryclinton (r/hc), and r/The_Donald (r/TD). r/TD grew to be a large subreddit accumulating a high number of active accounts, though appears similar to the other political subreddits across other perspectives.

r/TD grew to be one of the most active subreddits on reddit, despite having been created in June 2015. r/SFP was created in 2013 and r/hc in 2012, but both saw the majority of their activity from the beginning of 2015 onwards. In 2016, r/TD was the 3rd highest producer of content across all subreddits and 24th for the number of unique accounts that were active. In contrast, r/SFP was ranked 32nd and r/hc 81st in total content created in each. This level of growth from a new subreddit to one of the most active subreddits is surprising, especially given that many of the most active subreddits were created years before r/TD.

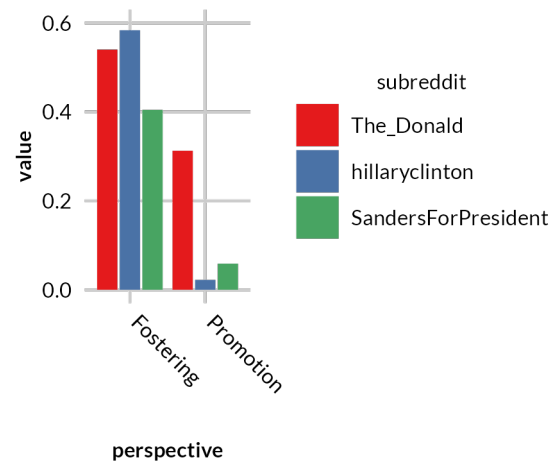


Figure 6.2: A comparison of fostered submissions and promotions. r/TD and r/hc have a majority of submissions created by fostering authors, while r/S4P has only about 40% of submissions created by fostering authors. r/TD received 31% inbound promotions relative to the highest number of inbound promotions across all subreddits.

Comparison of fostering submissions and external promotions reveal additional differences between the three political subreddits (Figure 6.2). Fostering accounts were a minority of the active accounts in all three subreddits but produced an outsized amount of the content (Table 6.2). Across all content (submissions and comments), fostering accounts created more than half of all subreddit content, despite consisting of only 5-7% of total accounts.

Focusing on fostered submissions, we can see that r/TD and r/hc had fostering accounts produce similar amounts of submissions (54% and 58% respectively) while fostering accounts in r/SFP produced only 40% (Figure 6.2). Submissions are important both structurally and behaviorally as they are structurally necessary for comments to occur, and are a pathway to visibility on the front page and encouraging further engagement with the subreddit for people who rely on the front page. One way to view this is as an inequality in the work of creating these submissions, and is captured in the Gini coefficient (Gini, 1912) which ranges from 0 to 1, where 0 is equal and 1 is complete inequality. r/hc and r/TD have a much higher level of this inequality with a relatively small percentage of people creating a majority of the submissions

and Gini for submissions 0.77 and 0.81 respectively. r/SFP has a lower level of inequality, but even at 40% is still a high level of inequality as seen from a Gini coefficient of 0.63.

A large difference was also observed in the number of on-platform promotions the subreddit received. In figure 6.2, we can see r/TD received more inbound promotions than the other two subreddits with 31% of the highest subreddit. The numbers in this graph are relative to the maximum value of inbound promotions across all subreddits (excluding r/AutoModerator²). Comparatively, r/hc had only 2% and r/SFP only 6%. The difference in the number of promotions between r/TD and r/SFP is 5 times larger, despite having only 1.6 times more active users and 2.8 times more content than SFP.

In the previous chapter, I found some supporting evidence for the idea that promotions and fostering may have value in predicting future growth. One problem with this is that the relationship between these may not be causal, and in fact the actual cause could be reversed. For example, the number of promotions may be a product of the size of the community. Since promotions are one potential way in which new members may discover a community, I explore the possibility that promotions precede new fostering authors joining and becoming active.

6.4 WHEN AND HOW DID R/TD GROW?

I first explored when and how the subreddit grew participating accounts. The participating accounts serve as a potential resource pool from which the subreddit collective used to accomplish some of their advocacy. Although anyone can contribute to collective goals and tasks, many online events often have people who are involved for longer periods, some of those which organize (Dailey and Starbird, 2014; Leavitt and J. A. Clark, 2014; White and Palen, 2015, e.g.). Fostering accounts are one way to focus on a higher level of engagement. I explore the growth of these engaged users over time.

²The highest number of promotions was for r/AutoModerator which I excluded. These were inflated by self-mentions and self-links in a large number of comments and submissions the bot created, which inflated the number to 3,279,703. Its number of promotions were three times higher than those of the next highest subreddit, r/conspiracy (1,069,547), which was used as the denominator to normalize.

6 Promotions and Fostering During a Political Campaign

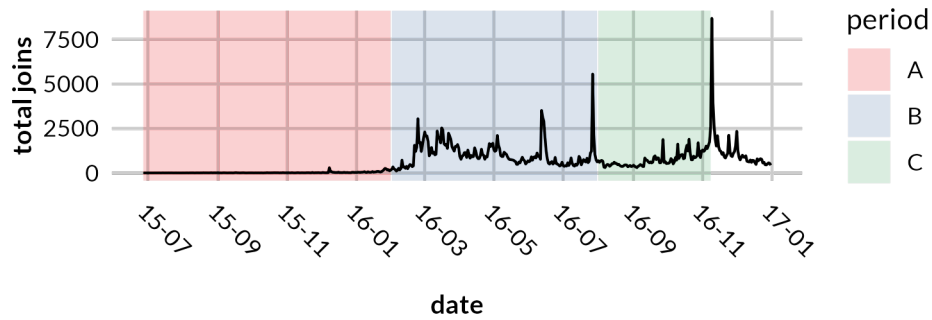


Figure 6.3: The daily number of new joins over time in r/The_Donald. The subreddit saw an increase in the number of daily new joins during February of 2016, after Trump won some early primaries. The trend after that was a decline through September, when an upward trend occurred to the election.

The number of new daily joins for all accounts varied through different phases of the campaign as seen in figure 6.3. The mean number of joins in Period A was 21, with a slight upward trend starting in December 2015 after Trump called for a ban on Muslims entering the US. There was a large spike in daily joins in mid February after Trump began winning early Republican primaries. At this point, Trump potentially became a viable candidate for many people on reddit, even for those that did not support him. This increase lasted through the entire period B with a mean of 1042 new daily joins, despite a general downward trend after the bump. The final period begins in a lull before an upward trend from mid-September through the election (mean= 742 new joins).

There are a number of visible bursts of increased joins (“spikes”) in the timeline. Thematic analysis of submissions during those time periods reveal that many of the largest spikes were related to real-world campaign events or news, both pro-Trump and anti-Hillary or anti-Sanders. For example, there were a few spikes in February 2016 after Trump won primary elections of individual states. One clear exception here was the Trump AMA in July 2016. The Pulse Nightclub shootings in June 2016 was a mix of external and internal events (relative to r/TD on reddit), where members of r/TD accused moderators in r/news of censorship for deleting posts identifying the perpetrator as a Muslim, leading r/TD to becoming a hot-spot for discussion of the topic which contributed over half of the 25 stories on r/all (Ohlheiser, 2016;

Peck, 2020). This may have been partially due to the moderators sticky-ing submissions in the subreddit to “boost” the number of votes some submissions received (Peck, 2020).

One possibility here is that the growth of r/TD was at least in part influenced by the ongoing campaign, likely resulting in the noticeable jump in daily joins in February, 2016. This may have lead to more people seeing Trump as a viable candidate and began participating in the subreddit. However, some events like the Trump AMA in r/TD and Pulse nightclub drama suggest that on-platform activity can also be a source of new account joins. However, of the events I analyzed, the majority of new joins were of accounts who less active than fostering accounts.

This leads to questions about how this growth developed. First, how many of these accounts remained active over time? Although the group can benefit from individual contributions, an extended amount of contributions may be needed to sustain the overall activity, especially with the declining trend in joins after February. Second, did promotions help people become aware of r/TD? Early in its life, r/TD’s name would not have been obvious to many people wanting to discuss Trump. Search is one possible solution, but so are r/all content and promotions.

6.5 JOINS BY LEVEL OF TENURE ENGAGEMENT

Were there patterns of joins based on the activity level of the account in r/TD? In this analysis, I look at different tenures and when they joined over time.

6.5.1 DATA

For this analysis, I employ descriptive statistics and induction to help answer the questions of join patterns by level of engagement. I combined both tenure and activity level into three groups: single-day, multi-day, and fostering. The single-day accounts were active less than 24 hours, the multi-day were active over 24 hours or more but did not meet fostering criteria, and fostering were accounts that created 5 or more pieces of content per month for 3 or more months. I considered an account as fostering if it had met these criteria at any point in its tenure.

6 Promotions and Fostering During a Political Campaign

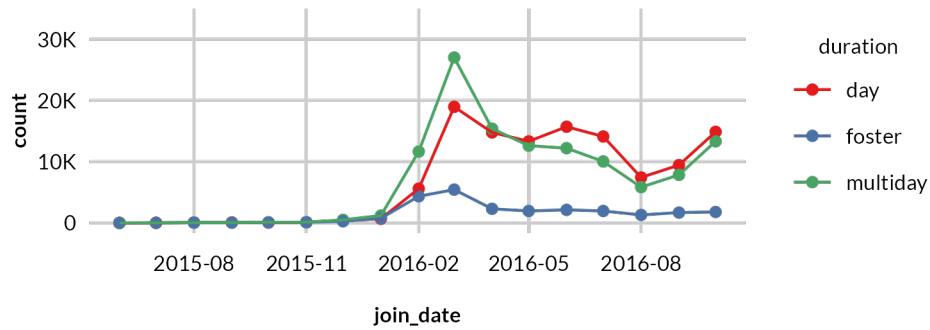


Figure 6.4: The monthly number of joins of accounts that are active for a single day, multiple days, or who fostered at least one period in their tenure. Of note, that the earlier life seemed to attract more fostering and multi-day accounts and fewer single-day accounts, but this reverses in May 2016. The method of operationalizing fostering has a 3-month time component, so I exclude the last 2 months of 2016 to avoid questions about why fostering appears to disappear.

6.5.2 RESULTS

There were two interesting patterns visible in figure 6.4 worth discussing. First, was the previously described sustained growth following the early primary wins. Second, there is a point in May 2016 where the number of joins of single-day participants surpasses the number of more engaged accounts. I must also point out that in figure 6.4, the drop-off in fostering accounts in the final two months is a product of how fostering is operationalized. I have included these two months on the graph only to include the election.

A large number of the fostering accounts joined early in 2016 after early Trump primary wins. A total of 46% of all cumulative fostering accounts joined as of April 1st, 2016, and this total crossed the 50% threshold on April 15th, 2016. The peak was in March where 5440 fostering accounts first became active (23% of all eventual fostering accounts). The multi-day accounts similarly peaked in March with 27029 accounts joining. Both of these groups began a downward trend after March, as the total number of joins declines through August³.

³It's also worth reiterating that the operationalization for groups have a time component and the dataset ends in 2016, meaning that some of the single day accounts that joined could participate in the subreddit beyond the dataset length and change these patterns. For the fostering data, I found that across fostering accounts in r/TD, over 50% began that behavior in their first month and 75% had begun by the end of the 2nd month. Excluding some surprise event that somehow spurred many accounts to participate again in the subreddit, the fostering counts through August may change very little. The multi-day

In May, there was a shift in joins towards less-engaged accounts that were active for only a single day. This trend continues through the end of the dataset. A similar trend was experienced in both r/hc and r/SFP. One possible explanation is that these are visitors that come from a stream of submissions that reached r/all. This could potentially attract a greater ratio of less engaged accounts (single day) in the form of accounts which are not subscribed to the subreddit and simply reply to individual submissions via comment.

This suggests that external events may have been a catalyst for some major growth events in r/TD, including growth of more engaged users. However, there is still a question of how these new members discovered the subreddit from an external event. It is possible that people were inspired by the wins and searched on reddit for a subreddit dedicated to Trump's campaign, but the subreddit's name is not the most obvious name. Other forms of awareness on reddit include explicit promotions and the visibility provided by reaching the front of r/all, which the moderators of r/TD appear to have hacked in order to get more visibility on r/all (Peck, 2020) ⁴. The latter is difficult to clearly track retroactively, and it is possible that many smaller subreddits may not be able to generate enough activity to reach the front page. Instead, I explore the viability of promotions being a potential cause of the growth.

6.6 ARE PROMOTIONS RELATED TO ACCOUNT JOINS?

My next question was about the relationship between promotions and account joins. I found evidence that promotions have some predictive power of future growth in the previous chapter, but the analysis was not causal. The promotions could have been the product of continued growth rather than a causal source of growth. In this analysis, I explored the temporal relationship in more depth. First, I looked for a correlation between the number of promotion events and the number of joins through August 31, 2016. I then explored the temporal relationship between the two by looking at joins before and after individual

operationalization is more sensitive as any single-day account can become multi-day by creating a single piece of content at any point later, though it may take rare events to radically change the ratios.

⁴This article notes this around June of 2016, but some reddit posts suggest this may have started around April 2016 (e.g. u/controversialideas, 2016; u/dwaxe, 2016)

6 Promotions and Fostering During a Political Campaign

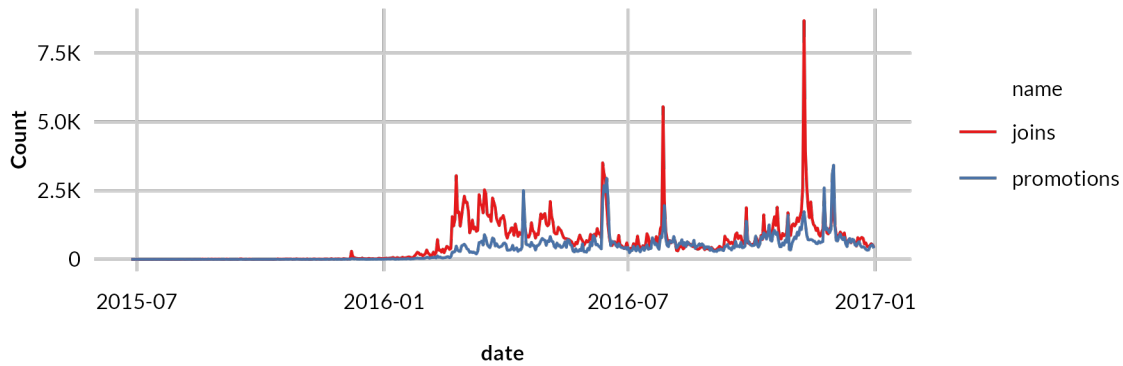


Figure 6.5: The number of promotions and joins per day. For the early primary wins, the number of joins are much higher than the number of promotions, but visually follow a similar pattern. After the early primary wins, the number of joins and promotions appear to show a slight correlation.

promotions. This is important because a casual relationship implies a temporal order, so if promotions do lead to an increase in joins then we should see a change in the number of joins after a promotion.

6.6.1 METHODS

To explore the temporal relationship between new accounts participating in the subreddit (“joins”) and promotions in other subreddits, I rely on descriptive statistics, cross-correlation, and Wilcoxon signed-rank tests of pre- and post-event groups.

To do the cross-correlation, I first tested for unit roots, which suggested trends exists, and removed the trends on both time series before running the cross-correlation. This was done because trends can inflate the correlation value between two time series.

The promotions were identified with regular expressions as previously described. Joins in this section are the first piece of content the accounts created.

6.6.2 RESULTS

Figure 6.5 shows the daily number of joins and the promotions. Visually, for a large time period in 2016, the number of joins appears to be roughly similar to the number of promotions, with a clear exception being the early part of 2016. The period of the early primaries through June has a greater number of daily

6 Promotions and Fostering During a Political Campaign

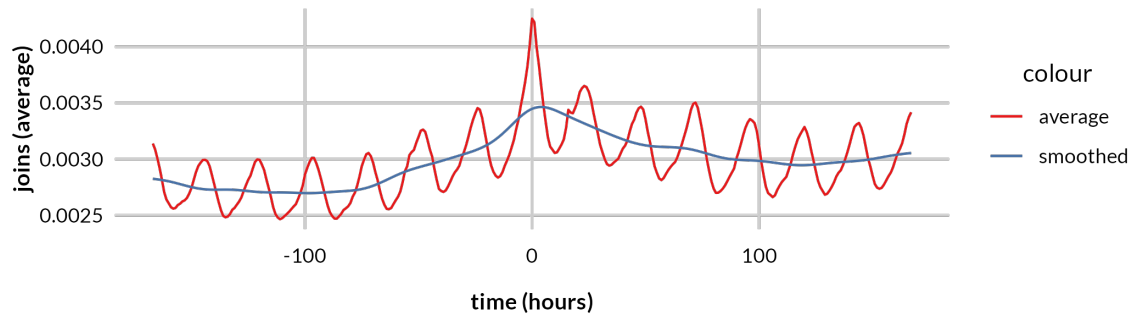


Figure 6.6: The average hourly number of joins relative to the time of a promotions within one week (168 hours) showing a slight increase in the number of joins after a promotion (time=zero). The graph contains both the raw average and a smoothed version. The raw average has a constant daily pattern driven by the large number of North American users on the site.

joins than promotions, though the pattern generally seems similar. From June onward, the number of joins and promotions appear similar excluding a few spikes, including the large one in June after the Pulse Night Club shooting and r/news drama I described earlier.

I first explored this temporal relationship through a cross-correlation. Auto-correlation analysis suggested a trend in both time series, which I removed by taking the first difference (lag=1) of both time series. Both the Augmented Dickey-Fuller test and the KPSS test suggest this yielded a stationary time series (ADF: $p < 0.01$ and KPSS $p > 0.1$). The cross-correlation of the de-trended timeseries had peak correlation coefficient of 0.35 on for the series with no lag ($t=0$). This is a relatively low level of correlation, which suggests that there may be other things affecting the relationship between the two. Additionally, the peak correlation with a lag of zero, suggests that the temporal relationship may be occur rather quickly – in less than 24 hours. If promotions was partially causal of some joins and the rate was slower, we might expect a lag of greater than zero, which is not the case.

Another way to test the temporal effect is to look at the hourly number of joins relative to a promotion. This provides an hour level of granularity across 7 days relative to the event (168 total hours before and after) to detect a temporal relationship. This employs a similar method to Krohn and Weninger (2022), who explored the relationship between promotions and activity. Similar to that work, I treat each promotional piece of content as an independent event. This includes links, mentions, and cross-posts. In total, there

were 111,963 unique promotions made toward r/TD through August 31, 2016. This date cutoff provides a large number of promotions, while still providing a larger window for accounts to become multi-day active or fostering accounts. The average number of joins by hour is plotted in figure 6.6, along with a smoothed value. The wave pattern is the result of daily cycles driven by the high number of active users in North America.

Both lines here suggest that there may be an increase in accounts at the promotion and the hours following it, possibly lasting beyond a single day. The median pre-promotion value was 1876 and the median after was 2029 and a paired Wilcoxon signed-rank test on there is a significant difference between the ranks of pre- and post-promotion groups ($V=2383096256$, $p<0.01$). This suggests that promotions of all types may precede a higher number of joins, but this is important to note that this alone is not necessarily indicative of a causal relationship.

