Tracking Reddit User Interests Over Time

Wendy Li, Hao Mao, Hao-Hsiang Song, and Ben Y. Zhao

Department of Computer Science, University of Chicago, Chicago, IL USA

{wendyli, mouhao, haohsiang, ravenben}@uchicago.edu

ABSTRACT

Data confidentiality policies of numerous popular social networks have severely limited researchers' access to large-scale datasets. While there have been some studies pertaining to detailed dynamics of the movement of communities in large online social networks over time, there has been very little research done in terms of monitoring a content dataset and how users' interest change over time. In this paper, we present the results of analyzing detailed dynamics of Reddit from July 2008 to Jan 2013. We analyze the dynamics at percommunity and network-wide granularities, determining how users' interests shift over time. We find that, contrary to intuition, Reddit exhibits online social network properties, in which users interact together in different categories.

1. INTRODUCTION

Online social networks (OSNs) like Reddit are among the most frequently visited websites on the Internet and have revolutionized the way people get news and information [2]. Traditionally, it was nearly impossible to analyze information propagations. Data like timestamps, volume, and participants were simply inaccessible by virtue of word-of-mouth. The coming of OSNs opens up opportunities to mine statistics like frequencies, disparities, and even fingerprints to discover how this information is propagated. Further understanding of how information is spread will enable us to more effectively use these OSNs to spread our ideas.

Past studies pertaining to OSNs and content have examined the relationship between the success of social media content and various factors such as its title, the community it is posted to, whether it has been posted before, and the time it is posted [7]. Others have sought to find the relationship between the popularity of an online article and its quality,

concluding that popularity on Reddit is a relatively strong reflection of intrinsic quality [13]. Further studies have even shown the importance of understanding these OSNs as reflections of the cultures they support [11]. Nonetheless, these content based papers fail to look at network topology and the change over time, missing vital findings about information spread that such analysis affords.

In fact, while some work has been done in terms of network growth and link prediction, these works fail to integrate content into their findings. For instance, one study presents the results of analyzing the detailed dynamics in a large Chinese social network over 2 years using different granularities to determine how much users are influenced by dynamics processes at different scales [17]. Another implements and evaluates link prediction algorithms, finding that coupling built filters with current algorithms dramatically improves prediction accuracy across all traces and algorithms [9].

Evidently, past research has focused on these two realms of content and network topology separately. In contrast, our study is different in that we are tying these two topics, which have traditionally been addressed as two different problems, together. Namely, in the realm of network topology, there is very little research done in terms of monitoring a content dataset and how users' interest change over time and that topic is what we are interested in. Towards this end, we utilize a Reddit dataset, a content dataset that spans a period of roughly five years [6]. Reddit is an American social news aggregation and discussion website. Users submit content such as links, images, or text to the site in the form of a post, which other users can comment on. Reddit is far from the traditional OSN, being much more discussion forum based, implying that users are expected to not interact together on different topics, staying confined to their own topic

of interest. Our content dataset enables us to not only monitor weighted edge changes and how communities or clusters in the network migrate and change over time, but map these structural changes to a tangible content category or topic, effectively enabling us to track users' shift in interest over time.

In this paper, we explore our questions empirically through analyzing Reddit at the level of user communities. We want to analyze user behavior pertaining to interest. Thus, we analyze a dynamic graph of a sequence of detailed time-stamped events capturing the ongoing interactions between users and growth of Reddit, a large discussion forum. In fact, in the month of December 2017 alone, Reddit obtained 1.7 billion unique visits [12]. Evidently, Reddit is a large OSN and a force to be reckoned with in terms of spreading information, being designed explicitly for sharing and discussing content. We focus our analysis to the time period between July 2008 to Jan 2013, right after Reddit became open-sourced in June 2008 [4] and after it began allowing users to create their own custom Reddits and Subreddits [10]. Ultimately, this particular time frame allows us to capture Reddit's initial burst of growth and sustained growth and popularity.

We used the HTML dataset obtained by Lakkaraju, McAuley, and J. Leskovec, 2013 of Reddit.com [7]. Specifically, the dataset consists of HTML files of all the original posts and comments, including timestamps for all these interactions. In addition, in total, it consists of 132,308 reddit.com submissions, making for a total of 6.2 million edges and 642,035 unique users. Thus, our dataset is notable because of its scale, its content, and it having the timestamps associated with each interaction event.

The way we analyze this dataset is on the cluster or community level, monitoring the creation, merging, deletions, and evolving of interest communities through time in the network. We consider the creation of a link between users as defined as a weighted interaction between two users. That is, a user commenting on another user's post is considered an interaction and the frequency determines the weight of the interaction. Then, we focus on tracking communities because, since the basic principle of Reddit consists of sharing posts pertaining to a particular topic and commenting on

those shared posts of a particular topic of interest, the behavior of a user is likely to be significantly impacted by the actions of users in the network that also share the same interests.

To track these user interest communities, we first define and track the evolution of user interest communities over time. Specifically, we use the Louvain algorithm to detect the various interest communities, tracking their creation, merging, dissolution, and transformation over time [17].

