Navigating the massive world of reddit: Using backbone networks to map user interests in social media

Randal Olson, Zachary P Neal

In the massive online worlds of social media, users frequently rely on organizing themselves around specific topics of interest to find and engage with like-minded people. However, navigating these massive worlds and finding topics of specific interest often proves difficult because the worlds are mostly organized haphazardly, leaving users to find relevant interests by word of mouth or using a basic search feature. Here, we report on a method using the backbone of a network to create a map of the primary topics of interest in any social network. To demonstrate the method, we build an interest map for the social news web site reddit and show how such a map could be used to navigate a social media world. Moreover, we analyze the network properties of the reddit social network and find that it has a scale-free, small-world, and modular community structure, much like other online social networks such as Facebook and Twitter. We suggest that the integration of interest maps into popular social media platforms will assist users in organizing themselves into more specific interest groups, which will help alleviate the overcrowding effect often observed in large online communities.

Navigating the massive world of reddit: Using backbone networks to map user interests in social media

Randal S. Olson[1,*] and Zachary P. Neal[2]

[1]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
[2]Department of Psychology, Michigan State University, East Lansing, MI 48824, USA
[*]E-mail: olsonran@msu.edu

Abstract

In the massive online worlds of social media, users frequently rely on organizing themselves around specific topics of interest to find and engage with like-minded people. However, navigating these massive worlds and finding topics of specific interest often proves difficult because the worlds are mostly organized haphazardly, leaving users to find relevant interests by word of mouth or using a basic search feature. Here, we report on a method using the backbone of a network to create a map of the primary topics of interest in any social network. To demonstrate the method, we build an interest map for the social news web site reddit and show how such a map could be used to navigate a social media world. Moreover, we analyze the network properties of the reddit social network and find that it has a scale-free, small-world, and modular community structure, much like other online social networks such as Facebook and Twitter. We suggest that the integration of interest maps into popular social media platforms will assist users in organizing themselves into more specific interest groups, which will help alleviate the overcrowding effect often observed in large online communities.

Introduction

In the past decade, social media platforms have grown from a pastime for teenagers into tools that pervade nearly all modern adults' lives [1]. Social media users typically organize themselves around specific interests, such as a sports team or hobby, which facilitates interactions with other users who share similar interests. For example, Facebook users subscribe to topic-specific "pages" [2], Twitter users classify their tweets using topic-specific "hashtags" [3], Del.ici.ous users bookmark links with topic-specific tags [4, 5], and reddit users post and subscribe to topic-specific sub-forums called "subreddits" [6].

These interest-based devices provide structure to the growing worlds of social media, and are essential for the long-term success of social media platforms because they make these big worlds feel small and navigable. However, navigation of social media is challenging because these worlds do not come with maps [7, 8]. Users are often left to discover pages, hashtags, or subreddits of interest haphazardly, by word of mouth, following other users' "votes" or "likes", or by using a basic search feature. Owing to the scale-free structure of most online social networks, these elementary navigation strategies result in users being funnelled into a few large and broad interest groups, while failing to discover more specific groups that may be of greater interest [9, 10]. This observation leads to the question: Is it possible to build a map to aid users in the discovery of relevant interests on social networks?

In this work, we combine techniques for network backbone extraction [11] and community detection [12] to construct a roadmap that provides an alternative method for social media users to navigate (i.e., find relevant interests) these social networks by identifying related interest groups and suggesting them to users. We implement this method for the social news web site reddit [13], one of the most visited social media platforms on the web [14], and produce an interactive map of all of the subreddits. An interactive version of the reddit interest map is available online [15].

Once we construct this map, we then ask: Does the reddit backbone network display the same complex network characteristics observed in many other real-world networks? By viewing subreddits as nodes linked by users with common interests, we find that the reddit social media world has a scale-free, small-world, and modular community structure. The scale-free property is the expected outcome of a preferential attachment process and helps explain the challenges of haphazard navigation. Additionally, the small-world property explains how the big world of reddit can seem small and navigable to users when it is mapped out. Finally, the modular community structure in which narrow interest-based subreddits (e.g., dubstep or rock music) are organized into broader communities (e.g., music) allows users to easily identify related interests by zooming in on a broader community. We suggest that the integration of such interest maps into popular social media platforms will assist users in organizing themselves into more specific interest groups, which will help alleviate the overcrowding effect often observed in large online communities [6].

Further, this work releases and provides an overview of a data set of over 850,000 anonymized reddit user's interests, thus establishing another standard real-world social network data set for researchers to study. This is useful because, although reddit is among the largest online social networks and has been identified as a starting point for the viral spread of memes and other online information [16], it has been relatively understudied [6, 17, 18]. This data set can be downloaded online at [19].