This pattern of increased joins after the promotion is seen across all levels of tenure. So this includes accounts active for a single day, multiple days, and accounts that eventually engage in fostering behavior. Wilcoxon rank-sum tests for all three groups yields significant differences ($p<0.01$ for each group), with the median higher in the after group than before, suggesting that promotions may precede many of the newly joining accounts. This suggests that there could be a causal effect for promotions to joins, but this alone does not exclude the possibility of other factors that may cause increases in both, such as external media.

Visual inspection of the graph (6.6) also shows an increase in the mean number of joins in the hours leading up to the promotion. One possible explanation here is that the events are not truly independent, as an increase in activity on r/TD can lead to content reaching the front page, which can lead to more promotions, and more activity (a cascade). This means that joins may have been increasing because of an earlier promotion that happened in a similar time window. Given the infamy of the r/TD, even external events relating to Trump's campaign could have yielded both promotions and increased joins. Subreddits like r/EnoughTrumpSpam or r/EnoughSandersSpam reacted, sometimes with promotional content, to content taking over front pages.

These results show with more granularity a possible temporal relationship between promotions and joins. The effect here does appear to happen for joining accounts of any activity level (from single day accounts to fostering accounts). This provides some evidence that they may lead people to join a subreddit after being exposed, though it may be difficult using these methods to attribute a join to any specific promotion. The one known possible confounder here is the front page ranking algorithms, which affects what they can see.

These results should be interpreted with caution. These data are not truly independent as we often see spikes of both promotions and joins near similar dates, which may lead to the increase in joins seen prior to the “event” in figure 6.6. This is not a full causal analysis, but rather one that find some supporting evidence of promotions preceding an increase in joins. I also cannot control for the effect of front-page submissions. Additionally, this was performed on a single subreddit, and although the number of promotions explored is large, it may be an outlier in how and when people join.

6.7 WHERE WERE NEW ACCOUNTS ACTIVE PRIOR TO JOINING R/TD?

Oberschall (1973) hypothesized that quick growth in collective action may be the result of “bloc recruitment”, wherein already connected groups with similar interests may join an ongoing effort. Many reddit users tend to gradually participate in new subreddits over time (Tan and L. Lee, 2015). We can consider an accounts past participation in subreddits as a form of weak membership. If bloc recruitment was utilized on reddit, we would expect a surge of joins from members who had previously participated in subreddits that would be sympathetic to Trump’s message. In this case, they could be conservative subreddits or those sharing concerns voiced by the campaign.

Prior activity across subreddits has been used it to understand the lineage of new subreddits (Tan, 2018) and mobilized attacks (Datta and Adar, 2019). People can post across similar subreddits, and the overlap of members has enabled interesting forms of discovery of similar subreddits (Martin, 2017).

6.7.1 METHODS

For the prior history, I focused my analysis on the accounts that were active in the campaign subreddits. I chose 30 days to focus on the most recent activity since interests and participation may shift over time. The data included all comments and submissions and related metadata. Using this data, I generated descriptive statistics and the accounts prior activity. I then perform an inductive analysis across the statistics and visualizations.

I combined this with a list of default subreddits. The default subreddits are ones the platform auto-subscribed new accounts to when they joined the platform. They include a variety of broad subreddits as a means of providing new users with interesting content like r/science, r/AskReddit, r/funny, and r/worldnews. The list was constructed using reddit blog posts, reddit submissions which detail the changes, and archive.org pages to attempt to verify. For the analyses using default data, I included all default subreddits including some that had been removed over the years, as some like r/politics had become large enough to have content in r/popular regularly.

6.7.2 RESULTS

Most of the fostering accounts were active across many subreddits before joining the political subreddit, including a number of default subreddits. In the 30 days prior to joining, the median number of subreddits an account participated in before joining r/TD was 22, and both r/SFP and r/hc had a median of 35. These median values are surprising, suggesting that accounts participating in r/TD may be less active across reddit accounts participating in r/SFP and r/hc. Additionally, a majority of accounts in all of the political subreddits had participated in a default subreddit prior to joining – 87% of all r/TD joins and 94% for both r/SFP and r/hc. Surprisingly, a smaller percentage of r/TD fostering authors had previously participated in default subreddits (76%) than in r/hc (87%) or in r/SFP (89%). The default subreddits are some of the most active subreddits on the platform, and they account for most of the top 20 subreddits authors were active in before joining. However, these tell us very little about the differences between the subreddits.

6 Promotions and Fostering During a Political Campaign

rank	The_Donald		hillarlyclinton		SandersForPresident	
	subreddit	count	subreddit	count	subreddit	count
1	SandersForPresident	2653	SandersForPresident	566	pcmasterrace	1010
2	pcmasterrace	1648	PoliticalDiscussion	549	nfl	783
3	TumblrInAction	1424	nfl	241	BlackPeopleTwitter	717
4	BlackPeopleTwitter	1418	SubredditDrama	206	trees	683
5	4chan	1354	enoughsandersspam	157	interestingasfuck	616
6	nfl	1241	nba	153	TumblrInAction	571
7	KotakuInAction	1181	The_Donald	148	nba	535
8	ImGoingToHellForThis	1065	EnoughTrumpSpam	129	woahdude	503
9	conspiracy	1055	BlackPeopleTwitter	99	Android	480
10	Conservative	1002	changemyview	99	reactiongifs	480
11	CringeAnarchy	954	baseball	98	StarWars	470
12	interestingasfuck	814	gameofthrones	93	cringepics	456
13	nba	787	me_irl	91	conspiracy	439
14	trees	757	circlebroke	90	Games	435
15	reactiongifs	731	CFB	88	leagueoflegends	413
16	cringepics	700	relationships	87	relationships	410
17	leagueoflegends	684	OutOfTheLoop	84	4chan	393
18	StarWars	656	StarWars	84	SubredditDrama	364
19	GlobalOffensive	630	NeutralPolitics	83	ImGoingToHellForThis	363
20	cringe	624	democrats	78	OutOfTheLoop	347

Table 6.3: The top 20 subreddits where fostering authors previously participated.

6 Promotions and Fostering During a Political Campaign

Excluding the default subreddits, I found the accounts also participated in a number of political and politically adjacent subreddits in addition to some non-default popular subreddits (Table 6.3). All three of the campaign subreddits had fostering authors who participated in popular subreddits on reddit like sports and gaming subreddits (e.g. nfl, nba, CFB, pcmasterrace, leagueoflegends, etc.). Other common subreddits appear among the histories including r/StarWars and r/BlackPeopleTwitter, the latter being a subreddit for sharing funny social media content from Black people.

The histories differ when we look at the political subreddits the accounts participated in prior to joining the campaign subreddits. This includes accounts that previously participated in one of the 3 campaign subreddits. The highest number of fostering authors in both r/TD and r/hc had previously participated in r/SFP. Similarly, r/hc also had a number of fostering authors who were previously active in r/TD. In the case of the r/hc and r/TD accounts that were previously active in r/SFP, these may have occurred after the campaign became less viable which is also when moderators locked the sub to encourage people to move to r/hc.

The histories also contained other political and cultural subreddits aligning to their respective political views. The fostering authors in r/TD had previously participated in r/Conservative while r/hc had some that had participated in r/democrats and r/NeutralPolitics. There was also some similarity between r/TD and r/SFP prior histories, where they share a number of similar subreddits including r/4chan, r/conspiracy, r/TumblrInAction, r/cringepics,

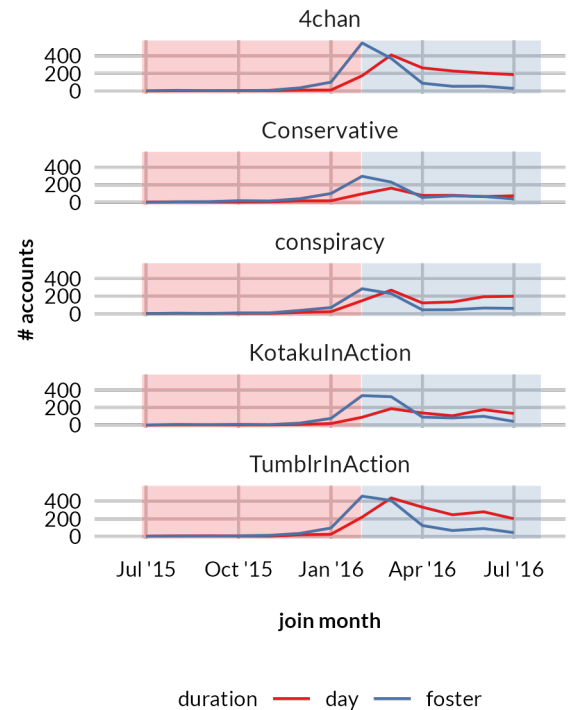


Figure 6.7: Counts of authors who participated in conservative subreddits prior to joining r/TD. The number of fostering joins begin rising prior to February when Trump won early primaries. The number of fostering joins from these subreddits exceed the joins for authors active only a single day early on, before eventually being overtaken later on.

r/ImGoingToHellForThis, and a few of the previously mentioned subreddits. The similarity between r/SFP and r/TD is surprising given that the campaigns were on opposite ends of the political spectrum. These might be attempts to recruit members or perhaps sharing negative news about Hillary Clinton.

Before the primaries, r/TD saw an increase in joins from some of the conservative subreddits as well as a few known for their anti-social behaviors. This harassment and some were known for being misogynistic, anti-LGBTQ, and anti-political-correctness (anti-pc). For example, r/KotakuInAction formed from members of GamerGate (Jhaver, Chan, and Bruckman, 2018) which took its name from r/TumblrInAction, and both were sources of harassment (Massanari, 2017). A number of these have been identified as sources of conflict on reddit including r/KIA, r/TD, r/conspiracy, r/CringeAnarchy, and r/ImGoingToHellForThis (Datta and Adar, 2019)⁵. In general, these were among a handful of subreddits that saw an elevated number of joins before February, 2016, though most peaked in February with the first primary wins. A total of 59% of the 258 fostering accounts that joined in December 2015 had participated in one of these subreddits and 56% of the 760 in January 2016. This suggests that some of the more-engaged accounts may have been previously sympathetic to Trump's message and joined early, and is evidence that a process similar to bloc recruitment may have occurred. This fell to 45% and 32% in March and April.

There were also a surprising percentage of accounts in r/TD that had no prior history before posting in r/TD (even beyond the 30-day window). In total, 43,966 accounts (13%) made their first comment or submission in r/TD. This was more twice the percentage of r/SFP and r/hc (6%). A majority (55%) of these accounts were active less than 24 hours. About 10% of the accounts that first posted in TD were first active in the Trump AMA on July 27th. Surprisingly, about 4360 (18%) of the accounts with no prior reddit activity history became fostering accounts in r/TD. Some of these are likely throwaway or alt accounts, intended to keep their activity in r/TD separate from their main profiles⁶, a practice that is not uncommon

⁵This paper also found r/SFP to be a source of conflict, which may align with these findings that r/SFP was the common prior history subreddit of r/TD fostering authors, but further analysis of the type of conflict is required.

⁶One possible reason for doing this was that a number of subreddits would auto-ban people who had participated in r/TD in an attempt to minimize trolling and other bad behavior.

on reddit (Leavitt, 2015; Wohn, Yuan, and Siri Jr, 2021). Some subreddits would automatically ban people who participated in subreddits like r/TD, so these accounts may have been alts to protect their main accounts.

6.8 DISCUSSION

Although this study was exploratory, there are a number of interesting findings about the evolution of r/TD and the campaign associated subreddits. The growth of the subreddits appears to be attached to the real-world events like the primary elections and news events. This is visible by the increased growth of r/TD in the first few months of 2016. This intuitively makes sense, given that the subreddits were aligned to follow the campaigns, but this deviates from some of the early discourse around online communities being distinct “spaces” as separate from the real-world. There were some internal events such as the platform discourse around censorship in the wake of the Pulse Night Club shooting and the Trump AMA, but this did not radically increase the number of fostering authors in the same way as the early primaries.

Relatedly, I observed that the a majority of overall fostering authors had joined by March, with a significant portion joining in at the beginning of the primary elections. This provides further evidence of the alignment with real-world events. Interestingly, the rise in joins of single-day only accounts such that it eventually overpasses fostering and multi-day is similarly surprising. This could be a product of scale, where these subreddits had reached a size that they were able to make it to the front of r/popular, thus leading to less-engaged accounts to participate only when submissions reached the front of r/popular. In the case of r/The_Donald, they had leveraged the subreddits activity to game the platform algorithm and score many front page articles by pinning submissions, a use case the platform had not intended. This allowed the subreddit to keep submissions visible longer on the front page of the subreddit, thus getting more votes than normal. However, with the dataset limited to only 2016 it is possible that these findings could change, with some accounts participating and becoming fostering authors after the dataset ended.

I also found evidence of a possible temporal relationship between promotions and new members joining. This furthers existing research around potential methods for growing communities. Previous research had found that these promotions could increase activity (Krohn and Weninger, 2022). While some

of the new activity or joins could be coordinated “attacks” from other subreddits (aka brigading on the platform), I specifically looked at the joins of fostering authors. Although I found some evidence of a possible temporal relationship with more joins typically occurring after promotions, this is not causal and needs more research. Some events lead to spikes in promotions which may mean promotions can cluster such that some may lag behind early joins which could lead to data similar to those observed here. More research is needed here to better understand this relationship and to further explore a causal effect.

In looking at the prior history of fostering accounts in r/TD, I found that many of the accounts had participated in subreddits that were conservative or potentially sympathetic to the causes. The growth from these sympathetic subreddits suggest that bloc recruitment may be a viable strategy employed in multi-community spaces. Reddit may be somewhat unique here compared to other platforms where personal networks are important. The high number of potentially sympathetic subreddits in users past history also aligns with the lineage theory of subreddits (Tan, 2018).

Lastly, perhaps the most surprising finding here was the prior participation in another campaign subreddit from that year. A number of participants from r/SFP had become fostering authors in r/hc and r/TD, which is surprising. Similarly, the prior participation in subreddits between r/TD and r/SFP fostering authors appeared similar, which suggest at least some surprising alignment between the two. A possible explanation might have been negative Clinton news and content, as this was a common opponent for both campaigns. However, this is an opportunity for future work to further explore.

6.8.1 LIMITATIONS

There were a number of limitations with this study that may affect the way the results are interpreted. Although I made some effort to control for some of these or to limit their effects, these controls may not have been sufficient in all cases.

The dataset used in this work extended only through the end of 2016 which presents a critical limitation for elements of this work that rely on the duration of participation within the subreddit. I limited several of these analyses to focus on the period through August 2016, based on analysis of fostering authors from the 2012 subreddits. There I found a large majority (> 80%) of accounts that became fostering

had done so within 4 months of their first participation in the subreddit. It is possible that these subreddits differ from those subreddits, and so these results could change with a longer dataset.

The join date is derived from the first piece of content created, which may not be the date a person subscribed or first participated. People can subscribe and vote prior in advance of actually writing a comment or submission, thus the effect of joins on some days could be underestimated (and potentially overestimated⁷)

This work also provides evidence that point towards a possible temporal relationship between promotions and growth, but this was not a causal analysis. This is presented as an exploratory case study, and although there are a large number of observations, these observations are not independent and appear to be related to external events.

Bots are also a problem for potentially inflating promotions and joins, and I did not fully control for this beyond ignoring promotions from u/AutoModerator. Similarly, the size of the audience where promotions occur could strongly influence their effectiveness. For example, a single post in a large subreddit could drive many more participants than 100 promotions in tiny subreddit. Future work should explore the effects of audience size on promotions and the potential effects of bots in this context.

6.8.2 CONCLUSION

In this chapter, I performed an exploratory case study on r/The_Donald with some comparisons to two of other political campaign subreddits from the 2016 US presidential election. I had a number of findings which help illustrate that both promotions and fostering may be valuable concepts to explore the evolution of other online communities to better understand how they grow. This also supports the idea of bloc theory as an eligible form of recruitment in the online space.

I found some evidence of a temporal relationship between promotions and fostering joins that suggest that the promotions may precede joins, suggesting that they could drive some amount of joins.

⁷This might be the case for some submissions or events being a catalyst for a large number of lurkers to start writing comments.

The results also suggest that promotions may be more complicated than a simple causal effect, and that some promotions may be reactionary to events or be made after a number of joins have occurred.

I additionally found that the pattern of joins in the subreddit aligned with campaign events, and most notably the primary elections. This suggests that these subreddits are connected to the real world, by way of experiencing some activity and growth of new users around events. This is not too surprising, as a body of research have found that many online forms of action such as protests and crisis events have been connected with the online activity (e.g. Dailey and Starbird, 2014; Leavitt and J. A. Clark, 2014; Starbird and Palen, 2011). In many of these cases, as events in the real world have changed, the online counterparts have also adjusted. In case of r/TD and the political subreddits, it is the building of a community and its overall capacity that reacted to changes in the campaign by adding new highly-engaged members. This has potential implications for a number of forms of online communities and fandoms, and while some research explores changes in these communities, more is needed.

Lastly, this work also added evidence of the value of history as a way to understand a community. In multi-community spaces like reddit, where members can participate in many subreddits over time, these can be used to understand and characterize these spaces. Tan (2018) first explored this on reddit as a means of understanding how communities can evolve from other communities that existed prior. Other research has looked these histories as a means of finding similarities between communities, but more work is needed in this area. I found some evidence that the percentage of people joining from five subreddits that were sympathetic were decreasing over time, and it is possible that these communities themselves may shift. Like this work, many analyses look only holistically at the complete set of members, potentially missing more nuanced temporal shifts that occur which may help understand the evolution.

Together this chapter has provided interesting insights into the evolution of political subreddits. However, there is still a significant amount of research that can be done. Some existing research and popular media articles have looked at the possible effects the subreddit may have had in influencing other online spaces, and possibly the campaign, but this is clearly an important space where more work needs to be done.

7 CONTRIBUTIONS & DISCUSSION

This work has explored the formation and early development of online communities. This work extends the body of work on online communities by providing a framework for analyzing and comparing communities, like subreddits, for the window of time between creation through the development of a community. This work also provides analysis of two specific behaviors that have less empirical data: fostering and promotion.

I want to provide a brief summary of each of the studies presented in this work along with the contributions I believe they make. In the following sections, I will discuss some of the key findings and the possible implications that I believe these could have. I will then conclude by suggesting some additional directions for future work in this space.

7.1 A FRAMEWORK FOR CHARACTERIZING ONLINE COMMUNITIES

In Chapter 4, I explored the evolution of some diverse online communities to understand some similarities and differences between them. These included a number individual behaviors and community level characteristics. Some of these, like the seeking behaviors, existed throughout the duration of the data I analyzed while others, like seeding content, were largely focused at the beginning. I grouped these attributes and behaviors into five themes (perspectives) which may function as a framework for analyzing online communities. This framework may be most valuable platform designers looking to support the development of new communities, and to create affordances for communities to better meet their desired goals.

7 Contributions & Discussion

The five perspectives were interaction, content, participation, fostering, and promotion. Interaction focuses on the social interactions occurring within the subreddit. Content characterizes on the type of the content shared within the community. Participation characterizes how people participate within the subreddit, including concepts like equality and tenure. Fostering focused on a number of intentional community building behaviors which were intended to develop or maintain the community. And lastly, promotion referred to methods for raising awareness of the community beyond its boundaries.