Key Findings. Through our analysis, we found several significant findings. First, we find that overall, the network grows exponentially, which is expected for an OSN. Secondly, we also found that as the network gets bigger and takes off in popularity, user retain rate is higher and more and more interest communities stick together instead of disappearing or disassociating. In addition, we can identify clear evolutions in user interest over time, with certain topics being added and deleted through time and other topics rising in popularity over time. Finally, by analyzing user interest dynamics over time, we conclude that, unlike what intuition would suggest, Reddit acts not only as discussion forum, but also as an OSN.

The rest of this paper is organized as follows. In the next section we describe how we constructed our graphs from the HTML files of the original Reddit posts and their comments. Section 3 describes the implementation of defining and tracking communities in our dataset. Section 4 details the evaluation methodology we use to quantify the movements of communities and maps these changes to shifts in user interests through time. Then, we summarize related research in Section 5 and conclude in Section 6.

2. EDGE CREATION

We begin by defining an edge between two users as an interaction between two users. For instance, a directed edge between A to B would consist of user A commenting on user B's post. To find all such interactions, we build an HTML parser that parses out the usernames of a post's author, the usernames of each commenter, the timestamps of each action, and the category of the post, denoted as the Subreddit tag. Then, after finding all such interactions and their respective attributes of time and category, we look

that the timestamps of all these actions throughout the timespan of the data, separating the events into four month long snapshots for further analysis.

Namely, we group the interaction data into the groups of the following snapshots according to its timestamp. For instance, the first and second snapshots would be from July 2008 to October 2008 and November 2008 to February 2009 respectively. Of note, the last snapshot spans from November 2012 to January 2013, being only 3 months in length instead of the usual 4 months.

In constructing weighted edges between two users, we aggregate all the interactions together. That is, we consider each comment between two users contributing a weight of 1 and mutual comments between two users adding an additional weight of 2 to these base weights of 1 per comment because such interactions are clearly stronger interactions as they are mutual. We then consolidate the directed interactions into nondirected interactions through a consolidation process, appending all categories that the user participated in to a list of categories for the user node, denoted as an attribute in the graph.

In performing this process for each snapshot, we achieved our snapshot graphs of our dataset and visualized using Python's Networkx. We then found key statistics of interest for each snapshot.
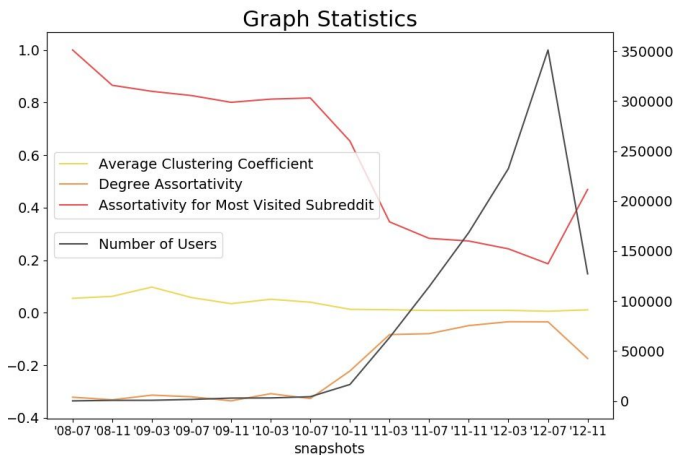

Figure 1. Assortativity for Most Visited Subreddit (red), Average Clustering Coefficient (yellow), Degree Assortativity (orange), and number of active users (black) for each four month snapshot from July 2008 to January 2013. '12-11 only consists of 3 months, partially explaining for lower number of active users.

As shown in Figure 1, interestingly, while the average degree assortativity in Reddit across the time period remains low, from around -0.34 to -0.03, the average assortativity for the most visited subreddits remain significantly higher, suggesting that popular users and categories tend to stick together. Furthermore, as shown in Figure 1, this assortativity drops by roughly 0.5 over 8 months. But, as indicated by the steep black line showing the number of active users, this drop is attributed to the explosion in user growth.

Finally, Reddit's average clustering coefficient of 3.3% suggests Reddit is not very tightly knit. Nevertheless, this coefficient is akin to OSNs like Twitter, Orkut, and Pokec, which have clustering coefficients of 2.15%, 4.13%, and 4.68%. It is slighter higher that of Google+'s, 0.371%, and slightly lower than that of Facebook's, 8.51% [15]. Other networks that are not OSNs have significantly different clustering coefficients. For instance, human social networks, real-world social networks between humans with offline ties, have much higher clustering coefficients, with some as high as 90.4% [15].

3. COMMUNITY CONSTRUCTION

Tracking communities in the presence of network dynamics is a vital step in our analysis of shifting user interest over time. Reddit consists of sharing posts of a particular topic of interest and commenting on those shared posts of a particular topic of interest. Thus, the behavior of a user is likely to be significantly impacted by the actions of users in the network that also share the same interests, thereby posting about those interests. Therefore, tracking the movement of communities of interest will be able to reveal the trends in shift user interest in a scalable manner without having to monitor each change in interest for a particular user.

Communities are defined as network structure as clusters of well-connected nodes. These areas essentially have dense connections inside of the communities, but few or sparse connections outside of the communities. Thus, such a grouping would help to isolate various communities interest groups from each other in such a structure like the large OSN structure of Reddit. Prior research has shown that dynamic community tracking is an NP-hard problem [14]. Thus, for scalability and efficiency, we use the Louvain algorithm to track communities in each snapshot.

