User Migration in Online Social Networks:
A Case Study on Reddit During a Period of Community Unrest

Edward Newell[1†], David Jurgens[2†], Haji Mohammad Saleem[1],
Hardik Vala[1], Jad Sassine[1], Caitrin Armstrong[1], and **Derek Ruths[1]**

{edward.newell, haji.saleem, hardik.vala, caitrin.armstrong}@mail.mcgill.ca
jurgens@stanford.edu, {jad.sassine, derek.ruths}@mcgill.ca
[1]McGill University, [2]Stanford University

Abstract

Platforms like Reddit have attracted large and vibrant communities, but the individuals in those communities are free to migrate to other platforms at any time. History has borne this out with the mass migration from Slashdot to Digg. The underlying motivations of individuals who migrate between platforms, and the conditions that favor migration online are not well-understood. We examine Reddit during a period of community unrest affecting millions of users in the summer of 2015, and analyze large-scale changes in user behavior and migration patterns to Reddit-like alternative platforms. Using self-reported statements from user comments, surveys, and a computational analysis of the activity of users with accounts on multiple platforms, we identify the primary motivations driving user migration. While a notable number of Reddit users left for other platforms, we found that an important pull factor that enabled Reddit to retain users was its long tail of niche content. Other platforms may reach critical mass to support popular or "mainstream" topics, but Reddit's large userbase provides a key advantage in supporting niche topics.

1 Introduction

Socially-curated websites such as Reddit depend on large communities for content creation and moderation (Lampe et al. 2010; Zhong et al. 2013; Stoddard 2015). Though one platform may dominate a given functionality, many smaller alternative platforms often co-exist. Occasionally, entire communities migrate from a popular platform to an alternative, as occurred from Slashdot to Digg (Garcia, Mavrodiev, and Schweitzer 2013; Liu and Xiao 2013).

In July 2015, a series of events on Reddit triggered a period of community unrest during which (a) many users publicly expressed a desire to leave and (b) several of the most-popular communities, referred to as subreddits, took their content offline in protest (Matias 2015). The scale of these events attracted coverage by major news media (Lynch and Swearingen 2015; Kamen 2015).

This period provides a remarkable opportunity to study the conditions and patterns of online migration, and investigate what motivations drive it. Users participate on Reddit and its alternatives mainly through public postings. This

[†]These authors made equal contributions.

makes it possible to study migration patterns using users' histories of activity. Meanwhile, many users actively express their opinions about the platforms and their reasons for choosing where to spend their time, providing insight into their motivations.

Social media platforms have been studied extensively (Kietzmann et al. 2011; Ruths and Pfeffer 2014; Lin and Lu 2011), but whether and how users participate in multiple platforms concurrently has only recently been explored (Malinen 2015). These studies have focused on small numbers of users, and coarse time scales, which do not reveal the dynamics of migration in full detail (Kumar, Zafarani, and Liu 2011). In contrast to these studies, cf. (Zhang et al. 2009; Hermosa and Hachtmann 2013), we investigate the motivational factors for a specific migration event and use quantitative analysis to contextualize our findings, connecting them to actual shifts in user activity.

We systematically analyze Reddit and 21 other platforms cited by Reddit users as alternatives. The scale of these alternatives range in size from a handful of users to hundreds of thousands. Our analysis provides wide coverage of the spectrum of possible destinations, uncovering the varied motivations that drive users to platforms of different sizes and designs. To investigate the interplay between users' motivations and behavior, we combine (a) surveys of users from various platforms, (b) an open-coded analysis of users' stated motivations in on-platform posts and comments, and (c) a large-scale computational analysis of user activity patterns.

Our work offers four key findings. First, user interests are sufficiently diverse that no one platform caters to all users, despite platform's near-identical functionality; indeed, the attraction of users to a particular platform follows from specific motivational interests, such as free speech or aversion to hate speech. Second, user behavior changes with migration events. Whereas an individual may contribute few posts and comments on Reddit, after migrating to a new platform, their level of contribution frequently increases. Thus, even if the primary content contributors of Reddit do not migrate, this behavior change can help platforms attain a critical level of activity. Third, a major draw of Reddit is its ability to support niche communities. Alternative platforms may attract sufficient users to aggregate content that appeals to a broad audience. But the very large user base of Reddit enables it

to support niche content which many users cite as an important attraction. And fourth, while smaller platforms have difficulty competing with respect to content-generation, they appear to foster a more congenial atmosphere, which may be an important driver of early growth.

2   The Reddit Controversy

Reddit is the most popular exemplar[1] of a class of websites known as *social content aggregators*, on which users can post new content as well as vote and comment on each other's content. Many users do not comment and fewer still create posts; these users represent a large silent majority, a phenomenon which is observed on many other types of social media (Benevenuto et al. 2009; Bernstein et al. 2013; Gong, Lim, and Zhu 2015). Like most social content aggregators, Reddit contains many topical communities that exist in parallel, called *subreddits*. Communities typically have rules that govern the content of posts and comments. Rules are enforced by *moderators*, who can delete comments and ban users. Moderators emerge by virtue of creating the subreddit, or are appointed by existing moderators.

Reddit was founded in 2005 with the intent of providing a discussion forum for all under the principle of free speech (Hill 2012). Reddit has since grown to receiving over 160 million unique views every month, making it among the most-visited websites[1]. Along with this growth has come a significant increase in content diversity; currently Reddit hosts over 350,000 subreddits. These include popular themes such as `/r/funny`, and `/r/AskScience`, but also controversial communities such as the recently banned community `/r/FatPeopleHate`, which engaged routinely in hate speech.

Reddit's growth has been accompanied by complaints by moderators of a lack of administrative support and by users of moderator's abuse of power (Alfonso 2014). The Reddit administration has gradually moved towards more restrictive content policies to protect users from harassment, hate speech, and offensive content. These policy changes have not been well-received by certain communities who believe that freedom of speech should be paramount. In a highly controversial event, on June 10th, 2015, Reddit administration banned five communities for hate speech and antagonizing other subreddits.

Another controversial event took place on July 2nd, 2015 when Reddit fired their Director of Communications, Victoria Taylor. No warning was given to moderators who depended on her for the operation of `/r/IAmA`, one of the most popular subreddits.