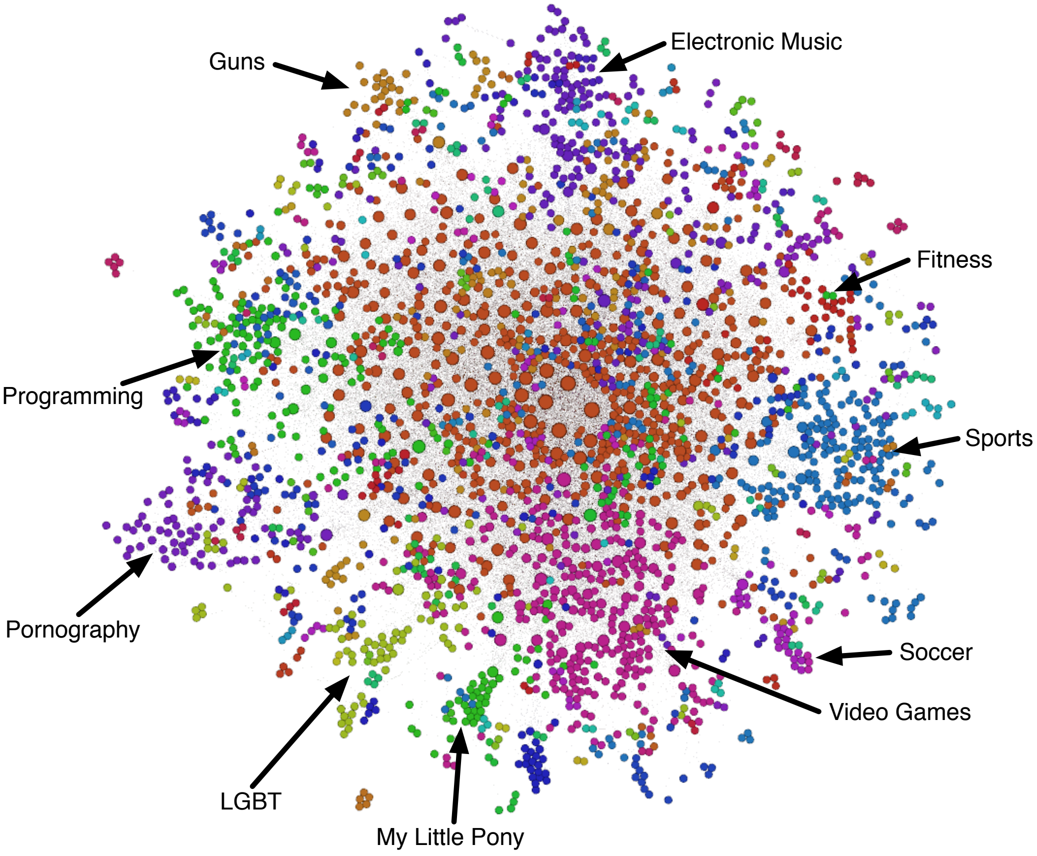

Figure 1: **Reddit interest network.** The largest components of the reddit interest network is shown with 10 interest meta-communities annotated; it closely matches the structure of other online social networks including Flickr and Yahoo360 [20]. Each node is a single subreddit, where color indicates the interest meta-community that the subreddit is a member of. Nodes are sized by their weighted PageRank to provide an indication of how likely a node is to be visited, and positioned according to the OpenOrd layout in Gephi to place related nodes together. An interactive version of the reddit interest map is available online at `http://rhiever.github.io/redditviz/clustered/`