In particular, the Louvain algorithm is a scalable community detection algorithm [9] that uses greedy local modularity optimization, making it significantly more scalable and efficient. In addition, it has greater modularity, and all these attributes are important in community detection in such a large dataset like our Reddit one [17]. In fact, it scales to graphs with tens of millions of nodes [17]. Thus, specifically, we utilize the Python community API, using this implementation of the Louvain algorithm to compute the communities in our snapshots. We use a resolution of 1, which represents the time described in [8] and finds the partitions of highest modularities. That is, the stability at time of 1 is equal to the modularity [8]. Furthermore, we also experimented with various resolutions, concluding that a resolution of 1 generated a good number of communities in our dataset that was relatively stable between snapshots. In fact, we found that lower resolutions were too sensitive, providing a too high number of communities.

4. COMMUNITY EVOLUTION

To track the evolution of communities through time, we essentially find the similar users across the snapshots and locate the interest community that they migrate to in the next snapshot.

That is, we compute the percentage of a cluster that goes from one snapshot to another given cluster in the next snapshot, also identifying the percentage of new users emerging in a given clusters and the percentage of users in each cluster that disappears or drops out.

To visualize our results, we utilize sankey diagrams depicting cluster movement from snapshot to snapshot.

Time Frame	Mean outgoing flows per cluster	Mean percentage of each cluster going inactive	Mean number of categories per cluster of source (left)	Mean number of categories per cluster of destination (right)
'08-07 to '08-11	3.14	85.17± 8.57	1.00±0.00	3.08±1.26
'08-11 to '09-03	5.82	88.87±4.33	3.08±1.26	2.45±0.99
'09-03 to '09-07	6.72	85.68±6.21	2.45±0.99	3.06±1.03
'09-07 to '09-11	6.64	86.62±10.46	3.06±1.03	4.41±1.86
'09-11 to '10-03	6.0	91.22±5.04	4.41±1.86	3.86±1.71
'10-03 to '10-07	6.39	91.21±4.75	3.86±1.71	4.15±1.37
'10-07 to '10-11	10.53	85.1±7.79	4.15±1.37	6.64±3.85
'10-11 to '11-03	12.79	75.43±16.97	6.64±3.85	7.64±4.92
'11-03 to '11-07	5.97	84.04±16.64	7.64±4.92	11.84±10.63
'11-07 to '11-11	3.7	82.55±20.28	11.84±10.63	10.80±11.28
'11-11 to '12-03	3.02	79.56±22.52	10.8±11.28	31.89±19.28
'12-03 to '12-07	3.96	69.77±21.00	31.89±19.28	23.95±53.00

Table 1. Cluster movement statistics for each four month snapshot from July 2008 to January 2013

Sankey Flow Graphs[1]:
Each graph depicts the cluster movement between two consecutive snapshots. The lines or bars on the left side of the graph represent the communities in a snapshot of four months. The lines or bars on the right side of the graph represent the communities in the next snapshot of four months. The clusters are normalized and the width of the flow leaving each cluster or bar corresponds to the percentage of users leaving that cluster going to a cluster or bar in the second second snapshot. These interactive sankey graphs include the exact percentages of flow and the large blue cluster on right represents the flow of users that become inactive from the last snapshot to the next.
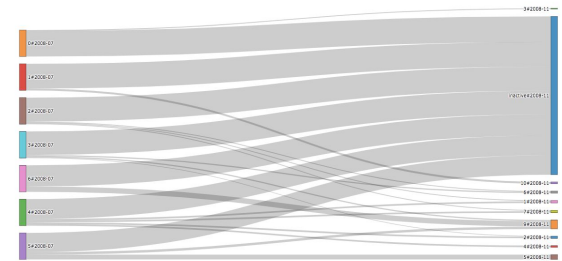

Figure 2. Cluster movement for the 2008-07 snapshot to the 2008-11 snapshot. For instance, 93% of cluster 1 flows to inactive, meaning 93% of that cluster drops out. The large blue cluster on right represents the flow of users from 2008-07 that become inactive in 2008-11.

---

[1] Graphs of 2011-07 to 2011-11, 2011-11 to 2012-03, and 2012-03 to 2012-07 are not depicted because of their scale. View in source code.
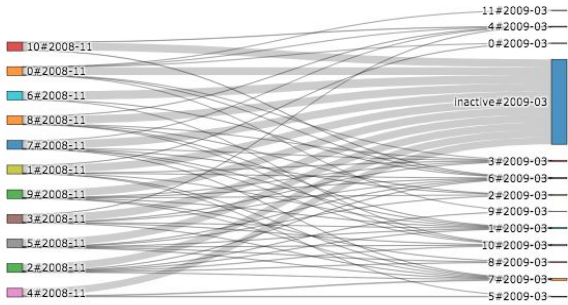
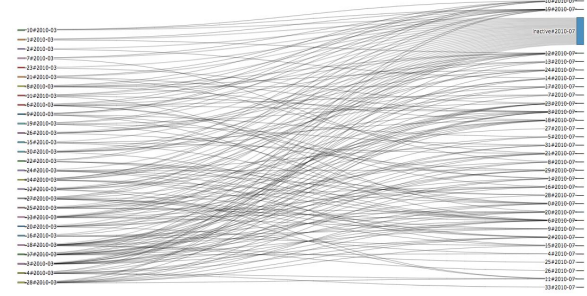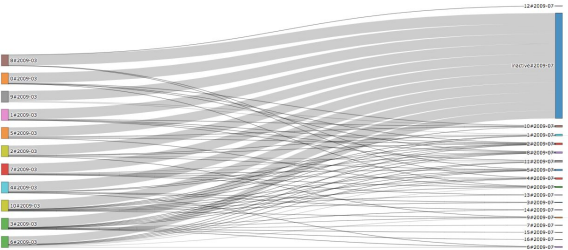**Figure 3. Cluster movement for the 2008-11 snapshot to the 2009-03 snapshot.**