Without her support, the `/r/IAmA` moderators shut down the subreddit to regroup, and by the end of the day, 2,278 subreddits had also shut down in protest, some of which had millions of members (Matias 2015). These actions were done "as a statement on the treatment of moderators by Reddit administrators, as well as a lack of communication and proper moderation tools" (Lynch and Swearingen 2015). This left millions of users without access to some of the most highly-trafficked subreddits for which alternatives
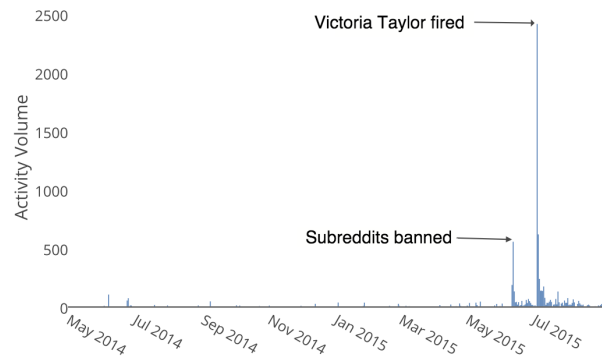
---

[1]by aggregate traffic, at the time of writting, `alexa.com`

Figure 1: Daily activity (number of posts and comments) in subreddit `/r/RedditAlternatives`.

were available on other platforms. Both events coincide with a surge in discussion among Reddit users of alternatives to Reddit (see Figure 1). As a result, we take these as key dates in the analysis of user migration.

We frame our analyses around three related research questions. First, we assess whether the events during July 2015 led to decreased activity on Reddit and whether this had a corresponding increase in activity on alternatives. Second, given that multiple alternative platforms were actively discussed in mass media, we assess whether individuals went to the same alternative. Third, among users who left, we assess the underlying motivations for their particular behavior.

3   Measuring the Reddit Migration

Given the peaked interest in the alternative platforms during events of July 2015, we examine whether individuals actually began to use these platforms and whether this usage was in addition to or as a replacement for Reddit.

Data and Methods

We begin by looking at bulk activity statistics for Reddit and its alternatives, and then analyze activity patterns of individual users having accounts on multiple platforms. Users contributing to the `/r/RedditAlternatives` community provided details of alternative platforms, from which we aggregated a list of 32 platforms. Of these, 11 platforms were eliminated because they lacked publicly accessible content, were shut-down midway through the study, or because it was not possible determine content posting dates. Reddit and each of the remaining 21 alternative platforms were crawled for all publicly available content.

To characterize movements between platforms in finer detail, we algorithmically match accounts on multiple platforms that likely belong to the same user. Given the lack of user-specific information on each platform, we use a conservative precision-oriented algorithm that links accounts with matching, uncommon usernames (e.g., two accounts with the name "joe" would not be linked since that username is likely to arise multiple times by chance). A double-blind evaluation of the algorithm establishes a lower bound for precision at 0.6, which is sufficient to identify trends in user
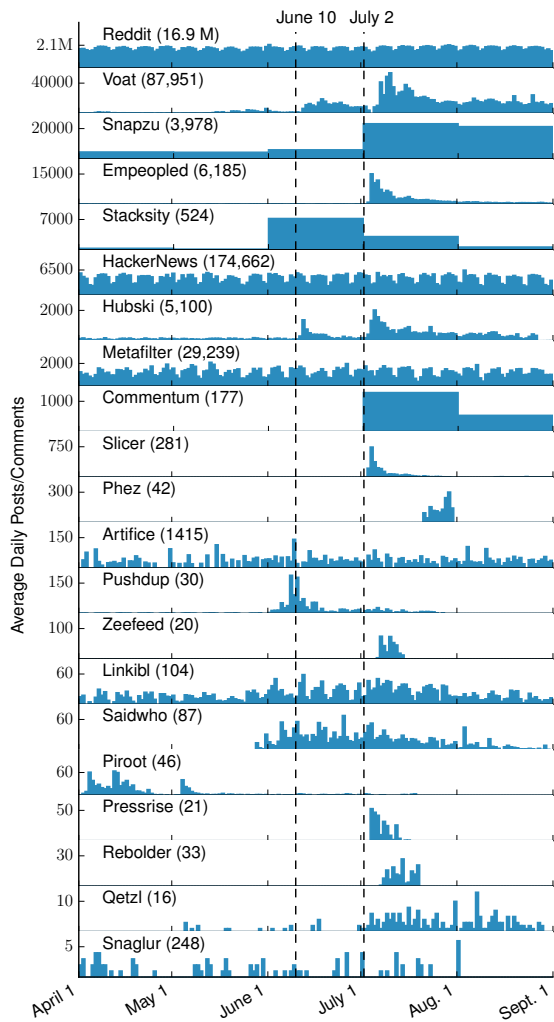
Figure 2: Daily post and comment volume per platform, or monthly when daily data is not available, from April 1st through September 1st, 2015. The number of active users is shown for each platform in parentheses next to its name.

behavior. Details of the algorithm and evaluation are given in Appendex A.

To analyze users' changes in activity among linked accounts, we categorize users into three classes: (1) migrants, (2) tourists, and (3) dual citizens. Using activities from June 1st through August 31st, 2015, we define each class as follows. Because user activity tends to be bursty, user activity rates were smoothed by taking a 7-day running average. We take migration to be a substantial shift in activity, wherein the user's smoothed activity is higher on alternatives than on Reddit for at least two weeks. In contrast, tourists exhibit a sudden burst in activity on Reddit alternatives and then no further activity there. Finally, dual citizens have activity on alternatives that was sustained for longer than one week, but their activity is not consistently higher on alternatives than Reddit. Small variations in smoothing window and periods of activity yielded similar results in user classifications.

Results

platform Activity. Many alternatives to Reddit saw a substantial increase in their relative post and comment volumes; however, the volume on Reddit was largely unchanged, indicating that the events had minimal effect on Reddit itself. As seen in Figure 2, a spike in activity appears on several alternatives directly after the events of June 10th and July 2nd, 2015. While the initial volumes of traffic largely subside, several alternatives do sustain a higher baseline of activity thereafter.