Together, these perspectives provide a holistic way of approaching analyses and understanding the evolution of communities. These were grounded in the observed behaviors from the first analysis. These behaviors and similar characterizations were grouped together to form the perspectives describing higher level themes.

This framework differs from some of those referenced in the background in a number of ways. First, because it is developed from a grounded analysis, it is built on observed behaviors rather than theory or synthesis of other work. This also means it's tailored to the specific data, and may need additional iterations to help it apply more broadly. Each of the perspectives was developed through behaviors I observed in the initial analysis, though in follow-up analyses, I apply some of the methods from other work for operationalizing specific constructs. However, it is worth noting that I was aware of the research before some of these analyses, which may have influenced how I built and named the themes. Second, it is aimed to be effective in the early life of online communities, even before they have developed into communities. The specific inclusion of fostering and promotion make it ideal for understanding the evolution on reddit, which is not true of other frameworks.

A few other works did capture elements of interaction and participation (e.g. Prinster et al., [2024](#)). Some of the ideas for promotion and fostering can be found in elements through other works (e.g. Kraut and P. Resnick, [2012](#); Preece, [2000](#)), though this was not part of their frameworks. Based on existing research (e.g. X. Wang, B. S. Butler, and Ren, [2013](#)), I chose to include content type and purpose as a selection criteria. However, I also observed distinct differences between the types of communities, and so found it a useful

addition to the framework. I broadened its scope slightly, such that it could be operationalized through methods used in other works.

A related contribution was a taxonomy for the forms of questions I found within the subreddits. This short taxonomy included questions for conversation starting, information-seeking, resource-seeking, and advice-seeking. I observed these across subreddits, and these illustrate some of the ways people consider communities as a resource. These also detail different forms of help or support they might receive from that community. There is some overlap with some existing question taxonomies (e.g. Harper et al., 2010; Liu and Jansen, 2015), which suggests it may have value to other researchers. It may also have value for industry designers, where they may want to surface information-seeking or resource-seeking questions beyond the community to help people get answers, but not surface more vulnerable support questions, where trust may be more of an issue.

Understanding the type of question may be important in providing responses and evaluating the quality of those responses. For example, it is possible to automate some aspects of online communities with a chat bot which handles information seeking questions, while submitting the advice seeking to the community at large. This could potentially reduce burn-out in communities which often see repetitive questions, like those for health support communities or programming communities. The types of questions a community answers, may also be valuable for understanding the type of community and for recommending the community to new users. If a person is looking for a support community for people with a health issue, recommending one that focuses only on factual information may not provide the support a person needs. However, such interventions should be handled with care given the importance of interaction for providing support and encouraging future participation, which could help the community adapt to participant's changing needs.

Two of the perspectives are also worth highlighting here as contributions to ongoing research discussions: fostering and promotion. Some prior work has covered elements of these, and the general behaviors have been discussed as important elements in building and sustaining online communities, but many often do this without empirical research. Some, like reddit's own documentation, mention these

as practices for growing communities, but rarely provide guidance or structure for doing them effectively. This has been improved as of 2026, though this largely appears to happen through email nudges from Reddit which are sent to creators of new subreddits.

I referred to a set of community-building behaviors as *fostering* which I observed across many of the subreddits. I observed fostering in the form of content-creation behaviors, where accounts would create a number of pieces of content over some sustained time. I interpret these as attempts to engage users through content creation or questions intended for starting discussions. Many were broad questions that anyone could answer or interesting shared content to start discussions. This included some of the initial content-creation, *seeding*, which was a push to generate content often near the creation of the subreddit. My focus was on the content-creation aspects of these behaviors, but I theorize that there is also a lot of hidden work that people do for communities that could be considered fostering behavior.

Promotion, on the other hand, took place outside the community. Promotion is important for raising awareness of the community and attracting new participants. Even when the creator or community members are not concerned with growth, they likely still intend to attract a few participants. Various forms of promotion have been discussed in past research (e.g. Kraut and P. Resnick, 2012; Preece, 2000) , and by the platform (Reddit, 2023a; Reddit, 2023b). I expand on these by describing some specific forms on reddit, including those that may not be intentional forms of promotion, but through platform affordances of linking can still act as a promotion.

7.2 PREDICTING "SUCCESS" & WHO DOES THE WORK

In the second study, I explored the framework's ability to predict three measures of success. Since creators and participants may have different goals for their communities, I attempted to use aspects of the framework to explore success. Like prior work, I found that different perspectives performed better at predicting different measures of success (e.g. Aumayr and Hayes, 2017; Cunha et al., 2019).

I found that participation and interaction had the highest performance for predicting retention. Participation's high performance in predicting retention may have been due to using retention as a feature,

which suggests that successful communities may be more likely to continue being successful one year later. Both also performed well at predicting new user growth. Promotion did well in predicting survival, but not growth. This was surprising, especially in the context of the third study suggesting that promotions may precede joins, and more research is needed to understand this seeming paradox.

Fostering generally had the lowest performance. This may be due, at least in part, to how I operationalized fostering which focused on visible content created by the accounts. I had hoped this might capture elements of loyalty and dedication to a subreddit, in terms of sustained activity, which encompassed an essence not captured in other metrics like month-to-month retention. The performance was not poor enough to exclude it, and it may have lagged in performance due to the high prevalence of bots and the inclusion of content that was not engaged with, which may have included spam.

A large portion of the work creators and moderators perform may be hidden (S. A. Gilbert, 2020; Li, Hecht, and Chancellor, 2022; Seering, T. Wang, et al., 2019), which would not have been reflected in my fostering metric. For example, decisions to remove content may involve discussions among moderators, which may lead to other actions being taken (S. A. Gilbert, 2020; Lo, 2018). Their work and its visibility can be complicated by a need to have different ways of handling different norm/rule violations (C. B. Smith, McLaughlin, and Osborne, 2006), increases in content which increase workload (B. Butler et al., 2002), and the members reacting in unforeseen ways (Squirrell, 2019).

Perhaps most importantly, I found the creator was not active initially in over a quarter of all of the subreddits I analyzed. I found that the creators contributed very little content over the subreddits lifetime in a substantial portion of these. Some communities still appeared to survive. In some cases, this may mirror the r/LinusTechTips example from the first study, where a new, committed account came along to build it into an active subreddit and the initial creator never participated. This suggests that with enough motivation, time, and platform support dead subreddits may be revived by any account, regardless of them being a creator, moderator, or even normal account. The key is that someone needs to do the work to attract and sustain new visitors.

Even in subreddits which experienced different forms of success, I found that creators tended to be less active in producing content as the community aged. This was surprising. A few possible explanations here can explain this effect. The first might be that moderators and normal accounts eventually produce enough content to sustain or grow the community, requiring less work of the creator. In this case, the creator may spend more time doing less visible work, such as behind-the-scenes organization, moderation, and management. Another possibility might be fatigue and burnout. My dataset unfortunately lacks some of the behind-the-scenes activity, to explore whether the problem of switching to less visible work or fatigue. Regardless, although creators can be important to the establishment and health of a community, but their activity may not be a silver bullet. Other participants can do this work when the platform enables and structures systems to allow it. Platforms could use this information to develop tools to enable newly arriving accounts an ability to help contribute to the success of the community.

7.3 PROMOTION AND FOSTERING IN A POLITICAL SUBREDDIT

In the last analysis, I applied two perspectives of the framework to the evolution of a political subreddit which quickly grew to one of the largest. I used the framework to both contextualize *r/The_Donald* and compare it to other campaign subreddits to illustrate how it differed in its growth. For example, I found much of the growth appeared to be related to real-world events that paralleled the campaign.

I also found that many of the most active accounts – those engaged in fostering – had joined fairly early in the life of the subreddit. As expected, some of the fostering accounts had been previously active in right-leaning subreddits, though a surprisingly high number had been previously active in the Bernie Sanders subreddit (*r/SandersForPresident*). In a way, *r/TD* may have acted as a meeting ground for some of the more right-leaning and anti-social accounts on the platform. These anti-social behaviors include more hostile behaviors like harassment and brigading other communities, which may be turnoffs for some people to join communities. However, these behaviors may be appealing to some people (Bernstein et al., 2011). Curiously, the growth of the subreddit appears to have aligned with a platform-wide overall shift towards the far-right on reddit that began in early 2016 (Waller and A. Anderson, 2021).

As the subreddit grew, the majority of newly joined accounts participated less than a day. Some of this was likely the result of being an infamous community known for their anti-social behavior and attempts to 'hack' the front page into having a lot of pro-Trump news. Some of the increase in single-day joins, may have been reactions from people in other subreddits to front-page content, brigading, or anti-social behaviors on reddit.

I also found evidence suggesting that promotions may precede increases in new joins. This analysis was not causal, but at least establishes a link such that promotions generally occurred before many new accounts became active. This is important because even unintentional promotions could attract new users. This has been generally assumed on the platform, by way of best-practice documents and advice posts, but those often provide no empirical evidence of it working. While the temporal relationship aligns with prior work (Krohn and Weninger, 2022), my results also showed a potential increase in joins of more engaged members as well. However, the increase may be the result of external events and a feedback effect, such that activity also generates more promotions. This may be a process similar to the Matthew effect, such that growing levels of activity generate more activity, in part due to the front-page algorithms which create more visibility.

7.4 SURFACING INVISIBLE WORK

This work also forefronts some of the less visible work in community building, especially that work which can happen outside the community itself. This contributes to a line of research on work that is invisible or less visible. This is work that may not be easily visible to other participants or outside viewers (e.g. researchers). For example, a single comment on reddit can take minutes or possibly hours to write, but the comment has only a single timestamp of when the comment was submitted to the system. Other instances may not be easily visible at all, such as some of the content moderation.

A body of work that has surfaced hidden work of content moderation across online platforms (e.g. Lo, 2018; S. T. Roberts, 2016). This includes work around content moderation, discussions about policies, implementing policies, community outreach, and community building (e.g. Lo, 2018; Seering and Kairam,

2023; Seering, T. Wang, et al., 2019; Squirrell, 2019; Wohn, 2019). Some have additionally noted various forms of maintenance and promotion (e.g. B. Butler et al., 2002). On reddit, a significant portion of a moderators work may not be easily visible to the members of the community (Li, Hecht, and Chancellor, 2022).

One of the challenges I fore-fronted was the effort of creating content within a subreddit. The work of creating submissions and content can get collapsed into a piece of content with a singular creation time, potentially hiding the true amount of time it took to create the content. This included the challenge of finding suitable content for subreddits, and the effort needed to create large resource threads such as r/WomensSoccer. This may be true of many other types of submissions and comments, including some of the detailed questions. This would align with prior work which found people could spend a lot of time composing questions in an attempt to get help (Turner and K. E. Fisher, 2006).

This suggests that fostering work may take a significant investment finding content that helps convey the purpose and character of the subreddit for those who invest in it. This includes seeded content, of course, which may require at least one large chunk of time. What appears as a few pieces of seeded content, may take minutes or even hours to find and share. This requires not just searching, but potentially consuming the content and making decisions about the quality of the content and the fit of the community.

I also surfaced promotion work, which is less visible as the traces exist outside of the subreddit. I found several instances of creators or moderators promoting the subreddit outside it, which may require searching or finding for subreddits which might be relevant to share in. This may also require communication with the moderators in those other communities to ask for permission to share as the platform suggests. Uncovering work like this requires other methods to explore the activity of specific accounts beyond the boundaries of the community traces.

Three of the perspectives in this framework hint about some of the less visible work being done on the subreddit. The framework would benefit from additional development of the types of work being invested to create content on the platform. However, as proposed the framework may still provide value to other researchers in trying to understand and compare this work between subreddits.

This work has focused entirely on the trace data from common usage on reddit, but there is undoubtedly more hidden work being done during the everyday usage of reddit. More research is needed to better understand the nature of this work, and to attempt to quantify the work being done. However, I leave this research to future projects, which I hope will better describe and account for some of the work being done to build these communities.

7.5 TAILORING SUCCESS

This work speaks suggests a need to consider how we measure success in online communities. Similarly, who should define what makes a community success and for whom is that successful? Like other researchers, I believe these metrics are best determined by the goals of the members.

Scholars have explored a number of measures including growth, survival, and retention (Backstrom, Huttenlocher, et al., 2006; Kraut and Fiore, 2014; Raban, Moldovan, and Jones, 2010; Tsugawa and Niida, 2019; X. Wang, B. S. Butler, and Ren, 2013). These are important to most communities to some extent, but research suggests communities may have other goals, some of which suggest that smaller communities have different values (Hwang and Foote, 2021; TeBlunthuis, Kiene, et al., 2022).

A line of research has focused on the many forms of online communities and the goals and values of their members (e.g. Foote, Gergle, and Shaw, 2017; Massimi et al., 2014; Weld, A. X. Zhang, and Althoff, 2022). Some communities are focused on providing support, news, discussion, and even entertainment. Many can provide multiple of these at the same time. Each community may operate in very different ways, and thus have different developmental goals. Many value the quality of the content shared (Leimeister, Sidiras, and Krcmar, 2006; TeBlunthuis, Kiene, et al., 2022; Weld, A. X. Zhang, and Althoff, 2022).

Growth is a common measure of success. It generally is measured by the number of members or volume of content. Some amount of growth is necessary for any community to function as a community, but not all communities need to be large. In Chapter 5, I included three metrics of success that have been used in prior works in an effort to engage this work with other works in the field, but in that work and the previous chapter I found many small and medium communities surviving with only small amounts

of growth. Large communities clearly have many values like having a larger audience for questions or feedback or possibly large amounts of content, and possibly higher quality content – all of which are things members value on reddit (Weld, A. X. Zhang, and Althoff, 2022). However, large communities also have downsides including a large increase in work (B. Butler et al., 2002).

Small communities can also have value to users and may focus more niche topics (Hwang and Foote, 2021; Leimeister, Sidiras, and Krcmar, 2006; TeBlunthuis, Kiene, et al., 2022). These smaller communities can also have a much lower signal-to-noise ratio enabling people to find content that they love. They may also feel more welcoming (Hwang and Foote, 2021), and people may feel more affinity to these smaller groups (Chancellor, Hu, and De Choudhury, 2018).

Another common metric is survival (Ducheneaut et al., 2007; Raban, Moldovan, and Jones, 2010; Tsugawa and Niida, 2019). This is often measured by defining a period of inactivity where the community is considered dead. However, in both the first and second study, I found a number of communities which became active after a potential death period. For example, the r/LinusTechTips was largely inactive until another user took control of the subreddit and began making regular posts and setting up a bot to help create these regular submissions. I also found that some communities were cyclical in nature, like r/WomensSoccer, where they became significantly more active in parts of the year before declining as the season ended. The cyclical nature would make the subreddit appear dead or nearly dead in off-seasons. While metrics like survival may still be useful for many communities, it may not be a simple one-size fits all.

Similarly, some subreddits may never be intended to exist for long periods of time. Political campaigns are one example of this. While people may participate beyond a campaign, it should also be reasonable for the community and people to invest their time elsewhere.

Several scholars argued for the inclusion of different measures of success beyond growth, as the ultimate goal should be meeting the needs of its members (e.g. Bruckman, 2006; Cunha et al., 2019; Massimi et al., 2014; Prinster et al., 2024). These can depend on the purpose of the community and what the community values. A community focused around skill development may initially value growth, but later may value

high-quality content and discussions about that skill, whereas a community for learning the programming language python may value the number of questions answered.

By expanding the available success metrics, we can build better tools to monitor and understand community health. This may also enable the development of better tools for enabling communities to reach those goals. This may include better statistics for moderators to monitor the health and make decisions about actions to take. I used a few simple metrics along with their larger concepts which may help predict different forms of success and add to the potential metrics of interest to moderators and their communities. These largely fell under 4 perspectives interaction, participation, fostering, and promotion. The metrics include number of fostering authors, number of promotions, retention, early community activity, Gini coefficient for participation inequality. Building tools which leverage these, can also provide targeted recommendations for specific goals the communities hope to achieve.

Small, niche communities may be particularly important for new platforms that are hoping to attract new users. Some research suggests that one of the unique values of reddit may be its smaller more niche communities, as redditors tended to increasingly explore smaller communities the longer they stay active on the platform (Tan and L. Lee, 2015). A portion of these can be off-shoots of larger existing communities (Tan, 2018). People who have attempted to migrate away from reddit often have trouble finding these niche communities, and it may be a reason for them to stay on the platform (Newell et al., 2016).

7.6 PLACEMAKING IN ONLINE COMMUNITIES

Some of the behaviors I observed in this work can be seen as forms of placemaking. It is a term intended to describe behaviors which maintain or establish a sense of place – something with social meaning. The term placemaking has been used across multiple fields including urban design (Friedmann, 2010), tourism (Dupre, 2019), and even human-computer interaction (Harrison and Dourish, 1996).

The concept relies on the distinction between space and place, where place has a social meaning or an identity through people using and interacting within the space. Through their use of the space or

interaction with others in the space, they can create a social meaning of the space, and thus making it feel like a place.

In the physical world, there are many actions people can do that make something a place. Daily living within a house can make it feel like a home through the interaction with the space and with others. People can also collectively participate in the design, building, and modification of that space to better fit their needs (Dovey, 1983). However, actions and the meanings they produce can be transitory such as homeless camps being swept (Ruddick, 2014) or nomadic peoples moving.

Behaviors like fostering can be seen as forms of placemaking, where people attempt to develop, adapt, or maintain a space through usage. Sustained content creation within a community can provide character and help shape the value and purpose of a community – similar to designing or decorating a space. This can also serve as an opportunity for others to engage. Seeded content acts as an initial set for establishing the identity, while enabling easy engagement for visitors. There are also specific forms of content I witnessed that may also help establish this, such as welcome messages, meta-discussions about the community, and supportive comments to help engage new submitters.

As new people begin to contribute, they can reaffirm the existing identity or shift it in new ways. When accepted by the larger community, this can shift the social meaning of the community and establish new niches for the place and its members. This can be useful in helping distinguish it from other similar communities. Over time, this can also lead to the establishment and enforcement of norms of acceptable content and behavior within the community.

Other behaviors like promotion can be viewed as a form of placemaking. When cross-posting or mentioning a subreddit, the choice of the target subreddit helps delineate the border of that community. In many cases, people may be likely to share mentions and links in more similar communities as was seen in Chapter 6, where many promotions were shared in politically aligned subreddits. This is also similar to border-shaping behavior of cross-posts on USENET (B. S. Butler and X. Wang, 2012). Though there were also instances of sharing links and mentions in an oppositional way, which could be used to distinguish a clear border between two communities through an example. In both instances, these can still serve to both

maintain a space and its borders or to attempt to modify the borders of the space by enforcing or moving a theoretical border. In the case of a positive promotion to a new space, this may lead to new people joining and shifts in the topics or interactions. Even the more oppositional promotions can lead to conflict (Datta and Adar, 2019; S. Kumar et al., 2018) and potentially change the feel of the community and how it responds.

As space and place scholars note placemaking must continually happen (Dourish et al., 1996) over time. It is not sufficient to establish a place and it exists forever. Active work must be done to continue to maintain and recreate the space over time. This means the place is continuously maintained or remade by those active in the space over time. The data here suggests that there may be a point at which some of the placemaking work may be shifted to other people. For example, the visible work of content creation may shift from creators to moderators and normal accounts after critical mass has been reached, such that the creator and moderators may focus their time elsewhere. That is not to imply they are not also doing placemaking work, but rather that the type of work they do may change over time.