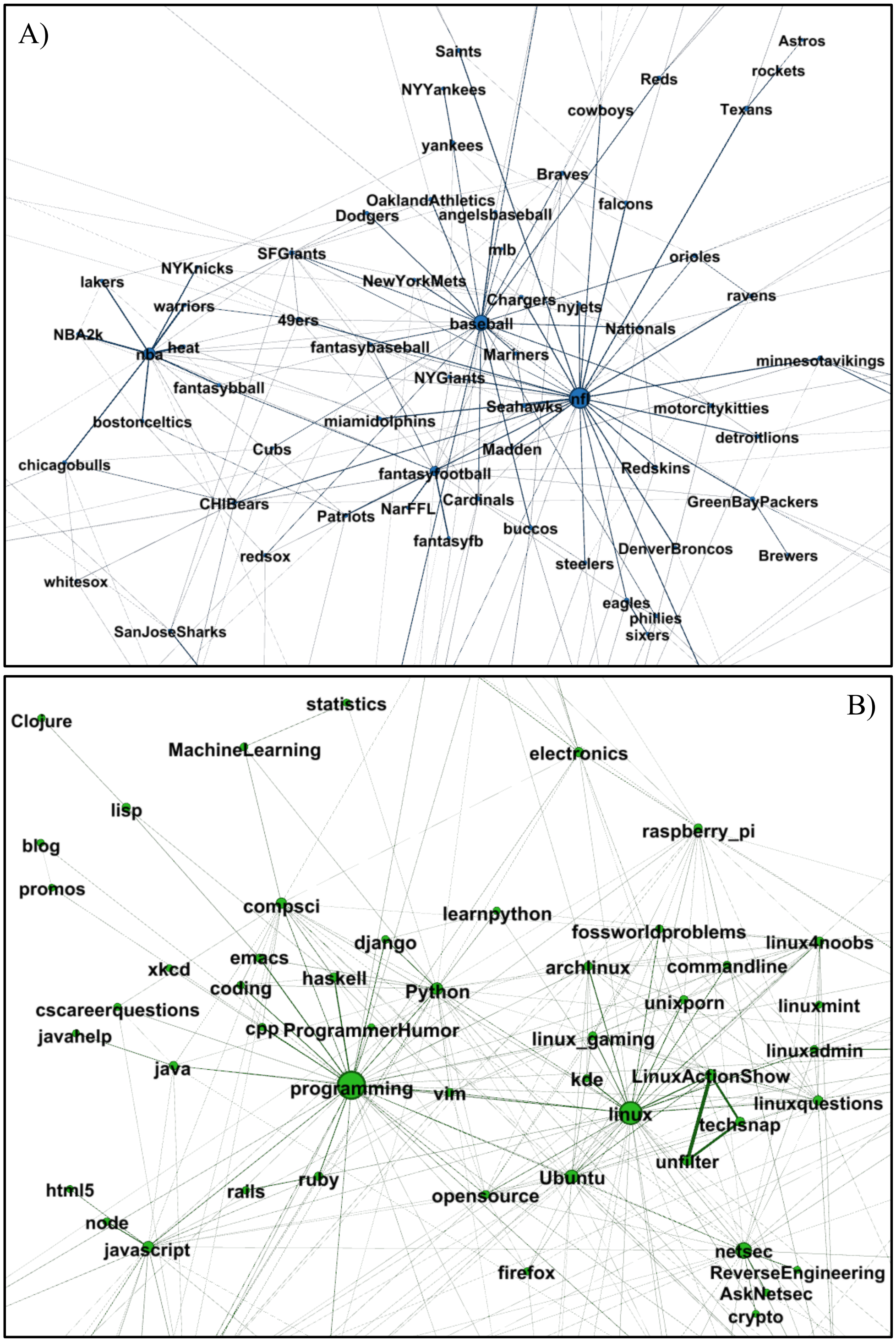

Figure 2: **Example reddit interest meta-communities.** Pictured are several topic-specific subreddits composing a meta-community around a broad topic such as sports (A) or programming (B). Each node is a subreddit, and each edge indicates that a significant portion of the posters in the two subreddits post in both subreddits (see Methods).

Results

Reddit interest map

For the final version of the reddit interest map, we use the backbone network produced with $\alpha = 0.05$ (see Methods). This results in a network with 59 distinct clusters, which we call *interest meta-communities*. In Figure 1, the nodes (i.e., subreddits) are sized by their weighted PageRank [21, 22] to provide an indication of how likely a node is to be visited, and positioned according to the OpenOrd layout in Gephi [23] to place related nodes together.

Through this method, we immediately see several distinct interest meta-communities, 10 of which are annotated in Figure 1. These interest meta-communities act as starting points in the interest map to show the broad interest categories that the entire reddit community is discussing. From these starting points, users can zoom in on a single broad interest category to find subreddits dedicated to more specific interests, as shown in Figure 2. Notably, there is a large, orange interest meta-community in the center of the interest map that overlaps with several other interest meta-communities. This orange interest meta-community represents the most popular, general interest subreddits (e.g., "pictures" and "videos") in which users of all backgrounds regularly participate, and thus are expected to have considerable overlap with many other communities.

Figure 2 depicts zoomed-in views of two interest meta-communities annotated in Figure 1. In Figure 2A, the "sports" meta-community, specific sports teams are organized around the corresponding sport that the teams play in. For example, subreddits dedicated to discussion of the Washington Redskins or Denver Broncos – relatively small, specific subreddits – are organized around the larger, more general interest NFL subreddit where users discuss the latest NFL news and games. Similarly in Figure 2B, the "programming" meta-community, subreddits dedicated to discussing programming languages such as Python and Java are organized around a more general programming subreddit, where users discuss more general programming topics.

This backbone network structure naturally lends itself to an intuitive interest recommendation system. Instead of requiring a user to provide prior information about their interests, the interest map provides a hierarchical view of all existing user interests in the social network. Further, instead of only suggesting interests immediately related to the user's current interest(s), the interest map recommends interests that are potentially two or more links away. For example in Figure 2A, although the Miami Heat and Miami Dolphins subreddits are not linked, Miami Heat fans may also be fans of the Miami Dolphins. A traditional recommendation system would only recommend NBA to a Miami Heat fan, whereas the interest map also recommends the Miami Dolphins subreddit because they are members of the same interest meta-community.

Network properties

In Figure 3, we show a series of network statistics to provide an overview of the backbone reddit interest network. These network statistics are plotted over a range of α cutoff values for the backbone reddit interest network (see Methods) to demonstrate that the interest network we chose in Figure 1 is robust to relevant α cutoff values.