Interestingly, the platforms that received the largest traffic spikes had low initial volume, whereas more established alternatives, such as HackerNews and Metafilter showed essentially no change. Note that the increased activity for Voat, Snapzu, and Empeopled increased well beyond the prior activity levels of these established platforms, so it is not merely the case that activity changes on the established platforms were masked by a higher baseline activity. We speculate that users had already considered using the established platforms at previous times, whereas at the time of the events in June and July 2015, they were interested in trying newer alternatives. In further discussions, we focus our analyses only on Voat, Snapzu, Empeopled, and Hubski, which received the majority of traffic from Reddit during the events.

User Behavior. In line with the trends seen in platform activity, among the linked accounts we observed relatively few instances of migrant users (1,895), with the majority class being tourists who simply visit, comment, and then never return (11,176 users). However, a substantial number of individuals became dual citizens (7,289) and while they continue to use Reddit, they regularly return to one alternative to contribute there as well. The migrants, tourists, and dual-citizens identified on Voat, Empeopled, Snapzu, and Hubski account for 18,296 (71.7%) of users that were linked to accounts on multiple platforms (see Table 1). The remaining 7,593 (29.3%) linked accounts were divided between the other platforms under analysis or belonged users that had been active on Reddit alternatives prior to June 1st but not during the period of analysis. Given that Reddit has tens of millions of users, such a migration is unlikely to have any effect on it. Nevertheless, the migrants, tourists, and dual citizens provide an opportunity to study user movements online.

Clear differences can be seen in the commenting frequency of each class of user. In Figure 3 average activity for each class of user is shown, after aligning users' activity profiles based on their first date of activity on an alternative platform, which we refer to as their departure date. While users of all three types show a peak in activity on the first day of joining an alternative platform, migrants show sustained usage, with nearly double the average number of comments as dual citizens. Figure 3 illustrates three notable trends. First, migrants have a significant activity increase on alternatives in the first week of moving. Our analysis found this increase is due two factors (1) users introducing themselves in content threads devoted to newcomers, which was most seen on Empeopled, and (2) users venting their frustration with Reddit policies and practices, mostly seen on Voat.
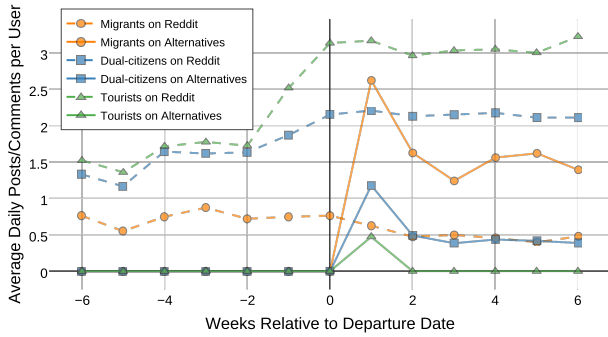
Figure 3: Average daily posting and commenting frequency per user on Reddit and its alternatives over time, for each of the three classes. Time is represented as the number of weeks relative to their estimated departure date from Reddit.

Activity subsides after the first week but for migrants activity on alternatives remains above that on Reddit. Second, dual-citizens and tourists had significantly higher initial activity rates on Reddit prior to trying an alternative platform, which suggests they were more actively involved in Reddit communities; such users might have more social capital on Reddit making them reluctant to sever their ties to Reddit. Third, tourists show a substantial increase in activity on Reddit around the departure date and afterwards, which we observed was due to complaints on Reddit and comments about trying to the alternative.

4 Understanding Motivations for Migrating

Our computational analysis of user behavior revealed that many individuals began using alternatives after the events of July 2nd, which raises the question of what specifically motivated them to seek out the other platforms and why they selected a particular platform. We draw on theories of geographic migration which describe the drivers of migration in terms of pushes and pulls. We collect a large number of comments from the five focal platforms (Reddit, Voat, Empeopled, Snapzu, and Hubski), and perform an open-coding analysis to identify the motivations users cite when evaluating platforms. We then design a survey inquiring into these motivations which we administer to users of the focal platforms. Using the self-reported motivations as a starting point enables us to identify motivations from the users themselves, while the survey helps us to obtain opinions that users might be less inclined to express publicly. Before discussing the survey and open-coding analysis, we provide brief background on migration theory.

	Voat	Empeopled	Snapzu	Hubski	**Total**
Migrant	1,439	58	38	35	1,570
Dual-Citizen	5,832	211	210	143	6,396
Tourist	9,096	694	312	228	10,330
Total	16,367	963	560	406	18,296

Table 1: Number of linked users of each class per platform.

Migration Theory

Physical migrations of human populations are usually driven by geographical and geopolitical disparities. Different theories emphasize different disparities: neoclassical economic theory cites wage differences, market theory distinguishes low-skilled and high-skilled labor, more recent labor migration models account for the role of family and household structures in moving spouses and children, and relative deprivation theory considers the perception of income differences among one's (geographically dispersed) peers and neighbors.

Everett Lee's migration laws unify these different models by observing that disparities create push and pull factors which influence the movement of individuals and communities (Lee 1966). In this view, migration is a response to conditions that force populations to leave (push) or entice them to move to (pull) a location.

In the context of online platforms, we can consider an array of push and pull factors which make people want to leave (e.g., one's community getting banned) or join (e.g., an agreeable moderation policy). To understand the factors influencing migration, we identified the aspects mentioned by users in posts on the focal platforms, then codified these into systematically-measurable variables suitable for use in a survey.

Data and Methods

Self-Reported Motivations. We culled a set of 2,987 comments potentially expressing opinions about platforms using three heuristics: (1) The comment had to name at least one platform or use the phrase "this site" or "that site". (2) The comment had to mention "leave", "left", "migrating", "migration", "migrated", or "exodus". (3) At least 3% of the comment words must be drawn from the words above.

Four annotators independently read a subset of 50 comments and identified the motivations therein to create a coding scheme. The annotators unified their schemes by consensus into a hierarchical scheme with 6 coarse-grained and 31 fine-grained motivational factors (additional details available at networkdynamics.org/pubs/2016/reddit-exodus/). Eight annotators applied the scheme to the remaining comments, with 10% annotated by all annotators to assess inter-rater reliability. Average pair-wise agreement was 76% for coarse factors and 74% for fine factors.