Understanding the placemaking behaviors, intentional or unintentional, may be helpful for people creating and establish new platforms and spaces. The earlier example of developing new platforms is a particularly salient one. The development of these spaces may require a significant amount of placemaking work initially to help make some groups engaged enough on the platform and help develop that sense of place. This may mean the owners or developers spending time on the platform creating content (e.g. Fostering). The reddit founders admit to having done exactly this during the creation of that platform (C. Johnston, 2012).

7.7 IMPLICATIONS AND FUTURE WORK

I have summarized some of the more important findings in the previous sections, but I want to provide a few ideas of how I see this work being used and how I hope it will be extended.

First, I join the other scholars in calling for more measures of success based on the different forms of goals communities will have. I believe there's some value in aspects of this framework at helping provide new metrics to better understand how communities can better meet their goals. This also lends itself to

providing some better tools to help new communities, beyond their creators, grow. In some instances, this may require us to adjust our perceptions of what online communities need to be. For example, it may be successful for communities to have a limited purpose, such as helping a campaign, before letting them die or closing them entirely.

One of the challenges with metrics like growth and survival is that they describe something a current state of the community without indication of what may be affecting these metrics. They may describe the overall health of a community but provide few insights for people to actually fix any problems that might lead to them. This is a challenge for both platforms and researchers interested in this area. This framework, and perhaps others like it, may provide a deeper level of understanding of the behaviors necessary to move some of these metrics. Since promotions and fostering may have some level of value in measuring different forms of success, they may also provide necessary tools for indicating how to reach those goals.

Second, the understanding of these different forms of success may also provide platforms with the knowledge to build tools to help communities with tasks they are already doing. One example could be tooling to aid content curation from across reddit that could be cross-posted, though obviously with specific moderator control to manage undesirable behavior. This may reduce some of the burden of seeking and finding additional content. Another might be suggestions for where and how to promote, including leveraging specially placed ads on people's front pages for new communities similar to those someone has already participated in.

More research is needed to be able to progress on developing newer metrics based on the framework and to develop tools from it. This includes more in-depth research of creators, moderators, and members about their intent in doing some of these behaviors and what they hope to achieve. For example, we need to understand more about the work that creators do in the early part of creating their subreddit, and about the types of things they do not know. We also need to better understand the range of goals that creators and communities have before we can design new metrics and interventions.

I have also largely avoided the central question of when these subreddits may become a community. I have taken a more participatory perspective for enabling communities to decide their own goals, which makes measuring success difficult for researchers that do not work for a platform¹. In the same vein, I believe that a community may best be determined by the perceptions of the community members (aka sense of community). A popular line of research in this topic has been around the sense of community (McMillan and Chavis, 1986), and the sense of virtual community (A. Blanchard and M. Markus, 2002; A. L. Blanchard and M. L. Markus, 2004). These are the feelings people get that lead them to describe something as a community. Platforms like Facebook relied for a period on some potential antecedents of the sense of community via metrics like 'meaningful interactions' (Litt et al., 2020). It's unclear how these may relate to a sense of community, and whether it's possible to use such metrics to understand when people may begin feeling like a sense of community has developed.

This work has been limited by my focus on trace data, which obviously lacks the psychological perspective of participants within the spaces. We need more research to better understand the relationship between feeling a sense of community and feeling a sense of place. In the context of an online community, do people feel the two at similar times? Are there times when someone might feel one of those but not the other? And perhaps most importantly, are their metrics which might correlate with people beginning to feel either of these? The latter could be useful in understanding the point at which point people see subreddits as communities without the need to continuously poll users.

Even in the case that no one uses the framework I have designed here, I hope that research continues on online communities, such that we can continue to build new solutions which better serve the groups that rely on them.

¹This is a problem of data access to some behavioral data needed. which may introduce privacy and security issues by making this publicly available. In part, this may

7.8 REFLECTIONS ON THIS WORK'S RELEVANCE TODAY

Perhaps the most important question about this work is whether the findings still relevant in 2026 as I write this. I would argue that despite being focused on reddit, many of the core findings are still relevant for online communities even today. Different parts of this work may be useful for platform designers, community creators, and other researchers who do work for or on online communities.

I developed the framework to be generalized and hopefully useful for analyzing and understanding other forms of online community. The framework is grounded within the data and developed through an inductive process. It has been influenced by my knowledge of existing research and personal experience across other online communities. As I illustrated in my second study, several of measures from the perspectives may be useful for predicting different forms of success which could be especially important to platforms, community creators, and their members.

The framework and the behavioral descriptions, especially those from the early period of reddit, may be useful to people creating platforms where communities may form. In recent years, this has become a more pressing issue for people concerned about the private ownership of the platforms underlying modern social media Coatney, [2019](#); Doctorow, [2025](#); Schulman and Mir, [2023](#). This has led some people to attempt to create open and decentralized alternatives (e.g. Mastodon and Bluesky). These alternatives could all benefit from using such a framework, and especially in helping communities better understand how to reach their goals. Because this work looks at the early part of reddit's evolution, it may be especially useful for people developing new platforms for online communities especially early in their development. The reddit that exists in 2026 has many more affordances that would be difficult for new platforms to build, and so this work highlights more of the core behaviors of the active members prior to many of the affordances available today.

The work may also be useful for some of the private companies that host these online communities. I worked in a large private company that hosted many online communities while working on this project. This company had a significant amount of internal research around online communities, but it still lacked some of the knowledge from this dissertation which could have been useful in helping communities

achieve their individual goals. The research they had was also somewhat dated and was based around some specific affordances that were difficult to implement because of cultural and legal changes since the platform started (e.g. world-wide privacy legislation, legal cases, etc). Portions of this work may have been useful in improving the success of some new forms of online communities on their platform.

Reddit's platform has also evolved since 2016, and some newer features may be seen as a form of validation for some of the ideas in this work. As of 2026, the platform provides a number of affordances for creating a subreddit that did not exist in 2016 or earlier. For example, they introduced a series of achievements to gamify some of the participation and encourage participants to engage in community-building behaviors.

The participant achievements include “repeat contributor” and “poster prodigy” which encourage continued content creation and higher quality submissions (as determined by vote totals). These align with the seeding and fostering behaviors I describe, which may be important in encouraging others to participate and join the community.

They also introduced some achievements specifically for creators and moderators to encourage many of the same behaviors. The “mod-hatchling” achievement encourages a collection of both content creation and promotion of both sustained submission creation for 3 days and inviting people to join the community.

However, technology has also advanced and there are opportunities for reducing the effort of community building. Modern recommendation algorithms could potentially lessen the burden of promotion of members needing to promote subreddits, and AI tools could help members curate new content from across the platform and the web at large. These would obviously need guardrails to prevent spam and abuse, but these could still help surface higher-quality relevant content for participants. I would also caution that one of the core values of the community is the interaction, and reducing interaction may also reduce the desire to post into the community, thus limiting future participation (and limiting future training data). These technologies could also be useful for new platform designers and developers to tackle these problems early in the platform's life.

8 APPENDIX A

8.1 QUALITATIVE CODE BOOK

Themes	Top-Level Code	Code	Old Codes	Description
Promotion	Promoting	advertising other sub	Advertising, alternative sub, recruit- ing, growth, related sub, sub announcement, (and xpost)	Advertising a different sub within this sub. Could also be advertising external forum
Promotion, Curation	Cross-posting	cross posting from external	xpost	cross posting content from a different sub here. While this is a form of advertising, it's also a form of aggregating content from sim- ilar subs that may be more targeted in this sub
Promotion	Cross-posting	cross posting to external	xpost, adver- tising	cross posting into another sub is a form of advertising. Given that coding is restricted to the subreddit in question, this code will only occur on content *suggesting* or *asking* users to cross post into other subreddits.
Promotion	Promoting	seeding	growth	An attempt by someone to seed a set of con- tent. This is a number of sequential posts from the same person (and perhaps a small number of people) in a short period of time. This often occurs at the beginning of a sub or after a transfer of power as a means of pro- viding a set of content for new users to see and engage with.

Themes	Top-Level Code	Code	Old Codes	Description
locality	local	seeking local resources	local	a submission or comment seeking resources in a specific location (geographically)
locality	local	meeting	meet	a submission or comment seeking to meet people in person. This may happen for the sake of discussing a topic, practicing a sport, etc
locality	local	local event sharing	local, meet	sharing a local event
content	content - question	advice seeking	Advice, opinion/feedback, advice/feedback, advice seeking, question	Users asking for advice on a topic (usually personal). This might be feedback about a particular technique they are seeking to improve, or an idea they have. Sometimes these aren't questions, but can just be general statements where people are looking for someone to share advice.
content	content - question	resource seeking	question	Users seeking a resource (eg an external link, a website, name, local group, etc) from the community. This request or question is less about subjective or opinion feedback and the expectation is something external to the subreddit. Though the subreddit may have expertise about where something is or what the best resource is, this code is explicitly for when the user is looking for something external to the subreddit.

Themes	Top-Level Code	Code	Old Codes	Description
content	content - ques- tion	information seek- ing	question	Users are seeking information from the com- munity. This differs from resource seeking because users believe the experts here are the community, and their past experiences are the resource. This will probably overlap a lot with starting a discussion especially for more broad conversations where people. I think the difference here is that the informa- tion seeking might be more factual about the experiences, and less about opinions. This is also more specific. So for example, asking for how someone handled an experi- ence versus the very broad “What changed when you turned 30?” This could also be simply looking for per- spective from other people with similar situ- ations, perhaps for the sake of contextualiz- ing their situation or not feeling alone. someone making a specific request or ask- ing for help in doing something. This may overlap with collective action, though unlike collective action the request can be satisfied by a single person
content	content - request	requesting		

Themes	Top-Level Code	Code	Old Codes	Description
content	content - external resource	sharing external resource	External chat, share, link, info sharing, xpost, alter-native sub, related sub	Sharing a link or information about an external resource. This can include an information resource, tool, or something like an external chat. This is more than just a simple news link, but a link that shares a schedule of games could be included with this. Links to streams could be included. This is something that might be consulted at a later point or used for some period of time (chat, streams, etc)
				QUESTION: Should this resource be a single piece of content including an image or should this be more than just a single piece? If the former, this would include almost any external link to any code of content. What exactly counts as a resource? Much of the content in r/EarthScience could be considered an external resource by this definition and would that differ from the videos shared in r/Eskrima?
content	content - internal resource	sharing internal resource		This can also be a created resource. For example, aggregations that collect a number of links or other pieces of information into a single post. This could also be links to other posts/wikis/etc that were created within the subreddit.

Themes	Top-Level Code	Code	Old Codes	Description
content	content - discussion starter	starting discussion	Discussion, share, video response, sequence/status, personal story, brag/share,	These are attempts to start a discussion around a general topic. While questions can do this, these are often less focused on getting a specific interest and starting a broader discussion and can include starter content like pictures or examples, or links to get general thoughts around. May still overlap with questions though. These are generally very broad
content	content - event	sharing event	match announcement, live match	Sharing an event that was relevant to subreddit. This may be an upcoming announcement or a discussion around an ongoing or past event
content	content - personal	sharing personal content	Personal story, brag/share, appreciation, feedback,	Someone sharing a personal story or anecdote. This may include a personal victory, the story of finding the subreddit, how the subreddit affects them. This may overlap with advice seeking in the instance where the user needs to share personal anecdotes to get advice.
content	content - community engagement	engaging community	awards, survey,	This is an attempt by someone to keep a community engaged, often from a moderator or active user.
content	content - xpost	cross-posting	xpost	a cross post from another community. This is a curation of content from other communities, typically larger. These are now structured through the reddit api, but early ones are manually flagged as “xpost” “cross post” etc or have a link to a different subreddit
content	content - update	updating	update	sharing an update to a previous post or information.

Themes	Top-Level Code	Code	Old Codes	Description
content	content - subredd- dit meta	discussing sub		meta discussions about the subreddit or changes around the subreddit. Perhaps either as requests or otherwise
low- activity/growth	growth	checking activity	alive/status, user plea	Users asking if the sub is still active or alive, or pleading for more activity
low- activity/growth	growth	growth push	user plea	User attempt to boost content. user plea for more activity
low- activity/growth	growth	takeover	takeover	A user announcing a takeover or a request to takeover the subreddit
Communication	top-down	top-down com- municating	Subreddit background, stats, an- nounce- ment, takeover, flair	These are communications that occur (usu-ally a moderator but there may be flexibility here) that communicate something about the community like some background on it's purpose, statistics about the subreddit, or other announcements related to the subredd- dit.
Communication	top-down	requesting com- munity feedback	Survey, ad- vertise, mod search, recruiting	This is seeking some sort of feedback from a moderator or other person in power of the A request by mods for volunteers to help lead initiatives or be moderators within the sub- reddit. This may also co-occur with requests for promotion in other subreddits
Communication	top-down	recruiting		

Themes	Top-Level Code	Code	Old Codes	Description
Collective Action	collective action	requesting help/support	Collective action, fundraiser, fundraising, call to action, recruiting, xpost	these are requests of the community to engage with some sort of action like a fundraiser, growth attempts, or help. In the case of growth attempts, this code will overlap with others. The distinction here is that collective action doesn't need to be top- down QUESTION: How would this differ from nor- mal requesting?

9 APPENDIX B

9.1 METHODOLOGICAL NOTES

9.1.1 DELETED AND REMOVED CONTENT

Removed and deleted content within the pushshift dataset is problematic, partially due to changing inconsistencies in how they have been handled by reddit. As the platform grew, so too did the needs of the platform which may have been handled inconsistently until formal affordances and processes were developed. Historical submissions and comments from reddit admins and developers suggest this was the case on reddit as it developed new tools and methods for moderating and handling unwanted content.

Additionally, the dataset was collected over time, which means that some attributes and details change over the course of collection. This can present some methodological challenges in terms of both analysis and interpretation of data that involves content related to some of these features. For example, self posts are treated differently when removed than link submissions. Self posts have their text changed to '[deleted]' or '[removed]'.

In 2011, a reddit admin and developer implemented a change to change the selftext attribute of a submission when an admin a self post (u/spladug, 2011). Prior to this change if there was content within the submission text that needed to be removed, admins had to edit the submission text themselves, and it is not clear if this was done in a consistent fashion or is even detectable through the API.

This change replaces the text of a removed self post submission with the text '[deleted]', thus standardizing the process in a formal method. This change was extended to moderator removed self text submissions shortly thereafter (u/spladug, 2024). The self text here was not completely removed internally,

but simply rendered on the site, application, and API as deleted. The original author, moderators, and admins could see the original text. If a moderator later 're-approved' a removed submission, the text would be restored. This implies there could be a short window where some data within the dataset is marked as '[deleted]' or '[removed]' within the pushshift data, but visible on the site which may have been observed by some (u/theoddjosh, 2020).

Comments from the same admin in these threads suggest that this only affects admin and moderator removals and not when users delete their accounts, though a number of bots and services did allow users to retrospectively edit their content.

On September 21st, 2015, reddit announced a change where removed comments would be shown as '[removed]' instead of '[deleted]' as was the case prior to the change (u/tdohz, 2015). This change also removed the contents of self-posts when they were deleted, which had not been the case prior to the change. So between 2011 and 2015, self posts which were removed had their text changed to 'deleted', but deleted posts kept their text. This change changes both the text to 'removed' instead of 'deleted'.

However, as one of the comments illustrates in that post, an author can still delete their comment after it has been removed to change the text from '[removed]' to '[deleted]', meaning that trying to count '[removed]' content may not always be accurate, especially for people who delete their content.

Later changes added new fields to partially clarify some aspects of deletions and removals, including reasons for removals, but these fields are not available within the dataset instance I have.

I opted for transparency and attempted to report deleted and removed content separately. This adds some complexity to the interpretation of a number of sections, but these need to be considered with caution about the data. I have opted to exclude content where the author is '[deleted]' from specific analysis and summaries that involve author content counts. For example, '[deleted]' would appear as a prolific author and one that is fostering in many subreddits.

9.1.2 SCORES

Multiple documents describe the score to be equal to the number of upvotes minus the number of downvotes (Baumgartner et al., 2020; reddit, 2016). Though this may apply to comments, this does not

entirely apply to submissions. The platform imposes a rule such that it does not display negative scores for submissions, and this appears to carry over data coming from the API. There have been adjustments on the platform, which have affected the displayed scores and it is unclear if these also affected the API (u/KeyserSosa, 2016).

To some extent, you can estimate the number of upvote and downvotes on a given submission or comment by using a field *upvote_ratio*, but this cannot work for negative score submissions. The pushshift data snapshot data did not appear to have *upvote_ratio* in the dumps I used, and it is not documented in the paper, though it may have been added later.

There is an additional question of temporal API changes. As the platform has evolved over time, there have been some API changes such as hiding the number of upvotes and downvotes (u/Deimorz, 2014) and finally adding a score value (u/umbrae, 2014).

Further, reddit has employed a vote-fuzzing score for sometime. This does affect the API, though it is not entirely certain how that is or if it has changed over the years. This has also affected elements of the site such as the text “% like it” as described by this admin (u/Deimorz, 2014).

9.1.3 MENTION AND LINK REGULAR EXPRESSIONS

In multiple analyses, I look at the links and mentions within titles and the text of both comments and submissions. These are borrowed from what I believe was an early piece of reddit source and my testing found that they seemed to cover many common cases¹ So for example, these regular expressions detect mentions and links in multiple formats including r/subreddit and reddit.com/r/subreddit as well as full links like reddit.com/r/subreddit/submission_id/title/optional_comment_id .

Below is the regular expression modified for use in PostgreSQL was:

```
(?:\| | (?<=[!\"#$%&'()*+,-\./:;<=>?@[\]\\^\^_‘\{\|\}\~\s]) | (?<=\A))r\|
```

¹In a good example of a methodological mistake, I made a comment in the code that it was from a reddit source update, but I forgot to paste the source URL into the source code when the source code was written. Newer attempts to find the source of this regular expression has failed. Most of the most recently available source code seems to parse links in a very different way using the URLParser code.

9 Appendix B

```
([\\w\\+\\-]{2,}|reddit\\.com)[\\/\\w\\-]*
```

The link break in this code is added for readability and is not present within code. During testing of the regular expression, I used many different forms of links which should match. This regular expression does match some instances of text which are not actually links, though these are relatively few. In the processing pipeline, I added additional checks that the length of the subreddit is reasonable, and for some analyses I did check that the subreddit is valid.

One limitation of this regular expression is that it does not properly resolve multi-reddit links. These are links in the form of *r/AskReddit+science+funny+videos* where multiple subreddits are linked in the same URL. These do occur on reddit but were not common.