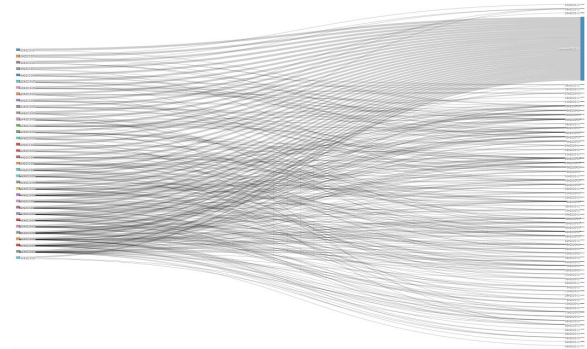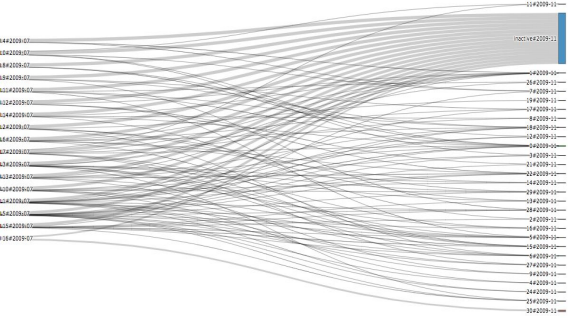**Figure 4. Cluster movement for the 2009-03 snapshot to the 2009-07 snapshot.**

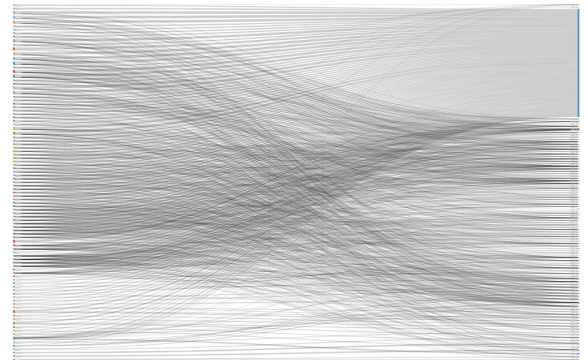**Figure 5. Cluster movement for the 2009-07 snapshot to the 2009-11 snapshot.**

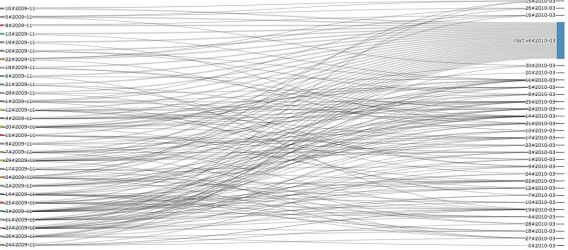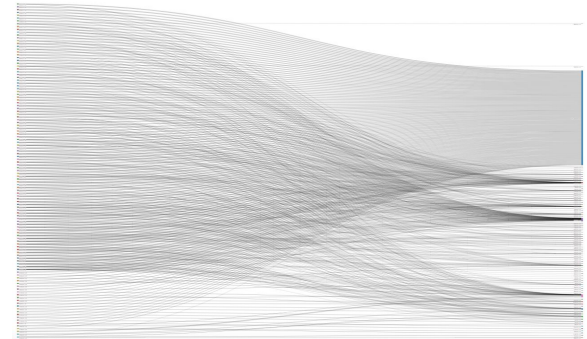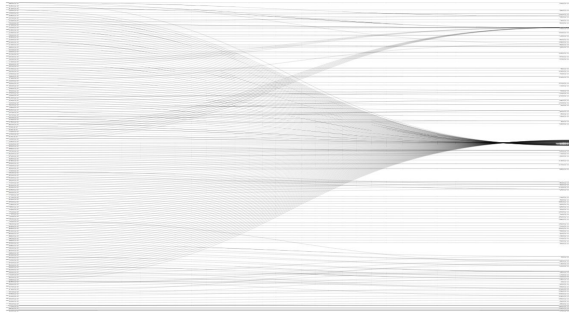**Figure 6. Cluster movement for the 2009-11 snapshot to the 2010-03 snapshot.**

Figure 7. Cluster movement for the 2010-03 snapshot to the 2010-07 snapshot.

Figure 8. Cluster movement for the 2010-07 snapshot to the 2010-11 snapshot.

Figure 9. Cluster movement for the 2010-11 snapshot to the 2011-03 snapshot.

Figure 10. Cluster movement for the 2011-03 snapshot to the 2011-07 snapshot.

Figure 11. Cluster movement for the 2012-07 snapshot to the 2012-11 snapshot.