As expected, the majority of the edges are pruned by an α cutoff of 0.05 (Figure 3, top left). This result demonstrates that the backbone interest network is stable with an α cutoff ≤ 0.05, which is the most relevant range of α cutoffs to explore. Surprisingly, 80% of the subreddits that we investigated – roughly 12,000 subreddits – do not have enough users that consistently post in another subreddit to maintain even a single edge with another subreddit. The majority of these 12,000 subreddits likely do not have any significant edges due to user inactivity, e.g., some subreddits have only a single user that frequently posts to them (Table **??**). Another factor that likely contributes to the 12,000 unlinked subreddits is temporary interests, i.e., an interest such as the U.S. Presidential election that temporarily draws a large number of people together, but eventually fades into obscurity again.

Next, we are interested in exploring whether the backbone reddit interest network is a scale-free network, where preferential attachment to subreddits results in a few extremely popular (i.e., connected) subreddits and mostly unpopular subreddits. As such, scale-free networks are known to have node degree distributions that fit a power law [9, 10]. Regardless of the α cutoff, we observed that the node degree distribution of all backbone reddit interest networks fit a power law ($R^2 \approx 0.91$ for $k \geq 50$; Figure 3, top right). This

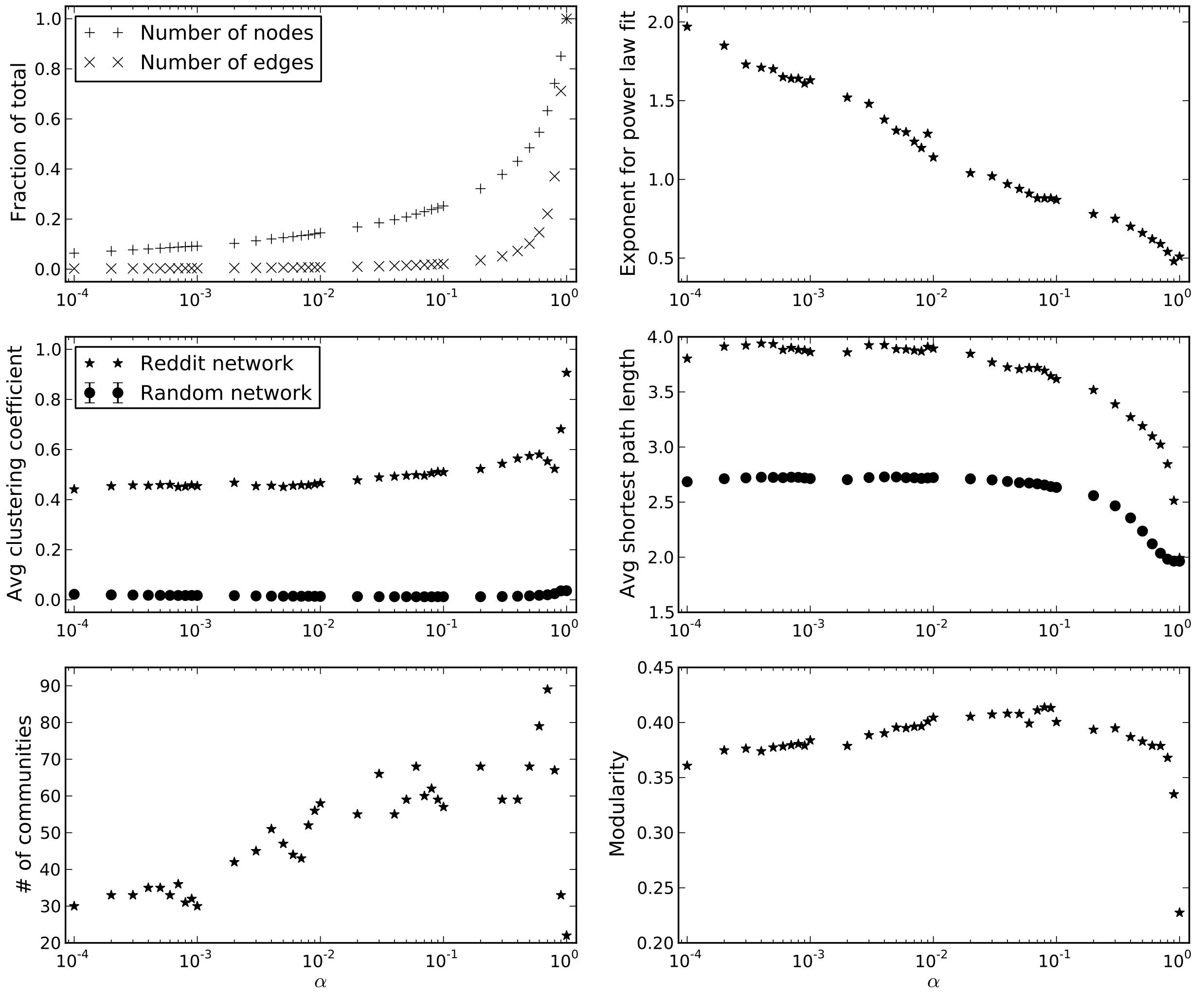

Figure 3: **Network statistics for the backbone network.** Sensitivity analysis of the largest connected component of the reddit interest network over a range of α cutoff values. Lower α means that fewer statistically significant edges are pruned. In general, this sensitivity analysis shows that the backbone interest network is stable for α cutoff values ≤ 0.05. Error bars for the Erdős-Rényi random networks are two standard deviations over 30 random networks, and are too small to show up on the graph. Note the logarithmic scale of the x-axis.