Survey. The survey addressed four key aspects of platform usage: (1) the user's history on Reddit and alternatives, (2) the user's opinions about the two key controversial events on Reddit, (3) demographic information, and (4) several facets of user experience on Reddit. Participants were allowed to submit incomplete surveys.

Surveys were first posted publicly to communities on Reddit, Voat, Hubski, Empeopled, Snapzu, Stacksity, Piroot, HackerNews, Linkibl, SaidWho and Qetzl. We also used private messaging (PM) features on Reddit and Voat to solicit participation from randomly-selected users. The anonymous survey was approved by our ethics review board.

Platform	Complete	Partial	ρ	p-value
Reddit	110	13	0.89	0.02
Voat	85	10	-0.20	0.70
Hubski	17	2	0.66	0.15
Empeopled	12	1	0.77	0.07
Snapzu	12	0	0.60	0.21
HackerNews	2	0	–	–
Linkibl	2	0	–	–
Quetzl	0	1	–	–

Table 2: Number of complete / partial survey responses on various platforms, and Spearman's rank correlation (ρ) between the ordering of coarse motivations by importance expressed in those responses compared to that obtained from the analysis of user comments (p-value indicates significance of correlation).

Results

Results from the survey and users' comments largely corroborated one another. In addition to the 2,987 analyzed comments, 267 users submitted valid surveys. More than 70% of the responses came from users under age 35. 74% of the responders were male and 63% of the responders were from United States. For both the coded comments and survey, we ranked which motivational factors were most important to users, based on which were mentioned most frequently (in the case of comments), and which were attributed the highest score on a Likert scale (in the case of surveys). Spearman's rank correlation shows that, for the most part, these rankings are similar (see Table 2). Because of the overall similarity between the survey and coded user comments, we discuss them together.

Voat is a notable exception, where users differed greatly in how they expressed themselves publicly and through the survey. In fact, Voat stood out in many ways. For one, whereas users frequently commented positively on the congeniality of the other alternatives to Reddit, in that respect, Voat resembled Reddit itself, receiving more negative comments about user behavior (see Figure 4). While our survey was well-received on the other Reddit alternatives, on Voat, the survey was met with a less positive reception publicly, despite positive and constructive private comments about the survey. One possible explanation for this discrepancy is the nature of the flow of users from Reddit to Voat. Several communities that were banned from Reddit on June 10th, 2015 moved en masse to Voat, carrying with them their grievances about Reddit and public perceptions of supporting hate speech, which may have influenced their attitudes towards public inquiries on their motivations for leaving.

Relative importance of motivational factors. Although platforms and users appear to vary in the degree to which they emphasize particular factors, we see relatively stable trends in what users regard as most important. Across all platforms, users mentioned two broad categories of aspects most frequently: (1) *site features* – including user interface / aesthetics, voting functionality, community size, platform availability, and level of commercialization and (2) *content*

– including the quality of posts, quality of discussions, presence of hate speech, frequency of new content, and uniqueness of content. Among comments users made about platform features, the platform's user interface was most often mentioned (Figure 4, Topic 9).

Surprisingly, although moderation issues appear to be very important to Reddit users in general, they recieved little emphasis from users comparing different platforms (Figure 4). It is possible that moderation is in fact less important than other factors, but another explanation is that assessing moderation is more difficult or takes more time than assessing other factors. Whereas a user might develop a preference for a particular voting functionality quickly, the relative merits of different moderation practices may be less clear initially, leading those issues to be discussed less. In any case, despite moderation being a relatively important issue in general, users did not comment on it frequently when comparing alternative platforms. On the other hand, issues related to platform administration (notably freedom of speech, see Figure 4, Topic 1) appears to have been the largest push away from Reddit. This is what we expected given that the key inciting events concerned administrative intervention. Correspondingly, site administration was a strong pull to Voat; interestingly this was not the case for other alternatives.

Objectionable content & behavior. Much of the controversy that sparked unrest among Reddit users related to how the administration interacted with communities that routinely posted objectionable content. Survey results and user comments show that the issue of balancing freedom of speech with the presence of objectionable content is highly divisive. While a majority of users who submitted surveys favor banning illegal content (99 versus 37), this trend reverses when it comes to the banning of objectionable, but legal, content (47 versus 126). As one user put it:

> "Shadow banning is wrong. Banning obscene content that's legal is wrong."

Among users who said they had reduced their Reddit usage in the survey, 50% viewed the banning of /r/FatPeopleHate negatively, while those who identified as continuing to use Reddit regularly viewed that action more positively (only 27% viewed it negatively). Respondents who said they continued to use the platform regularly also identified the presence of hate speech on Reddit as the aspect of the platform they disliked the most (accounting for 50% of the 80 responses).

On the other hand, users of Voat praise the platform's openness to objectionable content. We received 85 responses from users who reported declining use of Reddit in favor of Voat. They identified site administration (69) and site moderation (65) as the leading causes of dissatisfaction.

This response closely matched that from self-reported comments, as seen in the large percentage of comments on the importance of freedom of speech in Figure 4, Topic 1:

> "If you want a content aggregator that censors go to reddit. If you want a place to have free speech stay on voat."