Through analyzing our flow graphs for each snapshot of four months, it became evident that the majority of each cluster is dropping out from snapshot to snapshot, indicating the volatile nature of which users are active in Reddit. In fact, as shown in Table 1, this user drop out rate is significantly higher during the earlier snapshots, with a drop out rate of 89% of a cluster dropping out in the time frame from 2008-07 to 2008-11, compared to a 70% drop out rate from 2012-03 to 2012-07. This change is especially notable not only because of the significant change in percentage, but also because of the high deviation of 21.00 compared to 6.57 in the earlier snapshot, indicating that the dropout rate varies widely. In fact, 12 clusters in the last snapshot have a drop out rate of 0%, compared to 0 clusters having that drop out rate in the earlier snapshot.

Therefore, it seems that as time goes by and the OSN really takes off, there is a higher user retention rate. Although the mean cluster percentage of going inactive for the majority of the cluster still remains relatively high, notice that it is lower as the snapshots go on. Furthermore, notice how high the standard deviation is, indicating that not every cluster is getting such a high value of drop out.

In addition, as time goes by, the number of categories in each clusters increase while the number of flows per cluster decreases. This seems to suggest that visually, less flows can be seen, thereby making the graph seems sparser, because the social groups have become established. Namely, as shown in Figure 11, much of the clusters have a single horizontal flow to another corresponding cluster, indicating lack of user movement between clusters and therefore, lack of new user interaction. In contrast, Figure 9, which has an average flow of

12.79, is much more dense and cluttered with numerous crossing flows, indicating user movement and interactions across different communities. This visual display of clutter and density is not attributed to user number because, as shown in Figure 1, the snapshot of Figure 11 has more than 100,000 more users than that of Figure 9. If user number contributed, then we would expect to see greater clutter in Figure 11 because it has more users.

Thus, it also seems that in these figures, the visually observed density correlates with increased flow. Moreover, curved flow patterns correspond with interactions across multiple communities and straight lines and straight flow patterns correspond with isolated interactions.

Interest Continuity:

Interest continuity is defined as a user flow moving from one community to another community of the same interest or category and it is measured between two snapshots via flows. A flow is defined as users being grouped in a cluster in a previous snapshot forming a cluster, with or without others, in a later snapshot. When a group of users flows from a cluster to a future one, the dominating interests of the two source and destination clusters are compared, weighted by the relative size of the flow and then averaged. That is, the portion of users continuing on the same interest is calculated for each cluster. We make an adjustment because of the dynamic nature of the Reddit: we assign the weight of the size of active flows in which users are active in both snapshots. Finally, we plot the probability distribution of continuity rate across all clusters using the time span of four months.
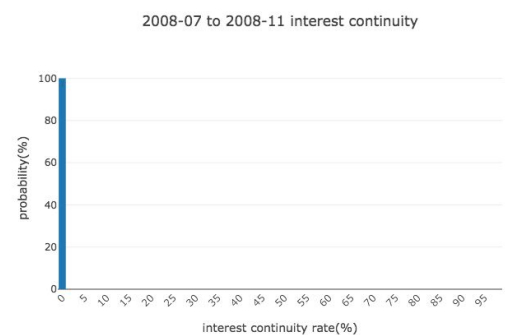

Figure 12. Distribution of interest continuity from the 2008-07 snapshot to the 2008-11 snapshot. For instance, about 100% of

all the clusters will have 0% of interest continuity. That is, all clusters exhibit 100% of its still active users in the next snapshot shifting interest.
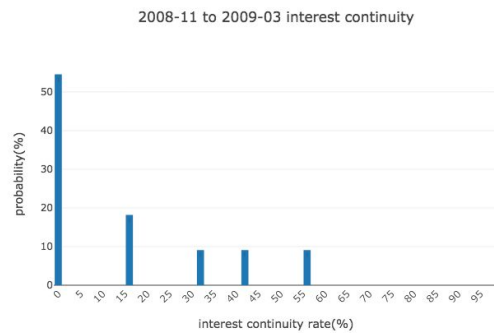

Figure 13. Distribution of interest continuity from the 2008-11 snapshot to the 2009-03 snapshot. For instance, about 18% of all the clusters will have 15% of the cluster maintain the same interest from the 2008-11 snapshot to the 2009-03 snapshot.
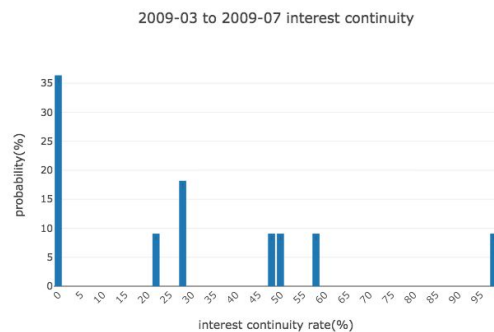

Figure 14. Distribution of interest continuity from the 2009-03 snapshot to the 2009-07 snapshot. For instance, about 36% of all the clusters will have 0% of the cluster maintain the same interest from the 2009-03 snapshot to the 2009-07 snapshot.
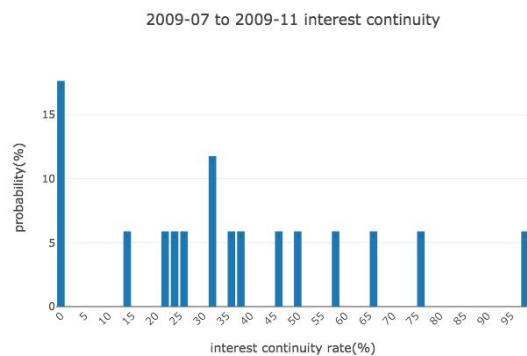