BIBLIOGRAPHY

- Ackerman, M. S. (2000). “The intellectual challenge of CSCW: the gap between social requirements and technical feasibility”. *Human–Computer Interaction* 15:2-3, pp. 179–203.
- Adamic, L. A. and B. A. Huberman (2000). “Power-law distribution of the world wide web”. *science* 287:5461, pp. 2115–2115.
- Adamic, L. A., J. Zhang, E. Bakshy, and M. S. Ackerman (2008). “Knowledge sharing and yahoo answers: everyone knows something”. In: *Proceedings of the 17th international conference on World Wide Web*, pp. 665–674.
- Alexa (Aug. 22, 2021). *The Top 500 sites on the web*. URL: <https://web.archive.org/web/20210822000208/https://www.alexa.com/topsites> (visited on 08/22/2021).
- Anderson, M. and J. Jiang (Nov. 2018). “Teens’ Social Media Habits and Experiences”. URL: https://www.pewinternet.org/wp-content/uploads/sites/9/2018/11/PI_2018.11.28_teens-social-media_FINAL2.pdf (visited on 05/22/2026).
- Andrews, D. C. (2002). “Audience-specific online community design”. *Communications of the ACM* 45:4, pp. 64–68.
- Arguello, J., B. S. Butler, E. Joyce, R. E. Kraut, K. S. Ling, C. Rosé, and X. Wang (2006). “Talk to me: foundations for successful individual-group interactions in online communities”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 959–968.
- Aumayr, E. and C. Hayes (2017). “The path to success: a study of user behaviour and success criteria in online communities”. In: *Proceedings of the International Conference on Web Intelligence*, pp. 219–226.

Bibliography

- Baborska-Narozny, M., E. Stirling, and F. Stevenson (2016). "Exploring the Relationship Between a Facebook Group and Face-to-Face Interactions in 'Weak-Tie' Residential Communities". In: *Proceedings of the 7th 2016 International Conference on Social Media & Society*, pp. 1–8.
- Backstrom, L., D. Huttenlocher, J. Kleinberg, and X. Lan (2006). "Group formation in large social networks: membership, growth, and evolution". In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 44–54.
- Backstrom, L., J. Kleinberg, L. Lee, and C. Danescu-Niculescu-Mizil (2013). "Characterizing and curating conversation threads: expansion, focus, volume, re-entry". In: *Proceedings of the sixth ACM international conference on Web search and data mining*, pp. 13–22.
- Backstrom, L., R. Kumar, C. Marlow, J. Novak, and A. Tomkins (2008). "Preferential behavior in online groups". In: *Proceedings of the 2008 international conference on web search and data mining*, pp. 117–128.
- Barabási, A.-L. and R. Albert (1999). "Emergence of scaling in random networks". *science* 286:5439, pp. 509–512.
- Barlow, J. P. (Feb. 8, 1996). *A Declaration of the Independence of Cyberspace*. [Online; accessed 8. Oct. 2022]. URL: <https://www.eff.org/cyberspace-independence> (visited on 10/08/2022).
- Barthel, M., G. Stocking, J. Holcomb, and A. Mitchell (Feb. 25, 2016). *Seven-in-Ten Reddit Users Get News on the Site*. URL: <https://www.pewresearch.org/journalism/2016/02/25/seven-in-ten-reddit-users-get-news-on-the-site/> (visited on 12/06/2024).
- Basu, T. (Apr. 2, 2020). "Reddit's automoderator is the future of the internet, and deeply imperfect". *MIT Technology Review*. URL: <https://www.technologyreview.com/2019/10/30/75229/reddits-automoderator-automod-catches-more-than-any-human-could-but-its-still-imperfect> (visited on 11/30/2024).
- Bateman, P. J. (2008). "Online community referrals and commitment: How two aspects of community life impact member participation". PhD thesis. University of Pittsburgh.

Bibliography

- Baumgartner, J., S. Zannettou, B. Keegan, M. Squire, and J. Blackburn (2020). “The Pushshift Reddit Dataset”. *arXiv preprint arXiv:2001.08435*.
- Baym, N. K. (1993). “Interpreting soap operas and creating community: Inside a computer-mediated fan culture”. *Journal of folklore research*, pp. 143–176.
- (1998). “The emergence of on-line community”. *cybersociety* 2:0, pp. 35–68.
- (2009). “A call for grounding in the face of blurred boundaries”. *Journal of Computer-Mediated Communication* 14:3, pp. 720–723.
- Beenen, G., K. Ling, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. E. Kraut (Nov. 6, 2004). “Using Social Psychology to Motivate Contributions to Online Communities”. In: *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. CSCW04: Computer Supported Cooperative Work. ACM, Chicago Illinois USA, pp. 212–221. ISBN: 978-1-58113-810-8. DOI: [10.1145/1031607.1031642](https://doi.org/10.1145/1031607.1031642). URL: <https://dl.acm.org/doi/10.1145/1031607.1031642> (visited on 05/31/2026).
- Bell, C. and H. Newby (1975). *Community studies: An introduction to the sociology of the local community*. George Allen & Unwin Ltd, London.
- Bender, T. (1978). *Community and social change in America*. Rutgers University Press New Brunswick, NJ.
- Bennett, W. L., A. Segerberg, and S. Walker (Feb. 7, 2014). “Organization in the Crowd: Peer Production in Large-Scale Networked Protests”. *Information, Communication & Society* 17:2, pp. 232–260. ISSN: 1369-118X, 1468-4462. DOI: [10.1080/1369118X.2013.870379](https://doi.org/10.1080/1369118X.2013.870379). URL: <http://www.tandfonline.com/doi/abs/10.1080/1369118X.2013.870379> (visited on 12/04/2024).
- Berente, N. and S. Seidel (2014). “Big data & inductive theory development: towards computational grounded theory?” In: *Proceedings of the Americas Conference on Information Systems*. Ed. by T. A. and R. B. Association for Information Systems, Savannah, GA, USA.
- Bernstein, M., A. Monroy-Hernández, D. Harry, P. André, K. Panovich, and G. Vargas (Aug. 2011). “4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community”. *Proceedings of the International AAAI Conference on Web and Social Media* 5:1, pp. 50–57. URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/14134>.

Bibliography

- Bishop, J. (July 2007). "Increasing Participation in Online Communities: A Framework for Human–Computer Interaction". *Computers in Human Behavior* 23:4, pp. 1881–1893. ISSN: 07475632. DOI: [10.1016/j.chb.2005.11.004](https://doi.org/10.1016/j.chb.2005.11.004). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563205000956> (visited on 12/04/2024).
- Blackwell, L., J. Hardy, T. Ammari, T. Veinot, C. Lampe, and S. Schoenebeck (2016). "LGBT parents and social media: Advocacy, privacy, and disclosure during shifting social movements". In: *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 610–622.
- Blanchard, A. and M. Markus (2002). "Sense of Virtual Community - Maintaining the Experience of Belonging". In: *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*. 35th Annual Hawaii International Conference on System Sciences. IEEE Comput. Soc, Big Island, HI, USA, pp. 3566–3575. ISBN: 978-0-7695-1435-2. DOI: [10.1109/HICSS.2002.994449](https://doi.org/10.1109/HICSS.2002.994449). URL: <http://ieeexplore.ieee.org/document/994449/> (visited on 05/06/2025).
- Blanchard, A. L. and M. L. Markus (2004). "The experienced" sense" of a virtual community: Characteristics and processes". *ACM Sigmis Database: the database for advances in information systems* 35:1, pp. 64–79.
- Blanchard, A. L. (Dec. 2007). "Developing a Sense of Virtual Community Measure". *CyberPsychology & Behavior* 10:6, pp. 827–830. ISSN: 1094-9313, 1557-8364. DOI: [10.1089/cpb.2007.9946](https://doi.org/10.1089/cpb.2007.9946). URL: <http://www.liebertpub.com/doi/10.1089/cpb.2007.9946> (visited on 12/04/2024).
- (Sept. 2008). "Testing a Model of Sense of Virtual Community". *Computers in Human Behavior* 24:5, pp. 2107–2123. ISSN: 07475632. DOI: [10.1016/j.chb.2007.10.002](https://doi.org/10.1016/j.chb.2007.10.002). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0747563207001562> (visited on 12/04/2024).
- Bruckman, A. (2006). "A new perspective on "community" and its implications for computer-mediated communication systems". In: *CHI'06 extended abstracts on Human factors in computing systems*, pp. 616–621.
- Bruckman, A. and C. Jensen (July 8, 2002). "The Mystery of the Death of MediaMOO: Seven Years of Evolution of an Online Community". In: *Building Virtual Communities*. Ed. by K. A. Renninger

Bibliography

- and W. Shumar. 1st ed. Cambridge University Press, pp. 21–33. ISBN: 978-0521785587. DOI: [10.1017/CB09780511606373.006](https://doi.org/10.1017/CB09780511606373.006). URL: https://www.cambridge.org/core/product/identifier/CB09780511606373A014/type/book_part (visited on 12/04/2024).
- Bruckman, A. and M. Resnick (Mar. 1995). “The MediaMOO Project: Constructionism and Professional Community”. *Convergence: The International Journal of Research into New Media Technologies* 1:1, pp. 94–109. ISSN: 1354-8565, 1748-7382. DOI: [10.1177/135485659500100110](https://doi.org/10.1177/135485659500100110). URL: <https://journals.sagepub.com/doi/10.1177/135485659500100110> (visited on 04/11/2026).
- Burke, M., R. E. Kraut, and E. Joyce (2010). “Membership claims and requests: Conversation-level newcomer socialization strategies in online groups”. *Small group research* 41:1, pp. 4–40.
- Butler, B., L. Sproull, S. Kiesler, and R. E. Kraut (2002). “Community effort in online groups: Who does the work and why”. *Leadership at a distance: Research in technologically supported work* 1, pp. 171–194.
- Butler, B. S. (2001). “Membership size, communication activity, and sustainability: A resource-based model of online social structures”. *Information systems research* 12:4, pp. 346–362.
- Butler, B. S. and X. Wang (2012). “The cross-purposes of cross-posting: Boundary reshaping behavior in online discussion communities”. *Information Systems Research* 23:3-part-2, pp. 993–1010.
- Butler, B. S. (1999). “The dynamics of cyberspace: Examining and modelling online social structure”. PhD thesis. Carnegie Mellon University.
- Butterworth, E. and D. Weir (1970). *The sociology of modern Britain: an introductory reader*. Fontana, London, p. 58.
- Carrasco, M. and A. Kerne (2018). “Queer visibility: Supporting LGBT+ selective visibility on social media”. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–12.
- Cassell, J. and D. Tversky (2005). “The language of online intercultural community formation”. *Journal of Computer-Mediated Communication* 10:2, JCMC1027.
- Chancellor, S., A. Hu, and M. De Choudhury (2018). “Norms matter: Contrasting social support around behavior change in online weight loss communities”. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–14.

Bibliography

- Chancellor, S., T. Mitra, and M. De Choudhury (2016). "Recovery amid pro-anorexia: Analysis of recovery in social media". In: *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 2111–2123.
- Chancellor, S., J.A. Pater, T. Clear, E. Gilbert, and M. De Choudhury (2016). "# thyghgapp: Instagram content moderation and lexical variation in pro-eating disorder communities". In: *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, pp. 1201–1213.
- Chandrasekharan, E., U. Pavalanathan, A. Srinivasan, A. Glynn, J. Eisenstein, and E. Gilbert (2017). "You can't stay here: The efficacy of reddit's 2015 ban examined through hate speech". *Proceedings of the ACM on Human-Computer Interaction* 1:CSCW, pp. 1–22.
- Chandrasekharan, E., M. Samory, S. Jhaver, H. Charvat, A. Bruckman, C. Lampe, J. Eisenstein, and E. Gilbert (2018). "The internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro Scales". *Proceedings of the ACM on Human-Computer Interaction* 2:CSCW, pp. 1–25.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
- Cheng, J., M. Bernstein, C. Danescu-Niculescu-Mizil, and J. Leskovec (2017). "Anyone can become a troll: Causes of trolling behavior in online discussions". In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 1217–1230.
- Clark, D. B. (1973). "The concept of community: A re-examination". *The Sociological Review* 21:3, pp. 397–416.
- Coatney, M. (Sept. 24, 2019). "We Need a PBS for Social Media". *The New York Times*. URL: <https://www.nytimes.com/2019/09/24/opinion/public-broadcasting-facebook.html> (visited on 09/19/2019).
- Collins, B. and K. Tenbarger (Sept. 22, 2022). "Kiwi Farms: Anti-trans stalkers chasing Keffals around the world". URL: <https://www.nbcnews.com/tech/internet/cloudflare-kiwi-farms-keffals-anti-trans-rcna44834> (visited on 12/27/2022).
- Constone, J. (Oct. 3, 2016). "Facebook launches Marketplace, a friendlier Craigslist". *TechCrunch*. URL: <https://techcrunch.com/2016/10/03/facebook-marketplace-2/> (visited on 03/18/2023).

Bibliography

- Corbin, J. (2009). "Taking an Analytic Journey". In: *Developing Grounded Theory: The Second Generation (1st ed.)* Ed. by J. Morse, P. Stern, J. Corbin, B. Bowers, K. Charmaz, and A. Clarke. Left Coast Press, Walnut Creek, CA, USA. Chap. 3, pp. 35–53.
- Crabtree, A., T. Rodden, and S. Benford (2005). "Moving with the times: IT research and the boundaries of CSCW". *Computer Supported Cooperative Work (CSCW)* 14, pp. 217–251.
- Crawford, K. (2009). "Following you: Disciplines of listening in social media". *Continuum* 23:4, pp. 525–535.
- (2011). "Listening, not lurking: The neglected form of participation". *Cultures of participation*, pp. 63–74.
- Cunha, T., D. Jurgens, C. Tan, and D. M. Romero (2019). "Are All Successful Communities Alike? Characterizing and Predicting the Success of Online Communities". In: *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. Ed. by L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, and L. Zia. ACM, pp. 318–328. DOI: [10.1145/3308558.3313689](https://doi.org/10.1145/3308558.3313689). URL: <https://doi.org/10.1145/3308558.3313689>.
- Dabbish, L., R. Farzan, R. E. Kraut, and T. Postmes (2012). "Fresh faces in the crowd: turnover, identity, and commitment in online groups". In: *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 245–248.
- Dailey, D. and K. Starbird (Dec. 2014). "Journalists as Crowdsourcerers: Responding to Crisis by Reporting with a Crowd". *Computer Supported Cooperative Work (CSCW)* 23:4-6, pp. 445–481. ISSN: 0925-9724, 1573-7551. DOI: [10.1007/s10606-014-9208-z](https://doi.org/10.1007/s10606-014-9208-z). (Visited on 02/19/2025).
- Danescu-Niculescu-Mizil, C., R. West, D. Jurafsky, J. Leskovec, and C. Potts (2013). "No country for old members: User lifecycle and linguistic change in online communities". In: *Proceedings of the 22nd international conference on World Wide Web*, pp. 307–318.
- Datta, S. and E. Adar (2019). "Extracting inter-community conflicts in reddit". In: *Proceedings of the international AAAI conference on Web and Social Media*. Vol. 13, pp. 146–157.

Bibliography

- Davies, C., J. Ashford, L. Espinosa-Anke, A. Preece, L. Turner, R. Whitaker, M. Srivatsa, and D. Felmlee (2021). "Multi-scale user migration on reddit". In: AAAI.
- De Bruyn, A. and G. L. Lilien (2004). "A multi-stage model of word of mouth through electronic referrals". *eBRC Research Paper Series (ref. 2004-02)*.
- De Souza, C. S. and J. Preece (2004). "A framework for analyzing and understanding online communities". *Interacting with computers* 16:3, pp. 579–610.
- Diakopoulos, N. and M. Naaman (2011). "Towards quality discourse in online news comments". In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 133–142.
- Dixon, J. and K. Durrheim (2004). "Dislocating identity: Desegregation and the transformation of place". *Journal of environmental psychology* 24:4, pp. 455–473.
- Doctorow, C. (Jan. 25, 2025). *Enshittification isn't caused by venture capital*. URL: <https://pluralistic.net/2025/01/20/capitalist-unrealism/> (visited on 05/31/2026).
- Doheny-Farina, S. (1996). *The wired neighborhood*. Yale University Press.
- Doolittle, R. J. and D. MacDonald (1978). "Communication and a sense of community in a metropolitan neighborhood: A factor analytic examination". *Communication quarterly* 26:3, pp. 2–7.
- Dourish, P. (2006). "Re-space-ing place: " place" and" space" ten years on". In: *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pp. 299–308.
- Dourish, P., A. Adler, V. Bellotti, and A. Henderson (1996). "Your place or mine? Learning from long-term use of audio-video communication". *Computer Supported Cooperative Work (CSCW)* 5, pp. 33–62.
- Dovey, K. (1983). "The Creation of a Sense of Place: The Case of Preshil". *Places* 1:2.
- Ducheneaut, N., N. Yee, E. Nickell, and R. J. Moore (2007). "The life and death of online gaming communities: a look at guilds in world of warcraft". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 839–848.
- Dupre, K. (Mar. 4, 2019). "Trends and Gaps in Place-Making in the Context of Urban Development and Tourism: 25 Years of Literature Review". *Journal of Place Management and Development* 12:1, pp. 102–

Bibliography

120. ISSN: 1753-8335. DOI: [10.1108/JPMD-07-2017-0072](https://doi.org/10.1108/JPMD-07-2017-0072). URL: <https://www.emerald.com/insight/content/doi/10.1108/JPMD-07-2017-0072/full/html> (visited on 05/19/2025).
- Ellis, K., M. Goldszmidt, G. Lanckriet, N. Mishra, and O. Reingold (Feb. 2016). "Equality and Social Mobility in Twitter Discussion Groups". In: *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, San Francisco California USA, pp. 523–532. ISBN: 978-1-4503-3716-8. DOI: [10.1145/2835776.2835814](https://doi.org/10.1145/2835776.2835814). (Visited on 12/04/2024).
- Ellison, N. B., P. Tri`eu, S. Schoenebeck, R. Brewer, and A. Israni (2020). "Why we don't click: Interrogating the relationship between viewing and clicking in social media contexts by exploring the "non-click"". *Journal of Computer-Mediated Communication* 25:6, pp. 402–426.
- Facebook (Oct. 25, 2021). *Facebook Reports Third Quarter 2021 Results*. [Online; accessed 25. Jun. 2022]. URL: <https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Third-Quarter-2021-Results/default.aspx> (visited on 10/25/2021).
- Fiesler, C. (Oct. 22, 2022). "Taylor Swift 'Midnights' album gives TikTok fan groups more fodder, and community". *NBC News*. URL: <https://www.nbcnews.com/think/opinion/taylor-swift-midnights-album-gives-tiktok-fan-groups-fodder-community-rcna53504> (visited on 01/29/2023).
- Fiesler, C. and B. Dym (2020). "Moving across lands: Online platform migration in fandom communities". *Proceedings of the ACM on Human-Computer Interaction* 4:CSCW1, pp. 1–25.
- Fiesler, C., J. McCann, K. Frye, and J. R. Brubaker (2018). "Reddit rules! characterizing an ecosystem of governance". In: *Twelfth International AAAI Conference on Web and Social Media*.
- Fiore, A. T., S. L. Tiernan, and M. A. Smith (2002). "Observed behavior and perceived value of authors in usenet newsgroups: bridging the gap". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 323–330.
- Fire, M. and C. Guestrin (2020). "The rise and fall of network stars: Analyzing 2.5 million graphs to reveal how high-degree vertices emerge over time". *Information Processing & Management* 57:2, p. 102041.