> "I think the universal opinion here [on Voat], which I fully agree with, is that people should be able to come together
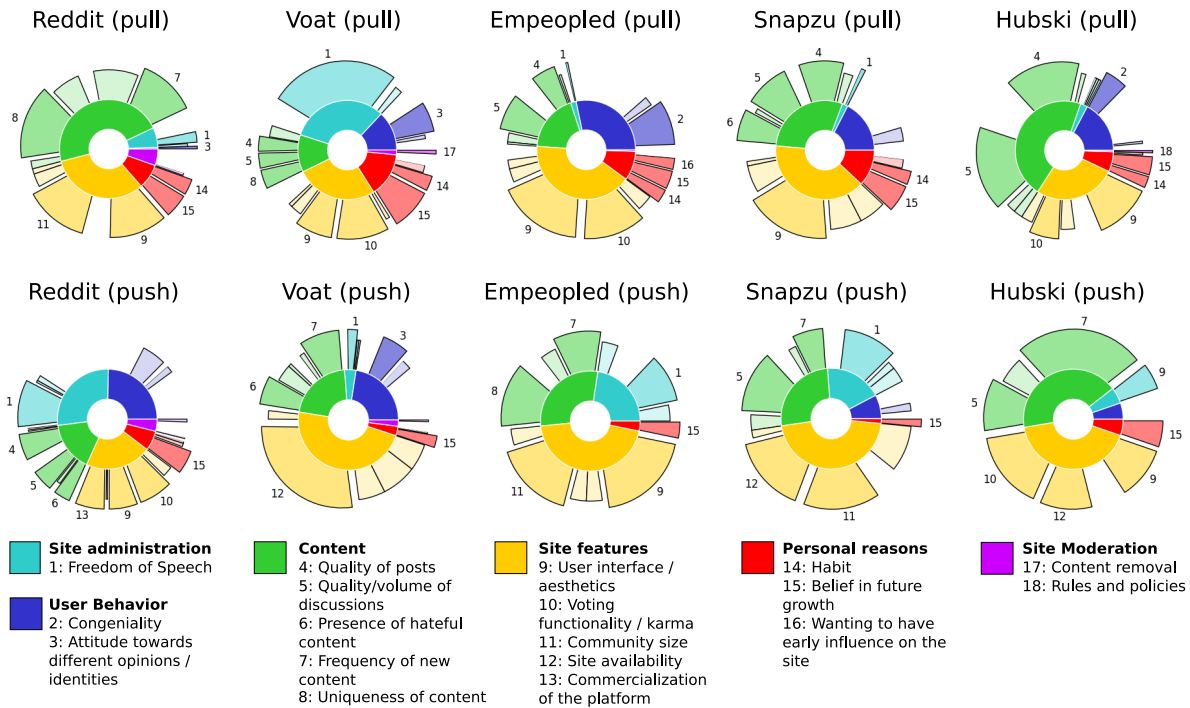
Figure 4: Relative frequencies with which users cite specific motivational factors as being advantageous (top row) or disadvantageous (bottom row) with regard to specific platforms, as determined through open-coding of comments collected from each platform. Inner circles show coarse motivational factors; outer circles show selected fine motivational factors.

and discuss anything, even offensive and disgusting opinions. Fuck Reddit and their thought policing."

It is interesting to note how Voat's emphasis on freedom of speech creates a strong pull that directly addresses the push away from Reddit due to perceived censorship. Our analysis of user traffic suggests that Voat absorbed the most users from Reddit (Table 1). Through the lens of Lee's push-pull theory of migration (1966), we can see this increased migratory flow as being facilitated by the alignment of a strong push from Reddit with a strong pull toward Voat along a single factor. Snapzu, in contrast, has attempted to avoid issues with objectionable content by establishing strict posting rules. While this was met with criticism, a subset of users seem to appreciate that approach (Figure 4, Topic 1).

Effects of platform scale. To a large extent, the size of a platform's userbase will determine the amount of original content that it produces, which stands as a major barrier to entry for new platforms seeking to grow their userbase. Users commented negatively about content more often on Reddit's alternatives, especially regarding the frequency with which new content was generated (Figure 4, Topic 7):

> "I was just thinking today about why I don't spend more time browsing snapzu like I do reddit... the big one that jumps out at me is that reddit has a lot more active users creating content."

In contrast, Reddit was the focus of a large number of positive comments regarding content, identifying Reddit's content as a significant pull factor.

However, users' comments also described advantages of smaller platforms. Despite complaints about content turnover, users valued Hubski for the *quality* of its content and discussions (Figure 4, Topics 4 and 5). Among participants who responded to the survey on Hubski (17), 47% indicated that loss of interest in the content on Reddit was a leading reason for their declining use of Reddit.

Frequently, users expressed the feeling that smaller platforms were more congenial, or that they felt that their contributions had more impact:

> "while Snapzu has nothing like Reddit's volume, I just like that it's generally a much more civil crew."

> "Reddit seems to have more content to look at, but. Voat has more people to talk to."

> "The difference between Reddit and Empeopled, is the same as going from a Metropolitan city to a progressive small town. Your presence simply matters more here..."

How users perceived congeniality can be seen in Figure 4, Topics 2 and 3. This shows that congeniality was a strong pull factor for the alternatives, whereas user behavior in general was criticized with regard to Reddit. Users who responded to the survey when it was posted on Snapzu, Empopled and Hubski cited good community as the most important feature for a platform (92% of 12 responses, 100% of 12 responses and 65% of the 17 responses respectively). Thus, while attracting a critical mass of active users is difficult for new platforms, they appear to be more able to fos-

ter a sense of connection and impact within their userbase, which may be an important avenue for early growth.

Migrating users increase engagement. As we will describe in the next section, our analysis of user activity revealed that users who migrate to new platforms tended to become more active on the new platform than they had previously been. Of the 144 users that completed the survey when it was posted on platforms other than Reddit, 67% report making daily contributions to that platform, but only 35% report the same for Reddit. Self-reported comments suggest that rapid content generation may dissuade users from contributing, feeling that their contribution will not be noticed:

> "If you are a newer poster your posts receive no attention. I posted on askreddit and received almost no response. It makes you not want to post anything at all."

Another factor that might inhibit contribution on larger platforms like Reddit could be that users perceive it to have a comparative lack of congeniality and greater issues with antisocial behavior ("trolling") (see Figure 4—User Behavior):

> "I found the level of trolling attacks made it nearly impossible to feel comfortable posting content or opinions [on Reddit]"

Niche communities play an important role in attracting users. The single most frequently cited pull factor for Reddit was its uniqueness of content (Figure 4):

> "I have to go back to Reddit more often keep up with niche subjects. Cause: There's no content on [Empeopled], and a lot of the topics I follow are unlikely to ever be ported over."

> "Reddit having such a huge userbase allows me to find the niche subreddits that are interesting to me, but places like Hubski I find somewhat barren in that regard"

Thus, it would appear that larger platforms not only have an advantage in generating greater quantities of original content, they are also able to provide more *diversified* content.