Figure 15. Distribution of interest continuity from the 2009-07 snapshot to the 2009-11 snapshot. For instance, about 6% of the clusters will have 50% of the cluster shift interest.
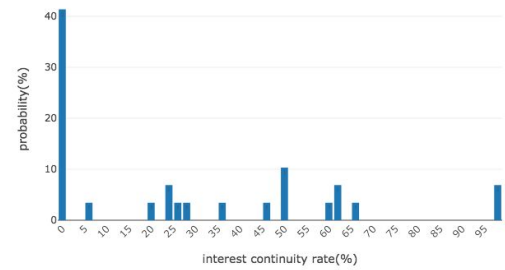

Figure 16. Distribution of interest continuity from the 2009-11 snapshot to the 2010-03 snapshot.
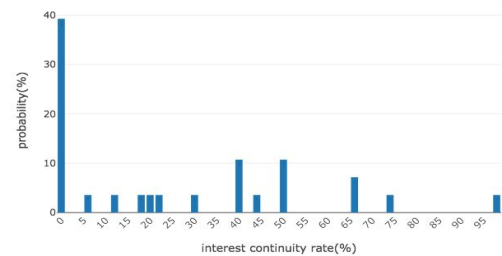
Figure 17.Distribution of interest continuity from the 2010-03 snapshot to the 2010-07 snapshot.
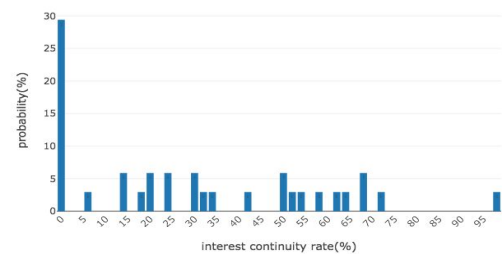
Figure 18. Distribution of interest continuity from the 2010-07 snapshot to the 2010-11 snapshot.
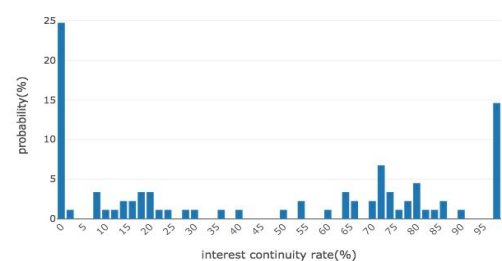

Figure 19. Distribution of interest continuity from the 2010-11 snapshot to the 2011-03 snapshot.

Figure 20. Distribution of interest continuity from the 2011-03 snapshot to the 2011-07 snapshot.

Figure 21. Distribution of interest continuity from the 2011-07 snapshot to the 2011-11 snapshot.

Figure 22. Distribution of interest continuity from the 2011-11 snapshot to the 2012-03 snapshot.
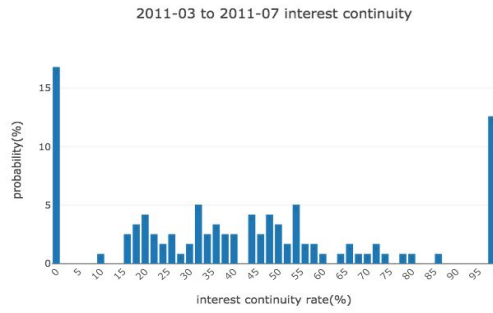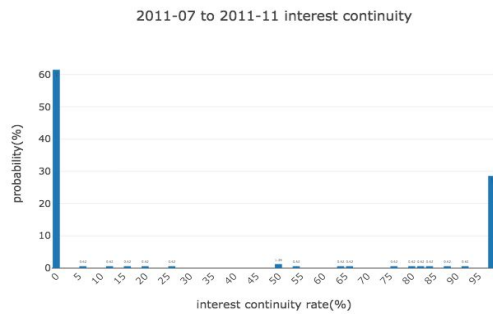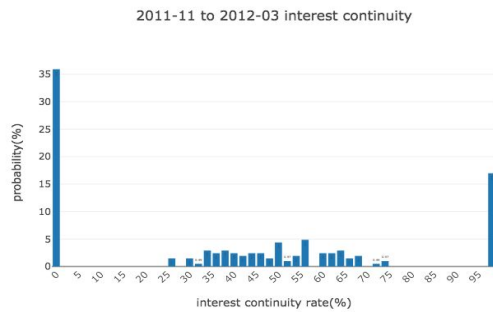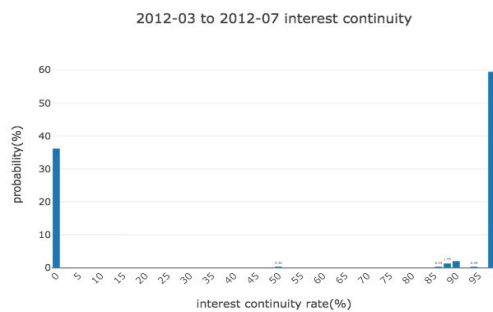

Figure 23. Distribution of interest continuity from the 2012-03 snapshot to the 2012-07 snapshot.
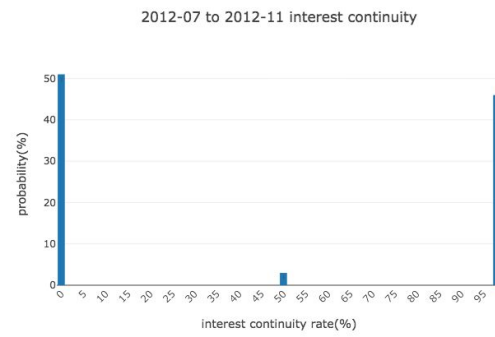

Figure 24. Distribution of interest continuity from the 2012-07 snapshot to the 2012-11 snapshot.

