

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

Randal S. Olson<sup>1</sup> and Zachary P. Neal<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA

<sup>2</sup> Department of Psychology, Michigan State University, East Lansing, MI, USA

## 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.

**Subjects** Network Science and Online Social Networks, Visual Analytics

**Keywords** Social networks, reddit, Mapping, Network backbone extraction, Social network navigation

## INTRODUCTION

In the past decade, social media platforms have grown from a pastime for teenagers into tools that pervade nearly all modern adults' lives ([Rainie & Wellman, 2012](#)). 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" ([Strand, 2011](#)), Twitter users classify their tweets using topic-specific "hashtags" ([Chang, 2010](#)), Del.ici.ous users bookmark links with topic-specific tags ([Golder & Huberman, 2006](#); [Halpin, Robu & Shepherd, 2007](#)), and reddit users post and subscribe to topic-specific sub-forums called "subreddits" ([Gilbert, 2013](#)).

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

Submitted 14 February 2015  
Accepted 11 May 2015  
Published 27 May 2015

Corresponding author

Randal S. Olson,  
[rso@randalolson.com](mailto:rso@randalolson.com)

Academic editor  
Massimiliano Zanin

Additional Information and  
Declarations can be found on  
page 10

DOI [10.7717/peerj-cs.4](https://doi.org/10.7717/peerj-cs.4)

© Copyright  
2015 Olson and Neal

Distributed under  
Creative Commons CC-BY 4.0

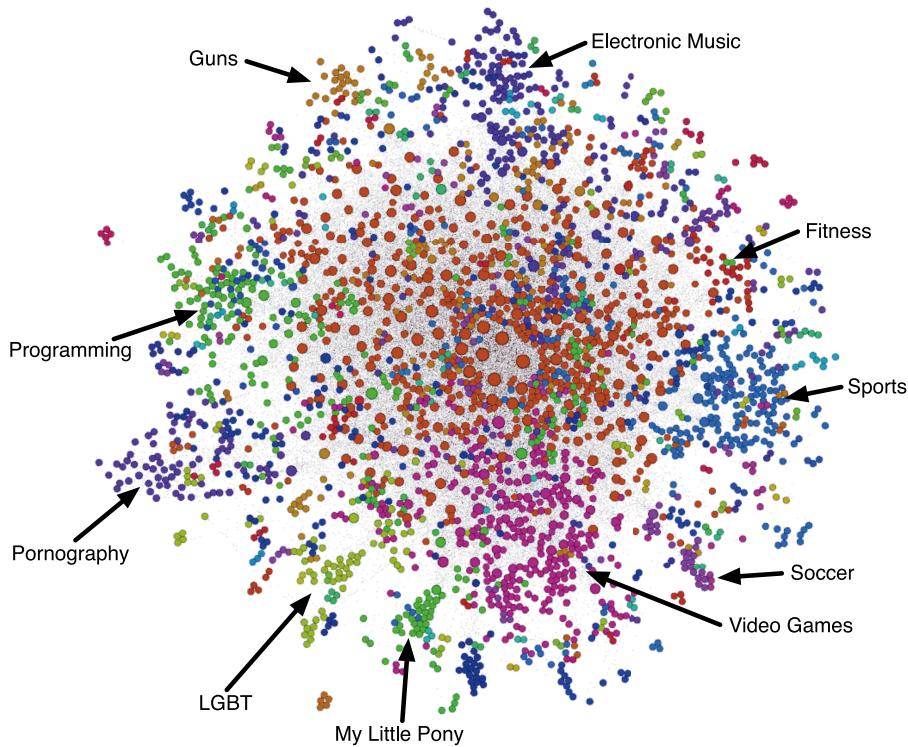
OPEN ACCESS

challenging because these worlds do not come with maps ([Boguna, Krioukov & Claffy, 2008](#); [Benevenuto et al., 2012](#)). 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 ([Albert, Jeong & Barabási, 1999](#); [Barabási, Albert & Jeong, 2000](#)). 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 ([Serrano, Boguñá & Vespignani, 2009](#)) and community detection ([Blondel et al., 2008](#)) 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 ([Anonymous, 2015a](#)), one of the most visited social media platforms on the web ([Anonymous, 2015b](#)), and produce an interactive map of all of the subreddits. An interactive version of the reddit interest map is available online ([Olson, 2013a](#)).