Bibliography

- Fisher, D., M.A. Smith, and H. T. Welser (2006). "You are who you talk to: Detecting roles in usenet newsgroups". In: *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. Vol. 3. IEEE, 59b–59b.
- Foote, J. D. (2019). "The Formation and Growth of Collaborative Online Organizations". PhD thesis. Northwestern University.
- Foote, J. D., D. Gergle, and A. Shaw (2017). "Starting online communities: motivations and goals of wiki founders". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6376–6380.
- Freeman, G. and D. Y. Wohn (2019). "Understanding eSports Team Formation and Coordination". *Computer Supported Cooperative Work* 28, pp. 95–126.
- Friedmann, J. (June 2010). "Place and Place-Making in Cities: A Global Perspective". *Planning Theory & Practice* 11:2, pp. 149–165. ISSN: 1464-9357, 1470-000X. DOI: [10.1080/14649351003759573](https://doi.org/10.1080/14649351003759573). URL: <http://www.tandfonline.com/doi/abs/10.1080/14649351003759573> (visited on 05/04/2025).
- Gaffney, D. and J. N. Matias (2018). "Caveat emptor, computational social science: Large-scale missing data in a widely-published Reddit corpus". *PloS one* 13:7.
- Galegher, J., L. Sproull, and S. Kiesler (1998). "Legitimacy, authority, and community in electronic support groups". *Written communication* 15:4, pp. 493–530.
- Ganley, D. and C. Lampe (2009). "The ties that bind: Social network principles in online communities". *Decision support systems* 47:3, pp. 266–274.
- Gaudiano, N. (July 2015). "Reddit boosts momentum for Sanders' campaign". *USA Today*. URL: <https://www.usatoday.com/story/news/politics/elections/2016/2015/07/24/reddit-boosts-momentum-sanders-campaign/30614507/> (visited on 12/05/2024).
- Gazan, R. (2009). "When online communities become self-aware". In: *2009 42nd Hawaii International Conference on System Sciences*. IEEE, pp. 1–10.
- Geiger, R. S. (2018). "The lives of bots". *arXiv preprint arXiv:1810.09590*.

Bibliography

- Geiger, R. S. and D. Ribes (2010). "The work of sustaining order in Wikipedia: The banning of a vandal". In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 117–126.
- (2011). "Trace ethnography: Following coordination through documentary practices". In: *2011 44th Hawaii international conference on system sciences*. IEEE, pp. 1–10.
- Gilbert, E. (2013). "Widespread underprovision on Reddit". In: *Proceedings of the 2013 conference on Computer supported cooperative work*, pp. 803–808.
- Gilbert, S. A. (May 28, 2020). "'I Run the World's Largest Historical Outreach Project and It's on a Cesspool of a Website.' Moderating a Public Scholarship Site on Reddit: A Case Study of r/AskHistorians". *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1, pp. 1–27. ISSN: 2573-0142. DOI: [10.1145/3392822](https://doi.org/10.1145/3392822). URL: <https://dl.acm.org/doi/10.1145/3392822> (visited on 05/06/2025).
- Gilbert, S. A. (2018). "Motivations for participating in online initiatives: exploring motivations across initiative types". PhD thesis. University of British Columbia.
- Gillespie, T. (2010). "The Politics of 'Platforms'". *New media & society* 12:3, pp. 347–364.
- Gini, C. (1912). *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche*. [Fasc. I.] Tipogr. di P. Cuppini.
- Glaser, B. G. and A. L. Strauss (1967). *Discovery of grounded theory: Strategies for qualitative research*. Aldine Publishing. ISBN: 0202302601.
- Golder, S. A. and J. Donath (2004). "Social roles in electronic communities". *Internet Research* 5:1, pp. 19–22.
- González-Bailón, S., J. Borge-Holthoefer, A. Rivero, and Y. Moreno (Dec. 2011). "The Dynamics of Protest Recruitment through an Online Network". *Scientific Reports* 1:1, p. 197. ISSN: 2045-2322. DOI: [10.1038/srep00197](https://doi.org/10.1038/srep00197). (Visited on 02/14/2025).
- Greif, I. (2019). "How we started CSCW". *Nature Electronics* 2:3, pp. 132–132.
- Hamilton, W. L., J. Zhang, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec (2017). "Loyalty in online communities". In: *Eleventh International AAAI Conference on Web and Social Media*.

Bibliography

- Harper, F. M., J. Weinberg, J. Logie, and J. A. Konstan (July 4, 2010). "Question Types in Social Q&A Sites". *First Monday*. ISSN: 1396-0466. DOI: [10.5210/fm.v15i7.2913](https://doi.org/10.5210/fm.v15i7.2913). URL: <https://journals.uic.edu/ojs/index.php/fm/article/view/2913> (visited on 05/17/2025).
- Harris, A. M., D. Gómez-Zarà, L. A. DeChurch, and N. S. Contractor (2019). "Joining together online: the trajectory of CSCW scholarship on group formation". *Proceedings of the ACM on Human-Computer Interaction* 3:CSCW, pp. 1–27.
- Harrison, S. and P. Dourish (1996). "Re-place-ing space: the roles of place and space in collaborative systems". In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pp. 67–76.
- Haythornthwaite, C. (2002). "Strong, weak, and latent ties and the impact of new media". *The information society* 18:5, pp. 385–401.
- Herrman, J. (Apr. 2016). *Donald Trump Finds Support in Reddit's Unruly Corners*. [Online; accessed 21. Jun. 2024]. URL: http://www.nytimes.com/2016/04/09/business/media/in-reddits-unruly-corners-trump-finds-support.html?_r=0 (visited on 04/08/2016).
- Hessel, J., C. Tan, and L. Lee (2016). "Science, askscience, and badscience: On the coexistence of highly related communities". In: *Proceedings of the international AAAI conference on web and social media*. Vol. 10. 1, pp. 171–180.
- Hillery, G. A. (1955). "Definition of community". *Rural sociology* 20, pp. 111–123.
- (1959). "A critique of selected community concepts". *Social Forces*, pp. 237–242.
- Hillman, S., J. Procyk, and C. Neustaedter (2014). "'alksjdf; Lksfd' tumblr and the fandom user experience". In: *Proceedings of the 2014 conference on Designing interactive systems*, pp. 775–784.
- Ho, T. K. (1995). "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE, pp. 278–282.
- Homans, G. C. (1951). *The human group*. Routledge.

Bibliography

- Horne, B. D., S. Adali, and S. Sikdar (2017). "Identifying the social signals that drive online discussions: A case study of Reddit communities". In: *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, pp. 1–9.
- hottest page* (Dec. 8, 2005). URL: <https://web.archive.org/web/20161221175624/https://redditblog.com/2005/12/08/hottest-page> (visited on 12/21/2016).
- Hwang, S. and J. D. Foote (2021). "Why do people participate in small online communities?" *Proceedings of the ACM on Human-Computer Interaction* 5:CSCW2, pp. 1–25.
- Iriberri, A. and G. Leroy (2009). "A life-cycle perspective on online community success". *ACM Computing Surveys (CSUR)* 41:2, pp. 1–29.
- Jhaver, S., D. S. Appling, E. Gilbert, and A. Bruckman (2019). "'Did You Suspect the Post Would be Removed?' Understanding User Reactions to Content Removals on Reddit". *Proceedings of the ACM on human-computer interaction* 3:CSCW, pp. 1–33.
- Jhaver, S., I. Birman, E. Gilbert, and A. Bruckman (2019). "Human-machine collaboration for content regulation: The case of Reddit Automoderator". *ACM Transactions on Computer-Human Interaction (TOCHI)* 26:5, pp. 1–35.
- Jhaver, S., A. Bruckman, and E. Gilbert (2019). "Does transparency in moderation really matter? User behavior after content removal explanations on reddit". *Proceedings of the ACM on Human-Computer Interaction* 3:CSCW, pp. 1–27.
- Jhaver, S., L. Chan, and A. Bruckman (2018). "The view from the other side: The border between controversial speech and harassment on Kotaku in Action". *First Monday* 23:2, p. 4.
- Johnson, S. L., S. Faraj, and S. Kudaravalli (2014). "Emergence of power laws in online communities". *Mis Quarterly* 38:3, 795–A13.
- Johnston, B. (Mar. 2007). *Online Community Metrics - Best Practices Survey*. San Francisco CA USA: Forum One Communications. URL: https://web.archive.org/web/20090219182916/http://www.onlinecommunityreport.com:80/reports/oc_metrics2007_final.pdf (visited on 04/30/2025).

Bibliography

- Johnston, C. (June 21, 2012). *reddit founders made hundreds of fake profiles so site looked popular*. [Online; accessed 23. Aug. 2024]. URL: <https://arstechnica.com/information-technology/2012/06/reddit-founders-made-hundreds-of-fake-profiles-so-site-looked-popular> (visited on 08/23/2024).
- Jones, Q. (Dec. 1997). "Virtual-Communities, Virtual Settlements & Cyber-Archaeology: a Theoretical Outline". *Journal of Computer-Mediated Communication* 3:3. JCMC331. ISSN: 1083-6101. DOI: [10.1111/j.1083-6101.1997.tb00075.x](https://doi.org/10.1111/j.1083-6101.1997.tb00075.x). URL: <https://doi.org/10.1111/j.1083-6101.1997.tb00075.x>.
- Jones, Q., G. Ravid, and S. Rafaeli (2004). "Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration". *Information systems research* 15:2, pp. 194–210.
- Joyce, E. and R. E. Kraut (2006). "Predicting continued participation in newsgroups". *Journal of Computer-Mediated Communication* 11:3, pp. 723–747.
- Kairam, S. R. and J. D. Foote (2024). "How founder motivations, goals, and actions influence early trajectories of online communities". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–11.
- Kairam, S. R., D. J. Wang, and J. Leskovec (Feb. 8, 2012). "The Life and Death of Online Groups: Predicting Group Growth and Longevity". In: *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*. WSDM'12: Fifth ACM International Conference on Web Search and Data Mining. ACM, Seattle Washington USA, pp. 673–682. ISBN: 978-1-4503-0747-5. DOI: [10.1145/2124295.2124374](https://doi.org/10.1145/2124295.2124374). URL: <https://dl.acm.org/doi/10.1145/2124295.2124374> (visited on 05/27/2026).
- Kang, K. W. and S. K. Shin (2006). "A Model for the Antecedents of Member Loyalty: The Case of USENET Newsgroup". *Proceedings of the Twelfth Americas Conference on Information Systems*. URL: <https://aisel.aisnet.org/amcis2006/533/>.
- Katz, J. (1998). *Luring the Lurkers*. URL: <https://news.slashdot.org/story/98/12/28/1745252/luring-the-lurkers> (visited on 09/24/2022).

Bibliography

- Kiene, C., A. Monroy-Hernández, and B. M. Hill (2016). "Surviving an" Eternal September" How an Online Community Managed a Surge of Newcomers". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 1152–1156.
- Kingod, N., B. Cleal, A. Wahlberg, and G. R. Husted (2017). "Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review". *Qualitative health research* 27:1, pp. 89–99.
- Klandermans, B. and D. Oegema (Aug. 1987). "Potentials, Networks, Motivations, and Barriers: Steps Towards Participation in Social Movements". *American Sociological Review* 52:4, p. 519. ISSN: 00031224. DOI: [10.2307/2095297](https://doi.org/10.2307/2095297). JSTOR: [2095297](https://www.jstor.org/stable/2095297). (Visited on 02/11/2025).
- Klassen, S., S. Kingsley, K. McCall, J. Weinberg, and C. Fiesler (2021). "More than a modern day Green book: Exploring the online community of Black Twitter". *Proceedings of the ACM on Human-Computer Interaction* 5:CSCW2, pp. 1–29.
- Koebler, J. (July 12, 2016). *How r/the_donald Became a Melting Pot of Frustration and Hate*. [Online; accessed 21. Jun. 2024]. URL: <http://motherboard.vice.com/read/what-is-rthedonald-donald-trump-subreddit> (visited on 06/21/2024).
- Koh, J., Y.-G. Kim, and Y.-G. Kim (Dec. 2003). "Sense of Virtual Community: A Conceptual Framework and Empirical Validation". *International Journal of Electronic Commerce* 8:2, pp. 75–94. ISSN: 1086-4415, 1557-9301. DOI: [10.1080/10864415.2003.11044295](https://doi.org/10.1080/10864415.2003.11044295). URL: <https://www.tandfonline.com/doi/full/10.1080/10864415.2003.11044295> (visited on 11/29/2025).
- Kollock, P. (2001). "The economies of online cooperation: Gifts and public goods in cyberspace". In: *Communities in cyberspace*. Ed. by P. Kollock and M. A. Smith. Routledge, London, UK, pp. 220–239.
- Kollock, P. and M. A. Smith (1996). "Managing the virtual commons". In: *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives*. Ed. by S. C. Herring. John Benjamins Philadelphia, PA, pp. 109–128.
- Kraut, R. E. and A. T. Fiore (2014). "The role of founders in building online groups". In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 722–732.

Bibliography

- Kraut, R. E., R. S. Fish, R. W. Root, B. L. Chalfonte, et al. (1990). "Informal communication in organizations: Form, function, and technology". In: *Human reactions to technology: Claremont symposium on applied social psychology*, pp. 145–199.
- Kraut, R. E., J. M. Levine, M. M. Escobar, and A. Herdağdelen (2020). "What makes people feel close to online groups? The roles of group attributes and group types". In: *Proceedings of the International AAI Conference on Web and Social Media*. Vol. 14, pp. 382–392.
- Kraut, R. E. and P. Resnick (2012). *Building successful online communities: Evidence-based social design*. Mit Press.
- Krohn, R. and T. Weninger (2022). "Subreddit Links Drive Community Creation and User Engagement on Reddit". In: *Proceedings of the International AAI Conference on Web and Social Media*. Vol. 16, pp. 536–547.
- Kuhn, M. (2008). "Building Predictive Models in R Using the caret Package". *Journal of Statistical Software* 28:5, pp. 1–26. DOI: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05). URL: <https://www.jstatsoft.org/index.php/jss/article/view/v028i05>.
- Kumar, S., W. L. Hamilton, J. Leskovec, and D. Jurafsky (2018). "Community Interaction and Conflict on the Web". In: *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*. The 2018 World Wide Web Conference. ACM Press, Lyon, France, pp. 933–943. ISBN: 978-1-4503-5639-8. DOI: [10.1145/3178876.3186141](https://doi.org/10.1145/3178876.3186141). URL: <http://dl.acm.org/citation.cfm?doid=3178876.3186141> (visited on 11/16/2025).
- Lagorio-Chafkin, C. (Nov. 2016). *Reddit and the God Emperor of the Internet*. [Online; accessed 21. Jun. 2024]. URL: <https://web.archive.org/web/20161124074424/http://www.nytimes.com/2016/11/20/opinion/sunday/reddit-and-the-god-emperor-of-the-internet.html> (visited on 11/19/2016).
- Lampe, C. and E. Johnston (2005). "Follow the (Slash) Dot: Effects of Feedback on New Members in an Online Community". In: *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work - GROUP '05*. The 2005 International ACM SIGGROUP Conference. ACM

Bibliography

- Press, Sanibel Island, Florida, USA, p. 11. ISBN: 978-1-59593-223-5. DOI: [10.1145/1099203.1099206](https://doi.org/10.1145/1099203.1099206). URL: <http://portal.acm.org/citation.cfm?doid=1099203.1099206> (visited on 12/04/2024).
- Lampe, C., R. Wash, A. Velasquez, and E. Ozkaya (Apr. 10, 2010). "Motivations to Participate in Online Communities". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10: CHI Conference on Human Factors in Computing Systems. ACM, Atlanta Georgia USA, pp. 1927–1936. ISBN: 978-1-60558-929-9. DOI: [10.1145/1753326.1753616](https://doi.org/10.1145/1753326.1753616). URL: <https://dl.acm.org/doi/10.1145/1753326.1753616> (visited on 12/04/2024).
- Lampe, C., P. Zube, J. Lee, C. H. Park, and E. Johnston (2014). "Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums". *Government Information Quarterly* 31:2, pp. 317–326.
- Lampe, C. A., E. Johnston, and P. Resnick (Apr. 29, 2007). "Follow the Reader: Filtering Comments on Slashdot". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI07: CHI Conference on Human Factors in Computing Systems. ACM, San Jose California USA, pp. 1253–1262. ISBN: 978-1-59593-593-9. DOI: [10.1145/1240624.1240815](https://doi.org/10.1145/1240624.1240815). URL: <https://dl.acm.org/doi/10.1145/1240624.1240815> (visited on 11/16/2025).
- Lazar, J. and J. Preece (1998). "Classification schema for online communities". *AMCIS 1998 Proceedings*, p. 30.
- Leavitt, A. (2015). "'This is a Throwaway Account' Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community". In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 317–327.
- Leavitt, A. and J. A. Clark (2014). "Upvoting hurricane Sandy: event-based news production processes on a social news site". In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1495–1504.
- Leavitt, A. and J. J. Robinson (2017a). "The role of information visibility in network gatekeeping: Information aggregation on Reddit during crisis events". In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 1246–1261.

Bibliography

- Leavitt, A. and J. J. Robinson (2017b). “Upvote My News: The Practices of Peer Information Aggregation for Breaking News on reddit. com”. *Proceedings of the ACM on Human-Computer Interaction* 1:CSCW, pp. 1–18.
- Lee, C. P., P. Dourish, and G. Mark (Nov. 4, 2006). “The Human Infrastructure of Cyberinfrastructure”. In: *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*. CSCW06: Computer Supported Cooperative Work. ACM, Banff Alberta Canada, pp. 483–492. ISBN: 978-1-59593-249-5. DOI: [10.1145/1180875.1180950](https://doi.org/10.1145/1180875.1180950). URL: <https://dl.acm.org/doi/10.1145/1180875.1180950> (visited on 05/30/2026).
- Lee, C. P. and D. Paine (Feb. 28, 2015). “From The Matrix to a Model of Coordinated Action (MoCA): A Conceptual Framework of and for CSCW”. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW ’15: Computer Supported Cooperative Work and Social Computing. ACM, Vancouver BC Canada, pp. 179–194. ISBN: 978-1-4503-2922-4. DOI: [10.1145/2675133.2675161](https://doi.org/10.1145/2675133.2675161). URL: <https://dl.acm.org/doi/10.1145/2675133.2675161> (visited on 11/23/2025).
- Leimeister, J. M., P. Sidiras, and H. Krcmar (2006). “Exploring success factors of virtual communities: The perspectives of members and operators”. *Journal of organizational computing and electronic commerce* 16:3-4, pp. 279–300.
- Leskovec, J., L. Backstrom, R. Kumar, and A. Tomkins (2008). “Microscopic evolution of social networks”. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 462–470.
- Li, H., B. Hecht, and S. Chancellor (May 31, 2022). “All That’s Happening behind the Scenes: Putting the Spotlight on Volunteer Moderator Labor in Reddit”. *Proceedings of the International AAAI Conference on Web and Social Media* 16, pp. 584–595. ISSN: 2334-0770, 2162-3449. DOI: [10.1609/icwsm.v16i1.19317](https://doi.org/10.1609/icwsm.v16i1.19317). URL: <https://ojs.aaai.org/index.php/ICWSM/article/view/19317> (visited on 05/17/2025).