Through analyzing interest continuity across all the snapshots, it is evident that in the beginning of Reddit's development, there is still a greater probability of 0% of the cluster showing interest continuity across all the earlier Figures from 12 to 18, meaning most users change interests. In these snapshots, for the rest of the interest continuities across clusters, there is also a more relatively even distribution, meaning some users in the clusters choose to maintain the same interest. Figure 15 show this trend particularly well. Namely, 0% is clearly the dominating interest continuity probability, but the rest of the interest continuity distributions seem rather uniform relative to each other.

However, especially starting from 2010-11, the time when Reddit begins its exponential growth in users, there becomes a much more polarized divide between user interest continuity as shown by Figure 19. As Figure 19 shows, roughly half of the users choose to stay in the same category, while the other half of the users choose to switch categories. That is, almost 50% of clusters will have 100% of the cluster shift categories. The distribution turns bimodal.

If Reddit was purely a discussion forum and exhibited discussion forum behavior, then we would expect for the majority of users to remain in the same major interest category across snapshots. Nevertheless, clearly, half of the communities of Reddit will change categories, suggesting the OSN behavior of Reddit. Therefore, Reddit seems to act as a OSN for those who want to use it as such and interact together with other users in different categories.

In addition, the increase in average categories of each cluster across time in Reddit also attests to the existence of the OSN behavior of Reddit. In general, as shown in Table 1, as time goes on, both the mean number of categories and standard deviation increase. For example, notice how the last snapshot has an average of 23.95±52.72 categories. The large standard deviation suggests the presence of some clusters with an immense number of categories, indicating the tight user interaction across communities of interest. In fact, the cluster with the largest number of categories has 416 categories, which strongly suggests the formation of a mini social network in Reddit. In contrast, in the same snapshot, there are also multiple clusters with only 1 or 2 categories. Thus, we conclude that Reddit is perhaps a OSN and discussion forum hybrid by the end of its development. We propose that those who want to find a OSN in Reddit can, but for those who wish to stick to a certain topic of interest, Reddit certainly supports those users as well.

Top Categories Breakdown:

Finally, we also analyze the top categories in the clusters across the snapshots. That is, for each snapshot, we extract the dominant category for each cluster in that snapshot and aggregate them. For instance, as Figure 25 shows, in the first snapshot of four months starting at 2008-07, 50% of the clusters have a dominant category of offbeat, while the other 50% of the clusters have the dominant category of pics.
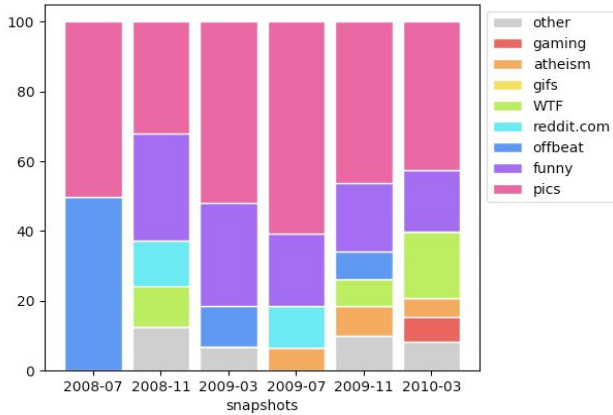

Figure 25. Distribution of Top Categories in clusters for snapshots from the 2008-07 to 2010-03.
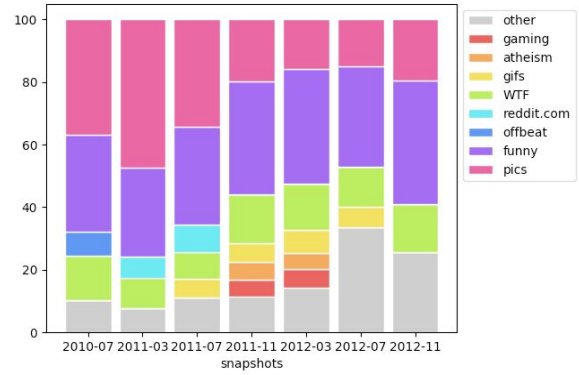

Figure 26. Distribution of Top Categories in clusters for snapshots from the 2010-07 to 2012-11.

From our analysis of the dominating categories in each cluster across the snapshots, it is evident that there is a trend that some categories are introduced and subsequently become an enduring popular category like the "funny" category. Others, like offbeat, are popular and then slowly die out. In fact, pics starts out very popular in the earlier snapshots, but slowly die out in the later snapshots. Adding on, some categories, like atheism, gaming, reddit.com, appear time to time but do not gain widespread popularity. Finally, it is notable that in such a large OSN like Reddit, especially in the later snapshots when there are over 100,000 active users, that there are only ever at most 7 most dominant categories in Reddit, suggesting that most people are only focusing and engaging on the few dominating and popular topics at a time. In short, user interest clearly changes dynamically and shifts throughout the snapshots.