5   Characterizing Multi-Site Users

Individuals cited multiple reasons for why they were motivated to leave Reddit and try a new platform. However, such statements may not reflect the underlying reality of their behavior. Therefore, using our set of linked users, we test for the effects of two stated trends: (1) niche communities kept users coming back to Reddit and (2) migration increased users' engagement.

Data and Methods

The posts and comments for all linked users were aggregated to identify the communities in which a user posts and their relative frequency of posting in each. To assess how popularity impacts contributions, we computed the ranking of each subreddit according to the number comments made to that community during June and July 2015. To compare users' behavior on Reddit with that on the alternative platforms, we leverage the fact that many alternatives feature subreddits with direct analogs to those seen on Reddit, e.g., AskReddit and AskEmpeopled. By mapping these communities, when a user posts to an alternative, we can identify how popular the corresponding subreddit would be on Reddit. These rankings reveal whether long-tail Reddit content is accessible on the alternative in its most popular communities.

Subreddits were mapped by matching names after lowercasing, removing punctuation, and replacing any usage of the platform's name with a platform-generic token. A small number of popular subreddits with idiosyncratic names were manually matched. This process matched the communities on alternative platforms that account for over 90% of the posting volume. The largest remaining unmatched communities were (a) those originally banned by Reddit, (b) those dedicated to talking about Reddit, and (c) those dedicated to functionality specific to the alternative platform.

Results

Our analysis of comments revealed that Reddit's greatest pull is its abundance of niche content that is not well substituted on alternative platforms. When looking at the communities in which dual citizens and tourists post on Reddit, we do find a slight tendency to post more often in smaller communities, relative to the average user and to users that chose to migrate. Figure 5 (left) shows the frequency with which users post to a particular community based on that community's activity rank, where communities are partitioned into two groups: (1) those in the top-100, which account for over 46% of all commenting volume and (2) those ranked 1000 and below, which are more specialized niche communities. Migrants and dual-citizen users are less likely to post to the top-100 communities, while dual citizens and tourists are more likely to post to long-tail communities than migrants and regular users. Although not statistically significant on their own, these differences are consistent with the self-reported statements made by users, and the trend was robust across different divisions of communities into popular and long-tail groups. Thus, posting patterns do lend support to the notion that would-be migrants are drawn back to Reddit preferentially to participate in smaller communities.

Meanwhile, on Reddit's alternatives, migrants show significantly higher posting rate in popular communities ($p < 0.05$), suggesting that the alternative platforms provide substitutes for Reddit's popular communities (Figure 5, right). Nevertheless, migrants and dual-citizen do still actively participate in communities on alternative platforms that correspond to long-tail or niche communities on Reddit.

To test for the second trend, we analyzed how individuals change their posting frequency when joining a new platform, relative to their initial participation on Reddit. Of the different user types, only migrant users increased their overall participation rates when joining a new platform, as shown in Table 3, confirming statements made in the surveys and in posts. In contrast, the other two user types saw substantially lower participation rates. This difference suggests that although alternative platforms may attract fewer users overall, they are better able to elicit active contribution from the users who do in fact migrate. This may help smaller platforms reach a critical level of content generation despite their size.
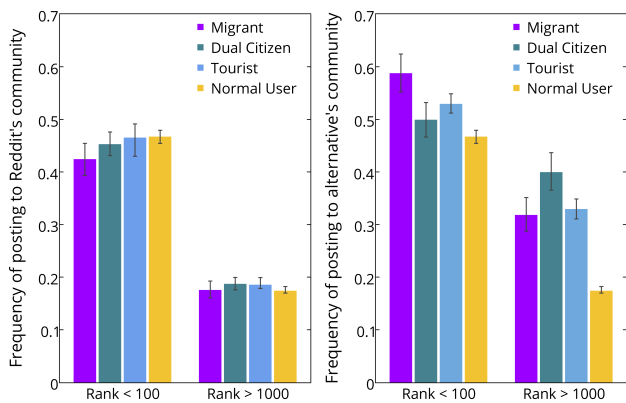
Figure 5: Probabilities of posting to communities according to popularity. Left: Posting probability for normal and multi-site users in Reddit communities. Right: Posting probability to alternative communities, classed based on the rank of the analogous community on Reddit.

User Type	Voat	Empeopled	Snapzu	Hubski
Migrant	**+81.7**	**+296.2**	+55.76	-2.2
Dual Citizen	**-75.3**	-51.2	**-81.6**	**-81.2**
Tourist	**-96.7**	**-90.1**	**-94.6**	**-93.9**

Table 3: Percent change in activity (number of posts and comments) on alternative platforms after users' departure date, relative to activity on Reddit prior to departure, for classes of user based on activity patterns. Statistically significant ($p < 0.001$) values are shown in boldface.

6   Related Work

Our investigation touches on three bodies of work. First, several studies have investigated user migration between social media platforms and communities from different angles. The first group of studies focuses on computational analyses of users to determine behavior change. Kumar, Zafarani, and Liu (2011) study user migration for 7.2K users on various categories of social platforms. Notably, they identify Reddit users as having a high propensity to move to alternate platforms. Similarly, Guimaraes, Couto da Silva, and Almeida (2015) model user flow in community question answering platforms. Both studies do not examine any of the motivations behind the user migration they demonstrate. In contrast, our work performs a similar computational analysis, but also identifies the platform and motivational factors involved. Zhu, Kraut, and Kittur (2014) examine community survival as a function of multiple memberships within Wikia communities. Their work found that higher levels of joint memberships between Wikia communities was correlated with success. In contrast, our work examines a fundamentally different setting where communities are actively competing with each other for users and the unique content they bring. Zhu et al. (2014) take an ecological view to examine the importance of niche selection in community success. Our work contextualizes the importance of specialized communities in larger social content aggregators.