scale-free network structure is likely partially due to reddit's default subreddit system [24], where newly registered users are subscribed to a set of 20 subreddits by default. As expected, as α approaches zero, the power law exponent also approaches zero. This occurs because as the α cutoff used for extracting the backbone network approaches zero, more edges are retained in the network and the structure approaches a fully connected graph in which the degree distribution is flat (i.e., characterized by a trivial power law with an exponent of zero).

Furthermore, we want to confirm that the backbone reddit interest network is a small-world network [25]. Small-world networks are known to contain numerous clusters, as indicated by a high average clustering coefficient, with sparse edges between those clusters, which results in an average shortest path length between all nodes ($\mathrm{L_{sw}}$) that scales logarithmically with the number of nodes (N):

$$\mathrm{L_{sw}} \approx \log_{10}(\mathrm{N}) \tag{1}$$

Figure 3 (middle left and middle right) depicts the average clustering coefficient and shortest path length for all nodes in the backbone reddit interest network. Compared to Erdős-Rényi random networks with the same number of nodes and edges, the backbone network has a significantly higher average clustering coefficient. Similarly, the measured average shortest path length of the backbone network (α cutoff = 0.05) follows Equation 1, with $\mathrm{L_{sw}} = \log_{10}(2,347) = 3.37 \approx 3.71$ from Figure 3 (middle right). Thus, the backbone reddit interest backbone network qualitatively appears to exhibit small-world network properties.

To quantitatively determine whether the reddit interest network exhibits small-world network properties, we used the small-worldness score ($\mathrm{S_G}$) proposed in [26]:

$$\mathrm{S_G} = \frac{\mathrm{C_G}/\mathrm{C_{rand}}}{\mathrm{L_G}/\mathrm{L_{rand}}} \tag{2}$$

where C is the average clustering coefficient, L is the average shortest path length between all nodes, G is the network the small-worldness score is being computed for, and "rand" is an Erdős-Rényi random network with the same number of nodes and edges as G. If $\mathrm{S_G} > 1$, then the network is classified as a small-world network. For the backbone reddit network, we calculated $\mathrm{S_G} = 14.2$ ($\mathrm{P} < 0.001$), which indicates that the reddit interest network exhibits small-world network properties.

Now that we know that the backbone reddit interest network is scale-free and exhibits small-world network properties, we want to study the community structure of the backbone network. Shown in Figure 3 (bottom right), the backbone network exhibits a consistently high modularity score with an α cutoff as high as 0.9, implying that even a slight reduction in the number of edges in the backbone network reveals the reddit interest community structure. Correspondingly, depicted in Figure 3 (bottom left), the number of identified communities (i.e., clusters) remains relatively low until the α cutoff is reduced to ≤ 0.9. As the α cutoff is reduced, the number of identified communities generally decreases, which coincides with the loss of nodes as α decreases. Thus, the backbone reddit interest network has ≈ 30 core communities, and another ≈ 30 weakly linked communities that are lost as a more stringent α cutoff is applied.

Discussion

We have shown that backbone networks can be used to automatically map massive interest networks in social media based solely on user behavior. By viewing the big world of reddit as a hierarchical map, users can now explore related interests without providing any prior information about their own interests. Additionally, these maps provide a *dynamic* view of interests on the social network, owing to the fact that they are constructed from actual user behavior in the social network. Future applications of this method may also facilitate navigation of other popular social network platforms such as Facebook and Twitter.

Furthermore, such an interest map could allow social media users to self-organize into more specific interest forums, thus reducing preferential attachment to large, general interest forums and alleviating the issues that arise in overcrowded social network forums [6]. Given previous work that suggests network properties such as small-worldness and even modularity can result solely from network growth processes [27], it would be interesting in future work to observe what processes govern network growth when users have access to an interests map like those shown in Figures 1 and 2, and what network properties emerge from these growth processes.

Additionally, we explored the network properties of the backbone reddit interest network that we composed from the posting behavior of over 850,000 active reddit users. In this analysis, we found that the reddit interest network has a scale-free, small-world, and modular community structure, corroborating findings in many other online social networks [28,29]. Uniquely, reddit potentially enforces a scale-free network structure on its users by automatically subscribing all new users to the same set of 20 subreddits [24]. Exploring the effect of automatically subscribing users to a fixed set of interest-specific forums on social interest network structure could be another interesting venue of future work. To expedite future analyses of the reddit interest network, we have provided the raw, anonymized data set available to download online [19].