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. We hypothesize that the scale-free property is the result of a preferential attachment process in which new reddit users, or existing users who are exploring, are most likely to encounter and become active in the largest and most active sub-reddits. However, because other processes may also account for the scale-free property, this hypothesis requires further investigation. 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 ([Gilbert, 2013](#)).

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 ([Sanderson & Rigby, 2013](#)), it has been relatively understudied ([Merritt, 2012](#); [Gilbert, 2013](#); [Buntain & Golbeck, 2014](#); [Leavitt & Clark, 2014](#)). This data set can be downloaded online at [Olson \(2013b\)](#).



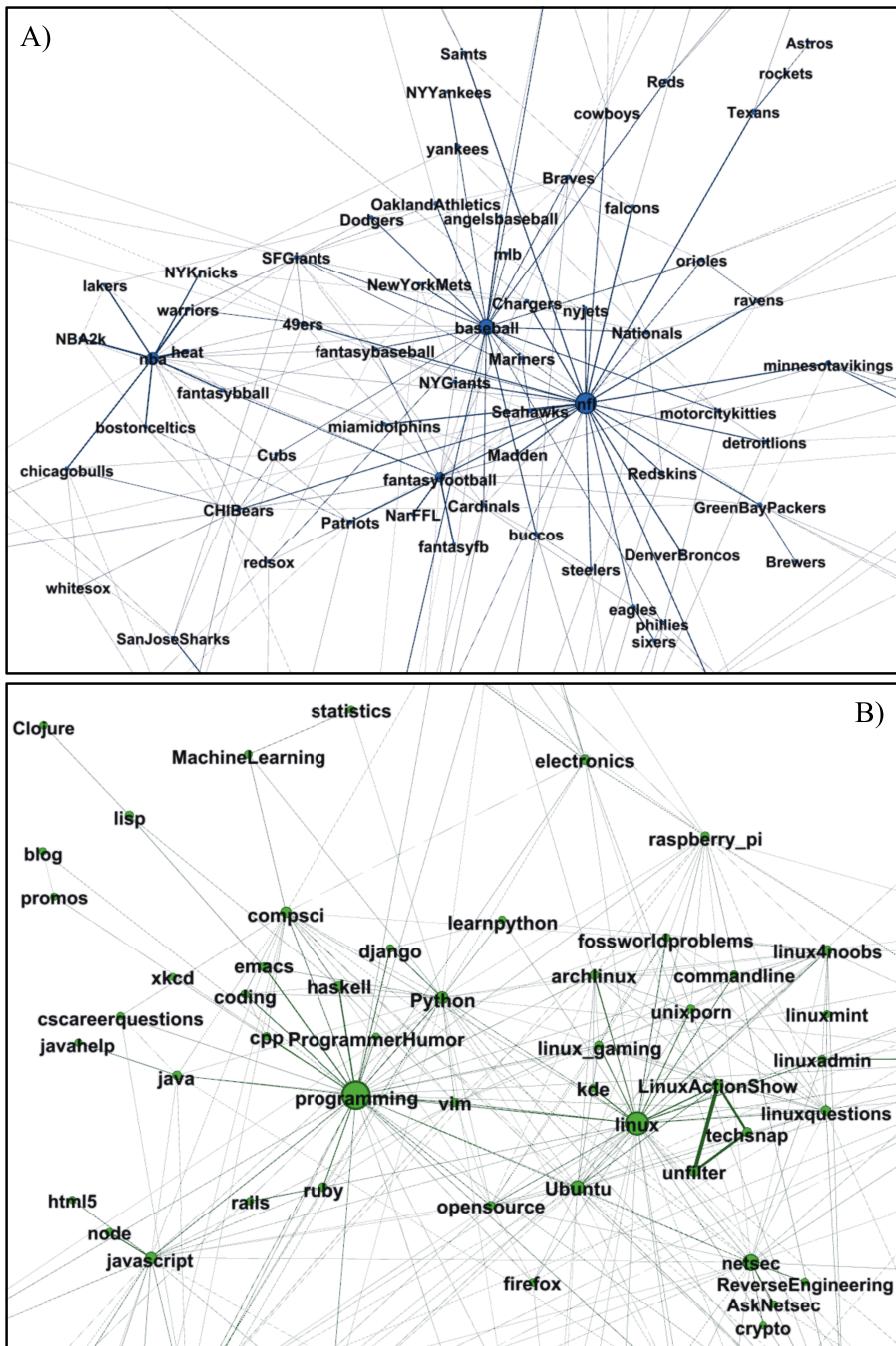
**Figure 1** Reddit interest network. The largest components of the reddit interest network is shown with 10 interest meta-communities annotated. Interactive version: <http://rheiver.github.io/redditzv/clustered/>.