Bibliography

- Liang, Y. and J. Introne (Feb. 2019). "Social Roles, Interactions and Community Sustainability in Social Q&A Sites: A Resource-based Perspective". In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. DOI: [10.24251/HICSS.2019.337](https://doi.org/10.24251/HICSS.2019.337).
- Licklider, J. C. and R. W. Taylor (1968). "The computer as a Communication Device". *Science and Technology* 76:2, pp. 1–3.
- Lieb, A., A. Chapekis, S. Shah, and A. Smith (Dec. 11, 2025). "How Parents Use Online Communities". URL: <https://www.pewresearch.org/data-labs/2025/11/12/how-parents-use-online-communities/> (visited on 05/22/2026).
- Lin, Z., N. Salehi, B. Yao, Y. Chen, and M. Bernstein (2017). "Better when it was smaller? community content and behavior after massive growth". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 11. 1, pp. 132–141.
- Litt, E., S. Zhao, R. Kraut, and M. Burke (July 2020). "What Are Meaningful Social Interactions in Today's Media Landscape? A Cross-Cultural Survey". *Social Media + Society* 6:3. ISSN: 2056-3051, 2056-3051. DOI: [10.1177/2056305120942888](https://doi.org/10.1177/2056305120942888). URL: <https://journals.sagepub.com/doi/10.1177/2056305120942888> (visited on 05/23/2025).
- Liu, Z. and B. J. Jansen (Apr. 18, 2015). "A Taxonomy for Classifying Questions Asked in Social Question and Answering". In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. CHI '15: CHI Conference on Human Factors in Computing Systems. ACM, Seoul Republic of Korea, pp. 1947–1952. ISBN: 978-1-4503-3146-3. DOI: [10.1145/2702613.2732928](https://doi.org/10.1145/2702613.2732928). URL: <https://dl.acm.org/doi/10.1145/2702613.2732928> (visited on 05/17/2025).
- Lo, C. C. W. Y. (2018). "When all you have is a banhammer: the social and communicative work of Volunteer moderators". PhD thesis. Massachusetts Institute of Technology.
- MacIver, R. M. and C. H. Page (1959). *Society: An introductory analysis*. Macmillan & Co Ltd, London.
- Martin, T. (2017). "community2vec: Vector representations of online communities encode semantic relationships". In: *Proceedings of the Second Workshop on NLP and Computational Social Science*, pp. 27–31.

Bibliography

- Marwick, A. and R. Lewis (2017). “Media manipulation and disinformation online”. *New York: Data & Society Research Institute*, pp. 7–19.
- Mason, B. (1999). “Issues in virtual ethnography”. In: *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities*, pp. 61–69. URL: <https://web.archive.org/web/20040102005340/http://www.ucs.mun.ca/~bmason/pubs/lime99.pdf>.
- Massanari, A. (Mar. 2017). “#Gamergate and The Fappening: How Reddit’s Algorithm, Governance, and Culture Support Toxic Technocultures”. *New Media & Society* 19:3, pp. 329–346. ISSN: 1461-4448, 1461-7315. DOI: [10.1177/1461444815608807](https://doi.org/10.1177/1461444815608807). URL: <https://journals.sagepub.com/doi/10.1177/1461444815608807> (visited on 12/04/2024).
- Massimi, M., J. L. Bender, H. O. Witteman, and O. H. Ahmed (2014). “Life transitions and online health communities: reflecting on adoption, use, and disengagement”. In: *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pp. 1491–1501.
- Mastroianni, A. M. and D. T. Gilbert (2023). “The illusion of moral decline”. *Nature* 618:7966, pp. 782–789.
- Matias, J. N. (May 7, 2016). “Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI’16: CHI Conference on Human Factors in Computing Systems. ACM, San Jose California USA, pp. 1138–1151. ISBN: 978-1-4503-3362-7. DOI: [10.1145/2858036.2858391](https://doi.org/10.1145/2858036.2858391). URL: <https://dl.acm.org/doi/10.1145/2858036.2858391> (visited on 12/03/2025).
- (Apr. 2019). “The Civic Labor of Volunteer Moderators Online”. *Social Media + Society* 5:2, p. 2056305119836778. ISSN: 2056-3051, 2056-3051. DOI: [10.1177/2056305119836778](https://doi.org/10.1177/2056305119836778). URL: <https://journals.sagepub.com/doi/10.1177/2056305119836778> (visited on 12/04/2024).
- Matthews, T., J. Chen, S. Whittaker, A. Pal, H. Zhu, H. Badenes, and B. Smith (2014). “Goals and perceived success of online enterprise communities: what is important to leaders & members?” In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 291–300.

Bibliography

- McAdam, D. (July 1986). "Recruitment to High-Risk Activism: The Case of Freedom Summer". *American Journal of Sociology* 92:1, pp. 64–90. ISSN: 0002-9602, 1537-5390. DOI: [10.1086/228463](https://doi.org/10.1086/228463). (Visited on 02/11/2025).
- McGlohon, M. and M. Hurst (2009). "Community structure and information flow in USENET: Improving analysis with a thread ownership model". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 3. 1, pp. 122–129.
- McMillan, D. W. and D. M. Chavis (Jan. 1986). "Sense of Community: A Definition and Theory". *Journal of Community Psychology* 14:1, pp. 6–23. ISSN: 00904392, 15206629. DOI: [10.1002/1520-6629\(198601\)14:1<6::AID-JCOP2290140103>3.0.CO;2-I](https://doi.org/10.1002/1520-6629(198601)14:1<6::AID-JCOP2290140103>3.0.CO;2-I). URL: [https://onlinelibrary.wiley.com/doi/10.1002/1520-6629\(198601\)14:1%3C6::AID-JCOP2290140103%3E3.0.CO;2-I](https://onlinelibrary.wiley.com/doi/10.1002/1520-6629(198601)14:1%3C6::AID-JCOP2290140103%3E3.0.CO;2-I) (visited on 11/29/2025).
- Mensah, H., L. Xiao, and S. Soundarajan (2020). "Characterizing the Evolution of Communities on Reddit". In: *International Conference on Social Media and Society*, pp. 58–64.
- Mills, R. A. (2018). "Pop-up political advocacy communities on Reddit. com: SandersForPresident and The Donald". *AI & SOCIETY* 33:1, pp. 39–54.
- Milton, A., L. Ajmani, M. A. DeVito, and S. Chancellor (2023). "'I See Me Here': Mental Health Content, Community, and Algorithmic Curation on TikTok". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–17.
- Moore, T. D. and M. A. Serva (2007). "Understanding member motivation for contributing to different types of virtual communities: a proposed framework". In: *Proceedings of the 2007 ACM SIGMIS CPR conference on Computer personnel research: The global information technology workforce*, pp. 153–158.
- Morse, J. M. (2009). "Tussles, Tensions, and Resolutions". In: *Developing Grounded Theory: The Second Generation (1st ed.)* Ed. by J. Morse, P. Stern, J. Corbin, B. Bowers, K. Charmaz, and A. Clarke. Left Coast Press, Walnut Creek, CA, USA. Chap. 1, pp. 13–22.

Bibliography

- Nelson, L. K. (2020). "Computational Grounded Theory: A Methodological Framework". *Sociological Methods & Research* 49:1, pp. 3–42. DOI: [10.1177/0049124117729703](https://doi.org/10.1177/0049124117729703). eprint: <https://doi.org/10.1177/0049124117729703>. URL: <https://doi.org/10.1177/0049124117729703>.
- New Sharing Features on Instagram: Notes, Group Profiles and More* (Dec. 2022). [Online; accessed 22. Feb. 2024]. URL: <https://about.fb.com/news/2022/12/sharing-features-on-instagram-notes-group-profiles-and-more>.
- Newell, E., D. Jurgens, H. Saleem, H. Vala, J. Sassine, C. Armstrong, and D. Ruths (2016). "User migration in online social networks: A case study on reddit during a period of community unrest". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 10. 1, pp. 279–288.
- Newton, C. (June 2020). "Reddit bans r/The_Donald and r/ChapoTrapHouse as part of a major expansion of its rules". *The Verge*. URL: <https://www.theverge.com/2020/6/29/21304947/reddit-ban-subreddits-the-donald-chapo-trap-house-new-content-policy-rules> (visited on 12/27/2022).
- Nguyen, D. and C. P. Rosé (2011). "Language use as a reflection of socialization in online communities". In: *Proceedings of the workshop on languages in social media*. Association for Computational Linguistics, pp. 76–85.
- Nonnecke, B., D. Andrews, and J. Preece (2006). "Non-public and public online community participation: Needs, attitudes and behavior". *Electronic Commerce Research* 6:1, pp. 7–20.
- Nonnecke, B. and J. Preece (1999). "Shedding light on lurkers in online communities". *Ethnographic studies in real and virtual environments: Inhabited information spaces and connected communities, Edinburgh* 123128.
- Nonnecke, B. and J. Preece (2000). "Lurker demographics: Counting the silent". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 73–80.
- (2001). "Why lurkers lurk". *AMCIS 2001 proceedings*, p. 294.
- Nov, O. (2007). "What motivates wikipedians?" *Communications of the ACM* 50:11, pp. 60–64.
- Noveck, B. S., J. Button, D. Gambrell, L. Paulson, P. Spada, and L. Subramanian (2021). "The power of virtual communities".

Bibliography

- O'Hallarn, B., M. Slavich, and B. Emmons (2023). "“I Used to Love Scheifele:” Dominant Narratives on Reddit About the Canadian Division of the Stanley Cup Playoffs". *Communication & Sport* 11:6, pp. 1203–1222.
- Oberschall, A. (1973). *Social Conflict and Social Movements*. Prentice-Hall Series in Sociology. Prentice-Hall, Englewood Cliffs., N.J. ISBN: 978-0-13-815761-6.
- Ohlheiser, A. (June 13, 2016). "Redditors Checked r/News for Updates on the Orlando Shooting. Instead, They Found a War." *Washington Post*. URL: <https://www.washingtonpost.com/news/the-intersect/wp/2016/06/13/redditors-checked-rnews-for-updates-on-the-orlando-shooting-instead-they-found-a-war/> (visited on 03/04/2025).
- Olteanu, A., C. Castillo, F. Diaz, and E. Kıcıman (2019). "Social data: Biases, methodological pitfalls, and ethical boundaries". *Frontiers in Big Data* 2, p. 13.
- Palinkas, L. A., S. M. Horwitz, C. A. Green, J. P. Wisdom, N. Duan, and K. Hoagwood (Sept. 2015). "Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research". *Adm Policy Ment Health* 42:5, pp. 533–544.
- Panek, E., C. Hollenbach, J. Yang, and T. Rhodes (Oct. 2018). "The Effects of Group Size and Time on the Formation of Online Communities: Evidence From Reddit". *Social Media + Society* 4:4, p. 2056305118815908. ISSN: 2056-3051, 2056-3051. DOI: [10.1177/2056305118815908](https://doi.org/10.1177/2056305118815908). URL: <https://journals.sagepub.com/doi/10.1177/2056305118815908> (visited on 12/04/2024).
- Park, A., M. Conway, and A. T. Chen (2018). "Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach". *Computers in human behavior* 78, pp. 98–112.
- Peck, R. (Aug. 3, 2020). "The Hate-Fueled Rise of r/The_Donald—and Its Epic Takedown". *WIRED*. URL: <https://www.wired.com/story/the-hate-fueled-rise-of-rthe-donald-and-its-epic-takedown> (visited on 02/26/2023).

Bibliography

- Ploderer, B., S. Howard, and P. Thomas (2008). "Being online, living offline: the influence of social ties over the appropriation of social network sites". In: *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 333–342.
- Poplin, D. E. (1972). *Communities: A survey of theories and methods of research*. Macmillan.
- Preece, J. (2000). *Online communities: Designing usability and supporting socialbilty*. John Wiley & Sons, Inc.
- Preece, J., B. Nonnecke, and D. Andrews (2004). "The top five reasons for lurking: improving community experiences for everyone". *Computers in human behavior* 20:2, pp. 201–223.
- Prinster, G. H., C. E. Smith, C. Tan, and B. C. Keegan (2024). "Community Archetypes: An Empirical Framework for Guiding Research Methodologies to Reflect User Experiences of Sense of Virtual Community on Reddit". *Proceedings of the ACM on Human-Computer Interaction* 8:CSCW1, pp. 1–33.
- Proferes, N., N. Jones, S. Gilbert, C. Fiesler, and M. Zimmer (2021). "Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics". *Social Media+ Society* 7:2, p. 20563051211019004.
- Putnam, R. D. (2001). *Bowling Alone: The Collapse and Revival of American Community*. 1. Touchstone ed. A Touchstone Book. Simon & Schuster [u.a.], London. ISBN: 978-0-684-83283-8.
- Raban, D. R., M. Moldovan, and Q. Jones (2010). "An empirical study of critical mass and online community survival". In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pp. 71–80.
- Rafaeli, S., G. Ravid, and V. Soroka (2004). "De-lurking in virtual communities: A social communication network approach to measuring the effects of social and cultural capital". In: *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*. IEEE, 10–pp.
- Ransbotham, S. and G. C. Kane (2011). "Membership turnover and collaboration success in online communities: Explaining rises and falls from grace in Wikipedia". *MIS quarterly*, pp. 613–627.
- Reddit (Nov. 2023a). *Growing your community*. [Online; accessed 2. May 2024]. URL: <https://support.reddithelp.com/hc/en-us/articles/15484256976148-Growing-your-community> (visited on 11/28/2023).

Bibliography

- Reddit (Nov. 2023b). *How to Crosspost your community's content to other communities*. [Online; accessed 3. May 2024]. URL: <https://support.reddithelp.com/hc/en-us/articles/15484385684628-How-to-Crosspost-your-community-s-content-to-other-communities> (visited on 11/28/2023).
- (Nov. 2023c). *Seeding content for a new community*. [Online; accessed 2. May 2024]. URL: <https://support.reddithelp.com/hc/en-us/articles/15484260579348-Seeding-content-for-a-new-community> (visited on 11/28/2023).
- (Feb. 2025). *What constitutes spam? Am I a spammer?* [Online; accessed 2. May 2024]. URL: <https://support.reddithelp.com/hc/en-us/articles/360043504051-What-constitutes-spam-Am-I-a-spammer> (visited on 02/09/2024).
- (Apr. 30, 2026). *Homepage*. URL: <https://web.archive.org/web/20260430124512/https://redditinc.com/> (visited on 05/29/2026).
- reddit (Oct. 7, 2016). *JSON*. [Online; accessed 8. Jun. 2024]. URL: <https://github.com/reddit-archive/reddit/wiki/JSON> (visited on 06/08/2024).
- Reddit and GlobalWebIndex (Jan. 2020). *The era of We and the rise of online communities*. URL: <https://www.redditinc.com/assets/case-studies/TheEraOfWe.1.6.20.pdf> (visited on 11/01/2022).
- reddit is so normal* (Mar. 16, 2008). [Online; accessed 24. Apr. 2024]. URL: <https://web.archive.org/web/20180806012645/https://redditblog.com/2008/03/16/reddit-is-so-normal> (visited on 08/06/2018).
- reddit.com (July 2015). *reddit.com: about reddit*. URL: <http://web.archive.org/web/20150703012219/http://www.reddit.com/about> (visited on 03/29/2020).
- Reid, E. (1993). "Electronic chat: Social issues on internet relay chat". *Media Information Australia* 67:1, pp. 62–70.
- Rheingold, H. (1993). *The virtual community: Finding connection in a computerized world*. Addison-Wesley Longman Publishing Co., Inc.
- Ridings, C. M. and D. Gefen (2004). "Virtual community attraction: Why people hang out online". *Journal of Computer-mediated communication* 10:1, JCMC10110.

Bibliography

- Ridings, C. M., D. Gefen, and B. Arinze (2002). "Some antecedents and effects of trust in virtual communities". *The journal of strategic information systems* 11:3-4, pp. 271–295.
- Roberts, S. T. (2016). "Commercial content moderation: Digital laborers' dirty work". Ed. by S. U. Noble and B. Tynes.
- Roberts, T. L. (1998). "Are newsgroups virtual communities?" In: *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 360–367.
- Rothaermel, F. T. and S. Sugiyama (June 2001). "Virtual Internet Communities and Commercial Success: Individual and Community-Level Theory Grounded in the Atypical Case of TimeZone.Com". *Journal of Management* 27:3, pp. 297–312. ISSN: 0149-2063, 1557-1211. DOI: [10.1177/014920630102700305](https://doi.org/10.1177/014920630102700305). URL: <https://journals.sagepub.com/doi/10.1177/014920630102700305> (visited on 11/19/2025).
- Rotman, D. and J. Preece (2010). "The 'WeTube' in YouTube—creating an online community through video sharing". *International Journal of Web Based Communities* 6:3, pp. 317–333.
- Rotman, D., J. Preece, Y. He, and A. Druin (2012). "Extreme ethnography: Challenges for research in large scale online environments". In: *Proceedings of the 2012 iConference*, pp. 207–214.
- Ruddick, S. M. (May 22, 2014). "Heterotopias of the Homeless: Strategies and Tactics of Place-making in Los Angeles". In: *Young and Homeless In Hollywood*. 0th ed. Routledge, pp. 63–78. ISBN: 978-1-315-86645-1. DOI: [10.4324/9781315866451-11](https://doi.org/10.4324/9781315866451-11). URL: <https://www.taylorfrancis.com/books/9781317960751/chapters/10.4324/9781315866451-11> (visited on 05/21/2025).
- Sanderson, D. and R. A. Polson (1939). *Rural community organization*. Wiley & Sons., New York.
- Schmidt, K. (2011). "The concept of 'work' in CSCW". *Computer Supported Cooperative Work (CSCW)* 20:4, pp. 341–401.
- Schmidt, K. and L. Bannon (1992). "Taking CSCW seriously". *Computer Supported Cooperative Work (CSCW)* 1:1, pp. 7–40.
- Schoberth, T., J. Preece, and A. Heinzl (2003). "Online communities: A longitudinal analysis of communication activities". In: *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. IEEE, 10–pp.

Bibliography

- Schulman, R. and R. Mir (Dec. 31, 2023). *Taking Back the Web with Decentralization: 2023 in Review*. URL: <https://www.eff.org/deeplinks/2023/12/taking-back-web-decentralization-2023-review> (visited on 05/31/2026).
- Seering, J. and S. R. Kairam (2023). “Who moderates on Twitch and what do they do? Quantifying practices in community moderation on Twitch”. *Proceedings of the ACM on Human-Computer Interaction* 7:GROUP, pp. 1–18.
- Seering, J., T. Wang, J. Yoon, and G. Kaufman (2019). “Moderator engagement and community development in the age of algorithms”. *New media & society* 21:7, pp. 1417–1443.
- Semrush (Dec. 1, 2022). *reddit.com Website Traffic, Ranking, Analytics [October 2022]*. URL: <https://web.archive.org/web/20221201230151/https://www.semrush.com/website/reddit.com/overview/> (visited on 12/11/2022).
- Shrestha, N., T. Barik, and C. Parnin (2021). “Remote, but Connected: How# TidyTuesday Provides an Online Community of Practice for Data Scientists.” *Proceedings of the ACM on Human-Computer Interaction* 5:CSCW1, pp. 1–31.
- shuffman56 (Feb. 2006a). *crossposting*. [Online; accessed 2. May 2024]. URL: <https://web.archive.org/web/20161221203835/https://redditblog.com/2006/02/24/crossposting> (visited on 02/24/2006).
- (Jan. 2006b). *for those of you with a private office...* [Online; accessed 31. Mar. 2020]. URL: <https://redditblog.com/2006/01/19/for-those-of-you-with-a-private-office/>.
- Silverman, C., C. Timberg, J. Kao, and J. B. Merrill (Jan. 4, 2022). *Facebook Hosted Surge of Misinformation and Insurrection Threats in Months Leading Up to Jan. 6 Attack, Records Show*. [Online; accessed 15. Nov. 2022]. URL: <https://www.propublica.org/article/facebook-hosted-surge-of-misinformation-and-insurrection-threats-in-months-leading-up-to-jan-6-attack-records-show> (visited on 11/15/2022).