Interestingly, our findings corroborate earlier analyses of the Del.icio.us tag network focusing on collaborative tagging systems [4,5]. We show here that reddit's subreddit degree distribution also follows a power law, as predicted for collaborative tagging systems. Further, whereas [5] suggested that the long tail of infrequently-used tags could likely be ignored, we demonstrated that entire interest meta-communities exist

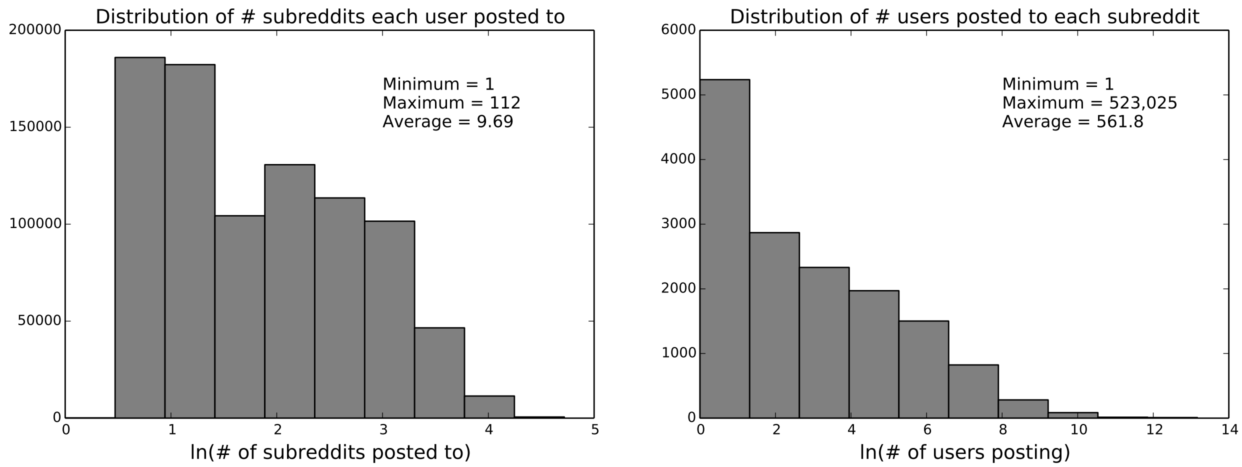

Figure 4: **Edge distribution in the bipartite (user-to-subreddit) network.** Note that the x-axes are log transformed to better display the distribution.

in that long tail that would not otherwise be discovered, including the sports, programming, and LGBT meta-communities. Finally, while [5] was limited to visualizing only subsets of the collaborative tagging network due to excess edges, the mapping method we present here suffers no such limitations due to the backbone network extraction method.

It is important to note that the sample of user behavior we have taken is cross-sectional, reflecting users' reddit posts and thus the relationships among reddit interests at a fixed point in time in mid-2013. However, as users' interests evolve, so too do the relationships among them [30]. In some cases, highly specialized and related subreddits may fuse into a single subreddit, while in other cases a general subreddit may split into multiple more specialized ones. Thus, such an interest map would require periodic (or, ideally, real-time) updating to accurately reflect dominant interests in the social network and their relationships to one another. Further, given that the network meta-communities are likely to change over time, it is not feasible to manually annotate the meta-communities as we did in Figure 1. In future work, it would be beneficial to improve this mapping method by implementing a programmatic annotation algorithm using automatic content analysis of the conversations in the subreddits.

Methods

To acquire the data for this study, we mined user posting behavior data from reddit by first gathering the user names of 876,961 active users that post to 15,122 distinct subreddits (see Figure 4 for more detail). reddit provides an open source API for anyone to freely mine data from the web site [31], and only requests that published compilations of reddit data be anonymized to protect the privacy of its users. We note that this data set represents a complete sample of all active users who posted one or more times on reddit between January 2013 and August 2013. For each of the users, we gathered their 1,000 most recent link submissions and comments, counted how many times they post to each subreddit, and registered them as interested in a subreddit only if they posted there at least 10 times. We applied this threshold of at least 10 posts to filter out users that are not active in a particular subreddit. Due to storage space limitations, the latter data format is the rawest form of data we were able to store long-term for this study.

From these data, we defined a bipartite network $\mathbf{X}$, where $X_{ij} = 1$ if user i is an active poster in subreddit j and otherwise is 0. We then projected this as a weighted unipartite network $\mathbf{Y}$ as $\mathbf{XX}'$, where Y_{ij} is the number of users that post in both subreddits i and j. This resulted in 4,520,054 non-zero, symmetric edges between the subreddits. Details of the raw weighted subreddit network are shown in Figure 5.