## RESULTS

### Reddit interest map

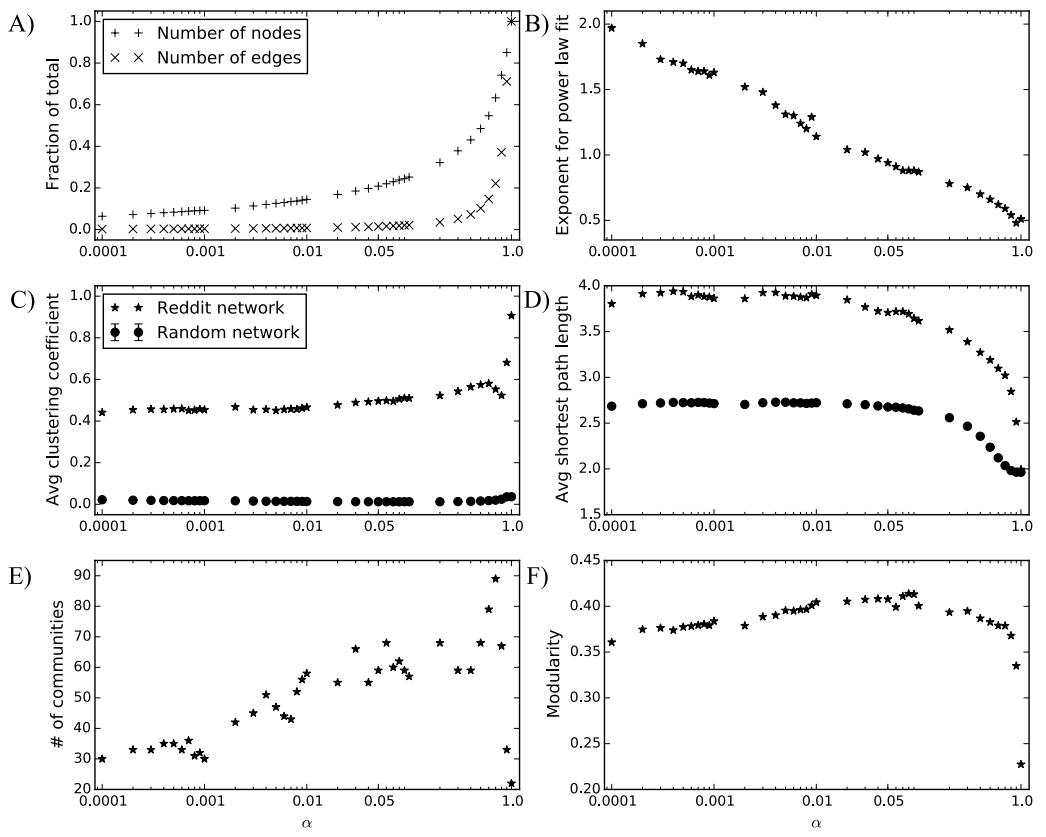
For the final version of the reddit interest map, we use the backbone network produced with  $\alpha = 0.05$  (see Methods). This results in a network with 59 distinct clusters of subreddits, which we call *interest meta-communities*. In Fig. 1, the nodes (i.e., subreddits) are sized by their weighted PageRank (Page et al., 1999; Grolmusz, 2012) to provide an indication of how likely a node is to be visited, and positioned according to the OpenOrd layout in Gephi (Bastian, Heymann & Jacomy, 2009) to place related nodes together.