5. RELATED WORK

Several studies have also measured the basic dynamic properties of weighted graphs. For instance, one study also uses temporal user interactions as graph edges in the analysis of FaceBook [16] and another evaluated the growth of a Korean OSN [1]. Moreover, another study performed analysis pertaining to detecting bursts in various temporal snapshots [5]. Nevertheless, unlike these studies, our work is different in that it analyzes the shifting of implicit user interest communities in a large-scale OSN with millions of edges that is densely connected.

Another study also performs a detailed analysis of user dynamics in a large Chinese online social network at the level of individual users, communities, and OSNs. [17] Nonetheless, this study uses user friendships as edges, being unable to relate content, and monitors a large Chinese online social network over a 25-month period, a period of time that is much shorter than our timeframe. Furthermore, their networks consists of an anomalous event of two OSNs merging [17].

Finally, another study uses network structure, specifically communities, to examine key social roles in Reddit, finding that the "answer-person" role is present in the Reddit community and can be identified using an automated method that is based solely on user interactions [3]. Nevertheless, unlike our study, this study does not study user interactions across time, using only the top 100 submissions from the month of July, 2013 [3].

6. CONCLUSION

To conclude, we analyze the dynamics at percommunity and network-wide granularities, determining how users' interests shift over time. We find that, contrary to intuition, Reddit exhibits properties of traditional OSNs, in which users interact together into different categories. Interestingly, as Reddit develops, there is a polarization in user interests shifts. Half of the users remain stagnant, concentrating on one topic of interest, while the other half of the users dramatically change interest categories from snapshot to snapshot, interacting with other users. Finally, user interest clearly change dynamically and shift throughout the snapshots.

Nonetheless, work still remains to be done. We have primarily focused on cluster movement between snapshots, distribution of user interest continuity, and distribution of top categories, but have not yet identified a scalable manner to tie and visualize these entities together. As future work, we will work to find a way to specific track flows across multiple snapshots in a scalable manner and map these flows as moving through specific categories of interest. In addition, we will further explore this shift in user interest through time and how it maps to real world events. Finally, we also want to further examine the implications of Reddit having OSN behavior.

Establishing understanding of the movement of user interest is an open problem, and the solution will enable us to further understand how content is introduced in these OSNs, how topics gains popularity, and how users interact with content.

7. REFERENCES

[1] Ahn, Y., Han, S., Kwak, H., Moon, S., and Jeong, H. Analysis of topological characteristics of huge online social networking services. In *WWW*, 2007.

[2] Amazon.com. The top 500 sites on the web. Retrieved January 13, 2018, from https://www.alexa.com/topsites

[3] Buntain, C., and Golbeck, J. Identifying Social Roles in reddit Using Network Structure. In *WWW*, 2014.

[4] Huffman, S. 2008. blog.reddit – what's new on reddit: reddit goes open source. Redditblog.com. Retrieved May 24, 2015.

[5] Kumar, R., Novak, J., Raghavan, P., and Tomkins, A. On the bursty evolution of blogspace. In *WWW*, 2005.

[6] Lakkaraju, H. Web data: Reddit submissions. Retrieved January 13, 2018, from https://snap.stanford.edu/data/web-Reddit.html

[7] Lakkaraju, H., McAuley, J., and Leskovec, J. What's in a name? Understanding the Interplay between Titles, Content, and Communities in Social Media. In *ICWSM*, 2013.

[8] Lambiotte, R., Delvenne J. C., Barahona, M. 2015. Laplacian Dynamics and Multiscale Modular Structure in Networks. In *IEEE Transactions on Network Science and Engineering*, 1, 2, 76-90.

[9] Liu, Q., Tang, S., Zhang, X., Zhao, X., Zhao, B. Y., and Zheng, H. Network Growth and Link Prediction Through an Empirical Lens. In *IMC*, 2016.

[10] Ohanian, A. 2008. blog.reddit – what's new on reddit: make your own reddit. Redditblog.com. Retrieved May 24, 2015.

[11] Potts, L., and Harrison, A. Interfaces as rhetorical constructions: reddit and 4chan during the boston marathon bombings. In *SIGDOC*, 2013.

[12] Statista. Combined desktop and mobile visits to Reddit.com from April 2017 to December 2017. https://www.statista.com/statistics/443332/reddit-monthly-visitors/

[13] Stoddard, G. Popularity and Quality in Social News Aggregators: A Study of Reddit and Hacker News. In *WWW*, 2015.

[14] Tantipathananandh, C., Berger-Wolf, T., and Kempe, D. A framework for community identification in dynamic social networks. In *SIGKDD*, 2007.

[15] The Koblenz Network Collection. http://konect.uni-koblenz.de/networks/

[16] Viswanath, B., Mislove, A., Cha, M., and Gummadi, K. P. On the evolution of user interaction in facebook. In *WOSN*, 2009.

[17] Zhao, X., Sala, A., Wilson, C., Wang, X., Gaito, S., Zheng, H., and Zhao, B. Y. Multi-scale Dynamics in a Massive Online Social Network. In *IMC*, 2012.