Due to the challenges associated with analyzing large weighted networks, we reduced the number of edges in the weighted subreddit network using a backbone extraction algorithm [11] that has previously been used

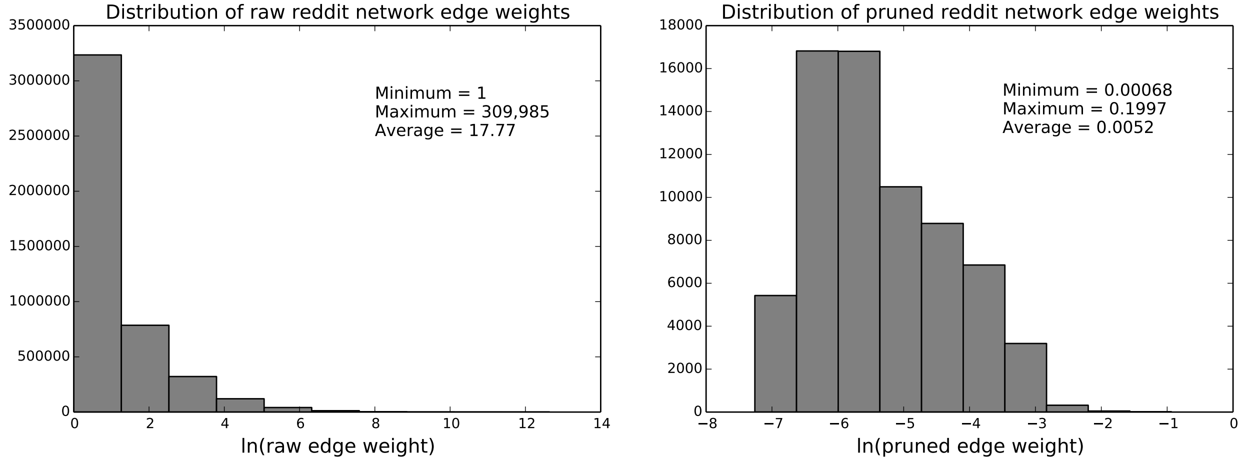

Figure 5: **Edge distribution in the raw and pruned reddit user interest network.** Note that the x-axes are log transformed to better display the distribution.

to reduce bipartite projections [32]. This algorithm begins by replacing symmetric valued edges (S_{ij}) with asymmetric weighted edges (A_{ij} and A_{ji}), where $A_{ij} = S_{ij}/i$'s degree and $A_{ji} = S_{ij}/j$'s degree. It then preserved edges whose weight is statistically incompatible, at a given level of significance α, with a null model in which edge weights are distributed uniformly at random. In our resulting backbone network, two subreddits are linked if the number of users who post in both of them is statistically significantly larger than expected in a null model, from the perspective of *both* subreddits. To recombine the directed edges between each two nodes, we replaced the two directed edges with a single undirected edge whose weight is the average of the two directed edges.

Thus, this technique defines a network of subreddit pathways along which there is a high probability users might traverse if they navigate reddit by following the posts of other users. Adjusting the α parameter allows the backbone network to include more (e.g., when α if larger) or fewer (e.g., when α is smaller) such pathways. Figure 3 summarizes the topological properties of backbones extracted using a range of α parameter values; in the findings and discussion we focus on a backbone extracted using $\alpha = 0.05$. Our choice of $\alpha = 0.05$ is arbitrary, but because the backbone extraction technique we use is rooted in probability theory, it nonetheless offers a precise interpretation: an edge is retained in the backbone if there is less than a 5% chance that an edge with the same weight or greater would appear under a null model in which all edge weights were conditionally (on node degree) random.

We used Python's PRAW package (Python Reddit API Wrapper: `https://github.com/praw-dev/praw`) to gather the data and Python's NetworkX package [33] to compute all network statistics. In the backbone graph, we focus only on the largest connected component. We detected network communities using [12], which aims to partition the nodes into mutually exclusive sets that maximize the graph's modularity. Other community detection algorithms exist and may yield slightly different partitionings, but all aim to achieve the same goal in principle [34]. We visualize the backbone network and detected communities using the OpenOrd node layout. Both the community detection and node layout routines are implemented in Gephi [23]. Meta-communities identified by the community detection algorithm, e.g. "sports" and "programming," were manually annotated using domain knowledge to identify a proper annotation.

Acknowledgments

We gratefully acknowledge the support of the Michigan State University High Performance Computing Center and the Institute for Cyber Enabled Research (iCER). We thank Arend Hintze, Christoph Adami, and Emily Weigel for helpful feedback during the preparation of this manuscript.

References

[1] Rainie, L. & Wellman, B. Networked: The new social operating system. (The MIT Press, Cambridge, MA, USA, 2012)

[2] Strand, J.L. Facebook: Trademarks, fan pages, and community pages. *Int Prop & Tech Law J* **23**: 10–13 (2011).

[3] Chang, H.C. A new perspective on Twitter hashtag use: diffusion of innovation theory. *P Am Soc Inform Sci* **47**: 1–4 (2010).

[4] Golder, S.A. & Huberman, B.A. Usage patterns of collaborative tagging systems. *J Inf Sci* **32**: 198–208 (2006).

[5] Halpin, H., Robu, V. & Shepherd, H. The complex dynamics of collaborative tagging. In: *Proc WWW 2007*. New York, NY, USA: ACM, WWW '07, pp. 211–220 (2007).