Bibliography

- Singer, P., F. Flöck, C. Meinhart, E. Zeitfogel, and M. Strohmaier (2014). “Evolution of reddit: from the front page of the internet to a self-referential community?” In: *Proceedings of the 23rd international conference on world wide web*, pp. 517–522.
- Singh, V., A. Johri, and R. Mitra (2011). “Types of newcomers in an online developer community”. In: *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, pp. 717–720.
- Smith, C. E., I. Alam, C. Tan, B. C. Keegan, and A. L. Blanchard (2022). “The Impact of Governance Bots on Sense of Virtual Community: Development and Validation of the GOV-BOTs Scale”. *Proceedings of the ACM on Human-Computer Interaction* 6:CSCW2, pp. 1–30.
- Smith, C. B., M. L. McLaughlin, and K. K. Osborne (June 23, 2006). “Conduct Control on Usenet”. *Journal of Computer-Mediated Communication* 2:4. ISSN: 10836101. DOI: [10.1111/j.1083-6101.1997.tb00197.x](https://doi.org/10.1111/j.1083-6101.1997.tb00197.x). URL: <https://academic.oup.com/jcmc/article/4584388> (visited on 05/05/2025).
- Smith, M. A. (1992). “Voices from the WELL: The logic of the virtual commons”. MA thesis. University of California: Los Angeles.
- Snow, D. A., L. A. Zurcher, and S. Eklund-Olson (Oct. 1980). “Social Networks and Social Movements: A Microstructural Approach to Differential Recruitment”. en. *American Sociological Review* 45:5, p. 787. ISSN: 00031224. DOI: [10.2307/2094895](https://doi.org/10.2307/2094895).
- Soliman, A., J. Hafer, and F. Lemmerich (Sept. 12, 2019). “A Characterization of Political Communities on Reddit”. In: *Proceedings of the 30th ACM Conference on Hypertext and Social Media*. HT ’19: 30th ACM Conference on Hypertext and Social Media. ACM, Hof Germany, pp. 259–263. ISBN: 978-1-4503-6885-8. DOI: [10.1145/3342220.3343662](https://doi.org/10.1145/3342220.3343662). URL: <https://dl.acm.org/doi/10.1145/3342220.3343662> (visited on 12/04/2024).
- Solomon, J. and R. Wash (2014). “Critical mass of what? Exploring community growth in WikiProjects”. In: *Eighth International AAAI Conference on Weblogs and Social Media*.
- Sorokin, P. A. (1969). *Society, culture, and personality: their structure and dynamics. A system of general sociology*. Cooper Square, New York.

Bibliography

- Squirrell, T. (2019). "Platform dialectics: The relationships between volunteer moderators and end users on reddit". *New Media & Society* 21:9, pp. 1910–1927.
- Stacey, M. (1969). "The myth of community studies". *The British Journal of Sociology* 20:2, pp. 134–147.
- Starbird, K. and L. Palen (May 7, 2011). "'Voluntweeters': Self-Organizing by Digital Volunteers in Times of Crisis". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11: CHI Conference on Human Factors in Computing Systems. ACM, Vancouver BC Canada, pp. 1071–1080. ISBN: 978-1-4503-0228-9. DOI: [10.1145/1978942.1979102](https://doi.org/10.1145/1978942.1979102). URL: <https://dl.acm.org/doi/10.1145/1978942.1979102> (visited on 02/19/2025).
- Stone, A. R. (1991). "Will the real body please stand up? boundary stories about virtual cultures". In: *Cyberspace: First Steps*. MIT Press, Cambridge, MA, USA, pp. 81–118. ISBN: 026202327X.
- Tan, C. (2018). "Tracing Community Genealogy: How New Communities Emerge from the Old". In: *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018*. AAAI Press, pp. 395–404. URL: <https://aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17811>.
- Tan, C. and L. Lee (2015). "All Who Wander: On the Prevalence and Characteristics of Multi-Community Engagement". In: *Proceedings of the 24th International Conference on World Wide Web*. WWW '15. International World Wide Web Conferences Steering Committee, Florence, Italy, pp. 1056–1066. ISBN: 9781450334693. DOI: [10.1145/2736277.2741661](https://doi.org/10.1145/2736277.2741661). URL: <https://doi.org/10.1145/2736277.2741661>.
- Tausczik, Y. R. and J. W. Pennebaker (2012). "Participation in an online mathematics community: differentiating motivations to add". In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 207–216.
- TeBlunthuis, N. and B. M. Hill (2022). "Identifying competition and mutualism between online groups". In: *Proceedings of the international aaai conference on web and social media*. Vol. 16, pp. 993–1004.

Bibliography

- TeBlunthuis, N., C. Kiene, I. Brown, L. Levi, N. McGinnis, and B.M. Hill (2022). “No community can do everything: why people participate in similar online communities”. *Proceedings of the ACM on Human-Computer Interaction* 6:CSCW1, pp. 1–25.
- Tedjamulia, S., D. Dean, D. Olsen, and C. Albrecht (2005). “Motivating Content Contributions to Online Communities: Toward a More Comprehensive Theory”. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. 38th Annual Hawaii International Conference on System Sciences. IEEE, Big Island, HI, USA, 193b–193b. ISBN: 978-0-7695-2268-5. DOI: [10.1109/HICSS.2005.444](https://doi.org/10.1109/HICSS.2005.444). URL: <http://ieeexplore.ieee.org/document/1385630/> (visited on 12/04/2024).
- Tilly, C. (1978). *From Mobilization to Revolution*. Addison-Wesley Pub. Co, Reading, Mass. ISBN: 978-0-201-07571-7.
- Timberg, C. and D. Harwell (Feb. 5, 2021). “TheDonald’s owner speaks out on why he finally pulled plug on hate-filled site”. *Washington Post*. ISSN: 0190-8286. URL: <https://www.washingtonpost.com/technology/2021/02/05/why-the-donald-moderator-left/> (visited on 02/26/2023).
- Toolis, E. E. and P. L. Hammack (2015). ““This is my community”: Reproducing and resisting boundaries of exclusion in contested public spaces”. *American Journal of Community Psychology* 56:3, pp. 368–382.
- Tsugawa, S. and S. Niida (2019). “The impact of social network structure on the growth and survival of online communities”. In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 1112–1119.
- Turner, T. C. and K. E. Fisher (2006). “Social Types in Technical Newsgroups: Implications for Information Flow”. *Virtual Communities*.
- u/controversialideas (Apr. 28, 2016). *Why do so many /r/The_Donald posts get onto the frontpage? I would have thought it'd be self-regulating in that if niche popular content is disagreeable to the majority, downvotes from /r/all would keep it off the frontpage. Does this mean Donald posts are actually supported by most /r/all readers?* URL: https://www.reddit.com/r/TheoryOfReddit/comments/4gsffj/why_do_so_many_rthe_donald_posts_get_onto_the/ (visited on 06/08/2026).

Bibliography

- u/Deimorz (Jan. 19, 2012a). *AutoModerator - a bot for automating straightforward reddit moderation tasks and improving upon the existing spam-filter*. URL: https://www.reddit.com/r/TheoryOfReddit/comments/onl2u/automoderator_a_bot_for_automating (visited on 11/30/2024).
- (Jan. 8, 2012b). *AutoModerator - a bot to automate straightforward moderation tasks*. URL: https://web.archive.org/web/20120117131005/http://www.reddit.com/r/PornOverlords/comments/o7qgx/automoderator_a_bot_to_automate_straightforward (visited on 11/30/2024).
- (June 2014). *reddit changes: individual up/down vote counts no longer visible, "% like it" closer to reality, major improvements to "controversial" sorting: r/announcements*. [Online; accessed 10. May 2024]. URL: https://www.reddit.com/r/announcements/comments/28hjga/reddit_changes_individual_updown_vote_counts_no (visited on 06/18/2014).
- (Mar. 31, 2015). *Moderators: AutoModerator is now built into reddit - new syntax and functionality: r/modnews*. [Online; accessed 10. Jan. 2024]. URL: https://www.reddit.com/r/modnews/comments/30ychn/moderators_automoderator_is_now_built_into_reddit (visited on 01/10/2024).
- u/dwaxe (Apr. 19, 2016). *This is how /r/The_Donald gets so many posts onto /r/all*. URL: https://www.reddit.com/r/TheoryOfReddit/comments/4fh8s9/this_is_how_rthe_donald_gets_so_many_posts_onto/ (visited on 06/08/2026).
- u/ketralnis (Nov. 14, 2022). *What happens to deleted posts?* URL: <https://www.reddit.com/r/redditdev/comments/yv9cat/comment/iwdf5nw/> (visited on 02/05/2023).
- u/KeyserSosa (May 2016). *Scores on posts are about to start going up: r/announcements*. [Online; accessed 10. May 2024]. URL: https://www.reddit.com/r/announcements/comments/5gvd6b/scores_on_posts_are_about_to_start_going_up (visited on 12/06/2016).
- u/simbawulf (Apr. 17, 2017). *Testing a new sign up experience*. [Online; accessed 3. Jun. 2026]. URL: https://www.reddit.com/r/changelog/comments/65x4a7/testing_a_new_sign_up_experience/ (visited on 06/03/2026).
- u/spez (June 29, 2020). *Update to Our Content Policy*. URL: https://www.reddit.com/r/announcements/comments/hi3oht/update_to_our_content_policy/ (visited on 12/27/2022).

Bibliography

- u/spladug (May 11, 2011). *Moderators: would you like self text to be replaced by [deleted] when you remove a self post? : r/modnews*. [Online; accessed 1. Aug. 2024]. URL: https://www.reddit.com/r/modnews/comments/hlvme/moderators_would_you_like_self_text_to_be (visited on 08/01/2024).
- (June 1, 2024). *Moderators: removing a self post will now hide the self text as well. : r/modnews*. [Online; accessed 1. Aug. 2024]. URL: https://www.reddit.com/r/modnews/comments/hpdxl/moderators_removing_a_self_post_will_now_hide_the (visited on 08/01/2024).
- u/tdohz (Sept. 2015). *[reddit change] Making removed & deleted content more consistent : r/changelog*. [Online; accessed 6. Jun. 2024]. URL: https://www.reddit.com/r/changelog/comments/3luvvy/reddit_change_making_removed_deleted_content_more (visited on 09/21/2015).
- u/theoddjosh (May 16, 2020). *Pushshift showing submission selftext as [removed] when the post isn't deleted : r/pushshift*. [Online; accessed 1. Aug. 2024]. URL: https://www.reddit.com/r/pushshift/comments/gl0nun/pushshift_showing_submission_selftext_as_removed (visited on 08/01/2024).
- u/umbrae (June 30, 2014). *[reddit change] API Availability: controversiality & score on comments, upvote_ratio on individual submissions : r/redditdev*. [Online; accessed 8. Jun. 2024]. URL: https://www.reddit.com/r/redditdev/comments/29i58s/reddit_change_api_availability_controversiality (visited on 06/08/2024).
- Unkel, J. and A. S. Kümpel (2020). “(A) synchronous Communication about TV Series on Social Media: A Multi-Method Investigation of Reddit Discussions”. *Media and Communication* 8:3, pp. 180–190.
- Waller, I. and A. Anderson (May 13, 2019). “Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms”. In: *The World Wide Web Conference*. WWW '19: The Web Conference. ACM, San Francisco CA USA, pp. 1954–1964. ISBN: 978-1-4503-6674-8. DOI: [10.1145/3308558.3313729](https://doi.org/10.1145/3308558.3313729). URL: <https://dl.acm.org/doi/10.1145/3308558.3313729> (visited on 04/20/2025).
- (Dec. 9, 2021). “Quantifying Social Organization and Political Polarization in Online Platforms”. *Nature* 600:7888, pp. 264–268. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/s41586-021-04167-x](https://doi.org/10.1038/s41586-021-04167-x). arXiv: [2010.00590](https://arxiv.org/abs/2010.00590) [cs]. URL: <http://arxiv.org/abs/2010.00590> (visited on 05/16/2025).

Bibliography

- Wang, Y.-C., R. E. Kraut, and J. M. Levine (2012). "To stay or leave? The relationship of emotional and informational support to commitment in online health support groups". In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 833–842.
- Wang, X., B. S. Butler, and Y. Ren (2013). "The impact of membership overlap on growth: An ecological competition view of online groups". *Organization Science* 24:2, pp. 414–431.
- Weinreich, F. (1997). "Establishing a point of view toward virtual communities". *CMC Magazine* 4:2.
- Welbourne, J. L., A. L. Blanchard, and M. B. Wadsworth (2013). "Motivations in virtual health communities and their relationship to community, connectedness and stress". *Computers in Human Behavior* 29:1. Including Special Section Youth, Internet, and Wellbeing, pp. 129–139. ISSN: 0747-5632. DOI: <https://doi.org/10.1016/j.chb.2012.07.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0747563212002270>.
- Weld, G., M. Glenski, and T. Althoff (May 10, 2022). *Political Bias and Factualness in News Sharing across More than 100,000 Online Communities*. DOI: [10.48550/arXiv.2102.08537](https://doi.org/10.48550/arXiv.2102.08537). arXiv: [2102.08537](https://arxiv.org/abs/2102.08537) [cs]. URL: <http://arxiv.org/abs/2102.08537> (visited on 04/19/2025). Pre-published.
- Weld, G., A. X. Zhang, and T. Althoff (2022). "What makes online communities 'better'? measuring values, consensus, and conflict across thousands of subreddits". In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 16, pp. 1121–1132.
- Wellman, B. (1979). "The community question: The intimate networks of East Yorkers". *American journal of Sociology* 84:5, pp. 1201–1231.
- (1987). "The community question re-evaluated".
- Wellman, B. and B. Leighton (1979). "Networks, neighborhoods, and communities: Approaches to the study of the community question". *Urban affairs quarterly* 14:3, pp. 363–390.
- Welser, H. T., E. Gleave, D. Fisher, and M. A. Smith (2007). "Visualizing the signatures of social roles in online discussion groups". *Journal of Social Structure* 8:2, pp. 1–32.
- Welser, H. T., E. Gleave, V. Barash, M. Smith, and J. Meckes (2009). "Whither the Experts? Social Affordances and the Cultivation of Experts in Community Q&A Systems". In: *2009 International*

Bibliography

- Conference on Computational Science and Engineering*. 2009 International Conference on Computational Science and Engineering. IEEE, Vancouver, BC, Canada, pp. 450–455. ISBN: 978-1-4244-5334-4. DOI: [10.1109/CSE.2009.254](https://doi.org/10.1109/CSE.2009.254). URL: <http://ieeexplore.ieee.org/document/5284180/> (visited on 05/05/2025).
- White, J. I. and L. Palen (Feb. 2015). “Expertise in the Wired Wild West”. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, Vancouver BC Canada, pp. 662–675. ISBN: 978-1-4503-2922-4. DOI: [10.1145/2675133.2675167](https://doi.org/10.1145/2675133.2675167). (Visited on 02/19/2025).
- Whittaker, S. (1996). “Talking to strangers: An evaluation of the factors affecting electronic collaboration”. In: *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, pp. 409–418.
- Whittaker, S., E. Isaacs, and V. O’Day (1997). “Widening the net: workshop report on the theory and practice of physical and network communities”. *ACM SIGCHI Bulletin* 29:3, pp. 27–30.
- Whittaker, S., L. Terveen, W. Hill, and L. Cherny (Nov. 1998). “The Dynamics of Mass Interaction”. In: *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. CSCW98: Computer Supported Cooperative Work. ACM, Seattle Washington USA, pp. 257–264. ISBN: 978-1-58113-009-6. DOI: [10.1145/289444.289500](https://doi.org/10.1145/289444.289500). URL: <https://dl.acm.org/doi/10.1145/289444.289500> (visited on 05/27/2026).
- Wohn, D. Y. (2019). “Volunteer moderators in twitch micro communities: How they get involved, the roles they play, and the emotional labor they experience”. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–13.
- Wohn, D. Y., P. Jough, P. Eskander, J. S. Siri, M. Shimobayashi, and P. Desai (2019). “Understanding digital patronage: why do people subscribe to streamers on twitch?” In: *Proceedings of the annual symposium on computer-human interaction in play*, pp. 99–110.
- Wohn, D. Y. and C. Lampe (2018). “Psychological wellbeing as an explanation of user engagement in the lifecycle of online community participation”. In: *Proceedings of the 2018 ACM International Conference on Supporting Group Work*, pp. 184–195.

Bibliography

- Wohn, D. Y., C. W. Yuan, and J. S. Siri Jr (2021). “The Many Facets of Me: Multiple Account Management on Reddit.” In: *HICSS*, pp. 1–10.
- World War Meme* (Apr. 2017). [Online; accessed 26. Feb. 2023]. URL: <https://www.politico.com/magazine/story/2017/03/memes-4chan-trump-supporters-trolls-internet-214856> (visited on 02/26/2023).
- Wright, M. N. and A. Ziegler (2017). “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R”. *Journal of Statistical Software* 77:1, pp. 1–17. DOI: [10.18637/jss.v077.i01](https://doi.org/10.18637/jss.v077.i01).
- Yang, D., R. E. Kraut, and J. M. Levine (2017). “Commitment of newcomers and old-timers to online health support communities”. In: *Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 6363–6375.
- Yu, B., X. Wang, A. Y. Lin, Y. Ren, L. Terveen, and H. Zhu (2017). “Out with the old, in with the new? unpacking member turnover in online production groups”. *Proceedings of the ACM on Human-Computer Interaction (CSCW)* 1:117, pp. 1–19.
- Yu, Y., J. Jiang, and P. Dhillon (Sept. 18, 2024). *Characterizing the Structure of Online Conversations Across Reddit*. DOI: [10.48550/arXiv.2209.14836](https://doi.org/10.48550/arXiv.2209.14836). arXiv: [2209.14836 \[cs\]](https://arxiv.org/abs/2209.14836). URL: <http://arxiv.org/abs/2209.14836> (visited on 12/04/2024). Pre-published.
- Zhang, J., W. L. Hamilton, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec (2017). “Community identity and user engagement in a multi-community landscape”. In: *Eleventh International AAAI Conference on Web and Social Media*.
- Zhu, H., J. Chen, T. Matthews, A. Pal, H. Badenes, and R. E. Kraut (2014). “Selecting an effective niche: an ecological view of the success of online communities”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 301–310.
- Zhu, H., R. E. Kraut, and A. Kittur (2012). “Effectiveness of shared leadership in online communities”. In: *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 407–416.
- (2014). “The impact of membership overlap on the survival of online communities”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 281–290.