A second group of studies pursued motivational factors for migration using surveys. Most related is the work of Hermosa and Hachtmann (2013), who performed surveys with users across multiple social media platform alternatives to assess the factors involved in using each, and found that connections with family and friends are a primary driver behind their usage of a particular platform. Polites and Karahanna (2012) survey users to investigate the interaction between habit, inertia, and switching costs in user switching intention between group collaboration systems. Zhang et al. (2009) identify satisfaction, sunk cost, and attractive alternatives as primary factors behind bloggers choosing to switch platforms. Hou et al. (2012) analyze the migration between social media platforms where site features and behaviors push users away from their current platform or pull them to a new one. In surveys of Facebook users, they find that pull mechanisms had the greater effect on behavior change. However, a common fault of survey-based studies was that analysis was not corroborated with data showing that behavior change actually occurred or to which platform users would have migrated, making it difficult to assess the impact of these factors.

A distinct but related body of work has examined top-down motivations that change within-platform user behaviour. Two common motivating factors are moderation and social norms. Wise, Hamman, and Thorson (2006) study the effects of moderation on online communities and find that an individual's motivation for using a particular platform impacts tolerance for offensive comments, with individual-centric motivators fostering less tolerance. Cosley et al. (2005) perform a field experiment with 87 MovieLens users and find that oversight improves both the quantity and overall quality of information, while reducing antisocial behaviour. Bergstrom (2011) notes that Reddit users have established social norms (particularly, account authenticity) and will react as a community when those norms are violated. Our work is notable in that it explores the comparative importance of these and other factors, instead of examining each in isolation.

Finally, a body of work has analyzed the growth and decline of social networking platforms such as Myspace and Friendster from a model theoretic standpoint. Backstrom et al. (2006) examine the degree to which social network transformation follows divisions along existing links, vs. exogenous factors. They propose a decision tree model which can identify significant network properties impacting the decision of an individual to join a community, or the likelihood of a community to grow. Similarly, Garcia, Mavrodiev, and Schweitzer (2013) examine the collapse of Friendster from a network dynamics perspective to understand the social conditions in which users leave the platform. They find that collapse is predicted by cascades of exiting users in the social network. However, due to their focus on the social network, these models do not provide analyses of where users go when they leave the network nor of what factors motivate leaving beyond the changes in their social network structure (Malinen 2015). Our work offers parallel insight for aggregation platforms (which do not feature an explicit social network) by identifying what factors, other than social structure, influences users to adopt a new platform.

7 Conclusion

By combining analyses of individual surveys, publicly-reported motivations, and broad temporal changes in user behavior, our study offers novel insights into the community profiles of Reddit and its alternatives during the summer of 2015, as well as insights into the phenomenon of community-driven platforms at large. Despite the hysteria concerning a mass exodus from Reddit, our behavior trend analysis shows that no such exodus occurred, though a small user migration was apparent. Furthermore, our study showed a diverse ecosystem of social content aggregation platforms tailoring to different user preferences, that user behaviors shift upon migration, and that Reddit's scale supports a diversity of unique content that helps to hold the attention of its userbase. Community-driven content systems form an important part of many people's online experience and our work highlights how the dynamics of these systems affect that experience.

Acknowledgements

A   Linking Users

Determining whether accounts on different platforms belong to the same person is difficult due to the lack of personally identifying information. However, individuals often reuse usernames to easily remember login credentials or to aid recognition (Liu et al. 2013; Perito et al. 2011).

Username similarity provides evidence that accounts are owned by the same person, but that evidence is only strong if the usernames are rare and so unlikely to be independently chosen. For example two accounts with username `pizza_wizard_1024` are more likely to belong to the same person than two accounts with username `Michael`. We model the probability of observing usernames similarly to Perito et al's (2011) method. First, we measure the frequency with which character 5-grams occur in the active userbases for all platforms. Then, we model the probability of a username u as $P(u) = \prod_{i=1}^{k} p(u[i])$, where k is the username length, and $p(u[i])$ is the frequency of the 5-gram ending on letter i. The surprisal, $-\log P(u)$, is higher for less frequent usernames.

To handle non-exact matches, we perform string alignment, then compute the surprisal accorded by the occurrence of 5-grams ending on letters in common to the usernames, while penalizing letters that differ:

$$S(u_1, u_2) = -\sum_{i=1}^{k_1}(-1)^{I(i,1)}\log p(u_1[i]) - \sum_{i=1}^{k_2}(-1)^{I(i,2)}\log p(u_2[i]),$$

where k_1 and k_2 are the lengths of the first and second username (u_1 and u_2), and $I(i, j)$ is 1 if letter i of u_j is common to both usernames, and 0 otherwise. Since the number of candidate account pairs is large, we only test pairs where the first two letters match.

When run, the algorithm produced a set of of 71,420 potentially linked identities having $S(u_1, u_2) > 200$. Systematic evaluation is difficult because no ground truth is available. By manual inspection, linkages with scores above 300 seemed plausible.

Match Score	Agreement with Algorithm	Linking Precision	α
200–299	0.71	0.41	0.41
300–599	0.75	0.51	0.39
500+	0.80	0.68	0.34
300+	0.76	0.54	0.37
300+ (filtered)	0.78	0.60	0.37

Table 4: Results for double-blind validation of account linking. Kripendorff's alpha (α) provides a measure of inter-annotator agreement. Filtering linkages with Jaro-Winkler similarity ≤ 0.9 improved precision (bottom two rows).

To estimate the algorithm's precision in the absence of ground truth, we devised a double-blind evaluation scheme. Eight annotators were shown, side-by-side, a random user account pair and an algorithmically linked pair, and were asked to identify the more similar pair. They were not told the account usernames nor which pair had been algorithmically linked. Annotators were shown the comments and URLs posted by each account, and the communities (e.g. subreddits) to which the accounts were subscribed. Assuming that (a) accounts belonging to the same person appear more similar than randomly paired accounts, and (b) that accounts incorrectly linked by the algorithm look statistically no more similar than randomly paired accounts, we can calculate the precision of the algorithm η based on the degree to which annotators select the algorithmically linked pairs as more similar, in excess of chance: $\eta = 2a - 1$, where a is the fraction of cases where annotators select the algorithmically linked pair; see Table 4. When annotators labeled 300 units in triplicate, they selected the algorithmically linked pairs in excess of chance ($p < 0.01$) for all ranges of match-score.

The annotation task was quite challenging, as evidenced by the relatively low inter-annotator agreement (see Table 4). The task's difficulty will tend to cause annotators to fail to recognize correctly linked accounts, driving their annotations toward chance. This unduly penalized the calculated algorithm precision, as annotation errors are attributed to incorrect algorithmic links. As a result, the calculated precisions presented in Table 4 reflect a *lower bound*.