[6] Gilbert, E. Widespread underprovision on reddit. In: *Proc CSCW 2013*. New York, NY, USA: ACM, CSCW '13, pp. 803–808 (2013). doi:10.1145/2441776.2441866.

[7] Boguna, M., Krioukov, D. & Claffy, K.C. Navigability of complex networks. *Nat Phys* **5**: 74–80 (2008).

[8] Benevenuto, F., Rodrigues, T., Cha, M. & Almeida, V. Characterizing user navigation and interactions in online social networks. *Inform Sciences* **195**: 1–24 (2012).

[9] Albert, R., Jeong, H. & Barabási. A.L. Internet: Diameter of the world-wide web. *Nature* **401**: 130–131. (1999)

[10] Barabási, A.L., Albert, R. & Jeong, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A* **281**: 69–77 (2000).

[11] Serrano, M., Boguñá, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *P Natl Acad Sci USA* **106**: 6483-6488 (2009).

[12] Blondel, V.D., Guillaume, J.L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J Stat Mech-Theory E*: P10008 (2008).

[13] n.p. What is reddit? (2014, June 5). Retrieved October 6, 2014, from `http://www.reddit.com/wiki/faq#wiki_what_is_reddit.3F`

[14] n.p. reddit Alexa ranking. (2014, October 6). Retrieved October 6, 2014, from `http://www.alexa.com/siteinfo/reddit.com`

[15] Olson, R.S. redditviz, the interactive reddit interest map. (2013). Retrieved October 6, 2014, from `http://rhiever.github.io/redditviz/clustered/`

[16] Sanderson, B. & Rigby, M. We've reddit, have you? What librarians can learn from a site full of memes. *Coll Res Libr* **74**: 518–521 (2013).

[17] Wasike, B.S. Framing social news sites: An analysis of the top ranked stories on reddit and Digg. *SW Mass Comm J* **27** (2011).

[18] Merritt, E. An Analysis of the Discourse of Internet Trolling: A Case Study of Reddit.com. Ph.D. thesis (2012).

[19] Olson, R.S. reddit user posting behavior (mid-2013). (2013). Retrieved October 6, 2014, from `http://dx.doi.org/10.6084/m9.figshare.874101`

[20] Kumar, R., Novak, J. & Tomkins, A. Structure and evolution of online social networks. In: *Link Mining: Models, Algorithms, and Applications*, Springer. pp. 337–357 (2010).

[21] Page, L., Brin, S., Motwani, R. & Winograd, T. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab (1999).

[22] Grolmusz, V. A note on the pagerank of undirected graphs. arXiv e-print. http://arxiv.org/abs/1205.1960. (2012)

[23] Bastian, M., Heymann, S. & Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In: *Proc ICWSM 2009*. San Jose, CA, USA: The AAAI Press, ICWSM '09, pp. 361–362 (2009).

[24] n.p. Saying goodbye to an old friend and revising the default subreddits. (2011). Retrieved October 6, 2014, from `http://blog.reddit.com/2011/10/saying-goodbye-to-old-friend-and.html`

[25] Barabási, A.L. & Albert, R. Emergence of scaling in random networks. *Science* **286**: 509–512 (1999).

[26] Humphries, M.D. & Gurney, K. Network "small-world-ness": A quantitative method for determining canonical network equivalence. *PLoS ONE* **3**: e0002051 (2008).

[27] Hintze, A. & Adami, C. Modularity and anti-modularity in networks with arbitrary degree distribution. *Biol Direct* **5**: 32+ (2010).

[28] Ahn, Y.Y., Han, S., Kwak, H., Moon, S. & Jeong, H. Analysis of topological characteristics of huge online social networking services. In: *Proc WWW 2007*. New York, NY, USA: ACM, WWW '07, pp. 835–844 (2007).

[29] Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P. & Bhattacharjee, B. Measurement and analysis of online social networks. In: *Proc IMC 2007*. New York, NY, USA: ACM, IMC '07, pp. 29–42 (2007).

[30] Banerjee, N., Chakraborty, D., Dasgupta, K., Mittal, S., Joshi, A., et al. User interests in social media sites: An exploration with micro-blogs. In: *Proc CIKM 2009*. New York, NY, USA: ACM, CIKM '09, pp. 1823–1826 (2009).

[31] n.p. reddit API documentation. (2014). Retrieved October 6, 2014, from `http://www.reddit.com/dev/api`

[32] Ahn, Y.Y., Ahnert, S.E., Bagrow, J.P. & Barabási, A.L. Flavor network and the principles of food pairing. *Sci Rep* **1** (2011).

[33] Hagberg, A.A., Schult, D.A., Swart, P.J. Exploring network structure, dynamics, and function using NetworkX. In: *Proc SciPy 2008*. Pasadena, CA USA, pp. 11–15 (2008).

[34] Fortunato, S. Community detection in graphs. *Phys Rep* **486**: 75–174 (2010).