Based on manual inspection of the usernames belonging to algorithmically linked accounts, we judged that in some cases, mismatching characters (i.e edit distance) were not being sufficiently penalized. As an additional step to increase precision, we kept only linked accounts whose usernames had a Jaro-Winkler string similarity greater than 0.9, which helped to increase precision (see Table 4, last two rows). The analysis presented in the paper was performed using algorithmically linked accounts with a match score of at least 300 and Jaro-Winkler string similarity greater than 0.9.

References

Alfonso, F. 2014. Reddit Strips r/technology's Default Status amid Moderator Turmoil. http://www.dailydot.com/news/reddit-censorship-technology-drama-default/. The Daily Dot.

Backstrom, L.; Huttenlocher, D.; Kleinberg, J.; and Lan, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *KDD*, 44–54. ACM.

Benevenuto, F.; Rodrigues, T.; Cha, M.; and Almeida, V. 2009. Characterizing user behavior in online social networks. In *9th ACM SIGCOMM Internet Meas*, 49–62. ACM.

Bergstrom, K. 2011. "dont feed the troll": Shutting down debate about community expectations on reddit.com. *First Monday* 16(8).

Bernstein, M. S.; Bakshy, E.; Burke, M.; and Karrer, B. 2013. Quantifying the invisible audience in social networks. In *CHI*.

Cosley, D.; Frankowski, D.; Kiesler, S.; Terveen, L.; and Riedl, J. 2005. How oversight improves member-maintained communities. In *CHI*, 11–20. ACM.

Garcia, D.; Mavrodiev, P.; and Schweitzer, F. 2013. Social resilience in online communities: The autopsy of friendster. In *1st ACM conf online social networks*, 39–50. ACM.

Gong, W.; Lim, E.-P.; and Zhu, F. 2015. Characterizing silent users in social media communities. In *ICWSM*. AAAI.

Guimaraes, A.; Couto da Silva, A. P.; and Almeida, J. M. 2015. Temporal analysis of inter-community user flows in online knowledge-sharing networks. In *SIGIR 2015 Workshop on Temp, Soc and Spatially-aware Info Access*.

Hermosa, M., and Hachtmann, F. 2013. The process of social media migration among young professionals: A grounded theory. In *Proceedings of the Conference of the American Academy of Advertising*.

Hill, K. 2012. "reddit co-founder alexis ohanian's rosy outlook on the future of politics". http://www.forbes.com/sites/kashmirhill/2012/02/02/reddit-co-founder-alexis-ohanians-rosy-outlook-on-the-future-of-politics/3/. Forbes.

Hou, A. C.; Chen, Y.-C.; Shang, R.-A.; and Chern, C.-C. 2012. The post adoption switching of social network service: A human migratory model. In *PACIS*.

Kamen, M. 2015. Reddit in revolt over anti-harrassment policy enforcement. http://www.wired.co.uk/news/archive/2015-06/11/reddit-revolt. Wired.

Kietzmann, J. H.; Hermkens, K.; McCarthy, I. P.; and Silvestre, B. S. 2011. Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons* 54(3):241–251.

Kumar, S.; Zafarani, R.; and Liu, H. 2011. Understanding user migration patterns in social media. In *AAAI*.

Lampe, C.; Wash, R.; Velasquez, A.; and Ozkaya, E. 2010. Motivations to participate in online communities. In *CHI*.

Lee, E. S. 1966. A theory of migration. *Demography* 3(1):47–57.

Lin, K.-Y., and Lu, H.-P. 2011. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. *Computers in Human Behavior* 27(3):1152–1161.

Liu, F., and Xiao, B. 2013. A theoretical model and empirical investigation of social networking site users' switching intention. In *PACIS*, 75.

Liu, J.; Zhang, F.; Song, X.; Song, Y.-I.; Lin, C.-Y.; and Hon, H.-W. 2013. What's in a name?: an unsupervised approach to link users across communities. In *WSDM*.

Lynch, B., and Swearingen, C. 2015. Why we shut down reddit "ask me anything" forum. http://www.nytimes.com/2015/07/08/opinion/why-we-shut-down-reddits-ask-me-anything-forum.html. The New York Times.

Malinen, S. 2015. Understanding user participation in online communities: A systematic literature review of empirical studies. *Computers in Human Behavior* 46:228–238.

Matias, J. N. 2015. What just happened on reddit? Understanding the moderator blackout. http://socialmediacollective.org/2015/07/09/what-just-happened-on-reddit-understanding-the-moderator-blackout/.

Perito, D.; Castelluccia, C.; Kaafar, M. A.; and Manils, P. 2011. How unique and traceable are usernames? In *Privacy Enhancing Technologies*, 1–17. Springer.

Polites, G. L., and Karahanna, E. 2012. Shackled to the status quo: The inhibiting effects of incumbent system habit, switching costs, and inertia on new system acceptance.

Ruths, D., and Pfeffer, J. 2014. Social media for large studies of behavior. *Science* 346(6213):1063–1064.

Stoddard, G. 2015. Popularity and quality in social news aggregators: A study of reddit and hacker news. In *WWW*.

Wise, K.; Hamman, B.; and Thorson, K. 2006. Moderation, response rate, and message interactivity: Features of online communities and their effects on intent to participate. *Journal of Computer-Mediated Communication* 12(1):24–41.

Zhang, K. Z.; Lee, M. K.; Cheung, C. M.; and Chen, H. 2009. Understanding the role of gender in bloggers' switching behavior. *Decision Support Systems* 47(4):540–546.

Zhong, C.; Shah, S.; Sundaravadivelan, K.; and Sastry, N. 2013. Sharing the loves: Understanding the how and why of online content curation. In *ICWSM*.

Zhu, H.; Chen, J.; Matthews, T.; Pal, A.; Badenes, H.; and Kraut, R. E. 2014. Selecting an effective niche: an ecological view of the success of online communities. In *CHI*.

Zhu, H.; Kraut, R. E.; and Kittur, A. 2014. The impact of membership overlap on the survival of online communities. In *CHI*.