Through this method, we immediately see several distinct interest meta-communities, 10 of which are annotated in Fig. 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 Fig. 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** Pictured are several topic-specific subreddits composing a meta-community around a broad topic such as sports (A) or programming (B).

[Figure 2](#) depicts zoomed-in views of two interest meta-communities annotated in [Fig. 1](#). In [Fig. 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



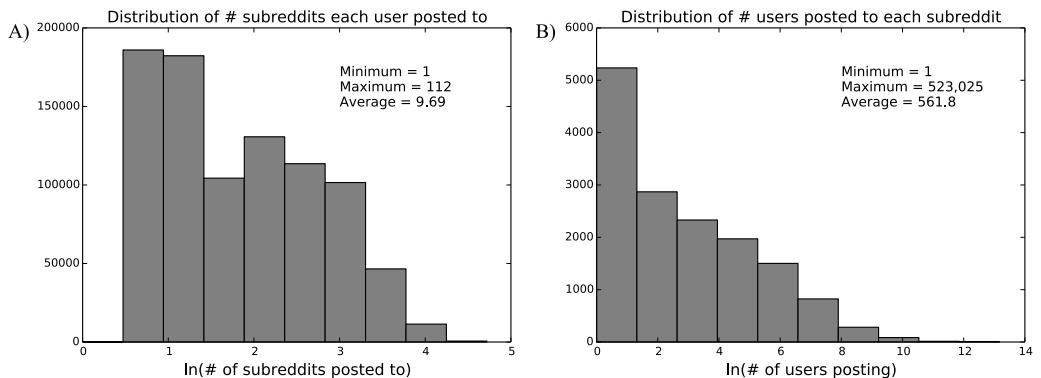
**Figure 3** Sensitivity analysis of the largest connected component of the reddit interest network over a range of alpha cutoff values.

users discuss the latest NFL news and games. Similarly in Fig. 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 Fig. 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 Fig. 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  $\alpha$  cutoff



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

values for the backbone reddit interest network (see Methods) to demonstrate that the interest network we chose in Fig. 1 is robust to relevant  $\alpha$  cutoff values. Since lower  $\alpha$  cutoff values result in many nodes becoming disconnected from the network, we only use the largest connected component of the network for all network statistics reported. As expected, the majority of the edges are pruned by an  $\alpha$  cutoff of 0.05 (Fig. 3A). This result demonstrates that the backbone interest network is stable with an  $\alpha$  cutoff  $\leq 0.05$ , which is the most relevant range of  $\alpha$  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 (Fig. 4). 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 can result 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 (Albert, Jeong & Barabási, 1999; Barabási & Albert, 1999; Barabási, Albert & Jeong, 2000; Albert & Barabási, 2002). Regardless of the  $\alpha$  cutoff, we observed that the node degree distribution of all backbone reddit interest networks fit a power law ( $R^2 \approx 0.91$  for  $k \geq 50$ ; Fig. 3B). This scale-free network structure is likely partially due to reddit's default subreddit system (Anonymous, 2011), where newly registered users are subscribed to a set of subreddits by default. As expected, as  $\alpha$  approaches 1.0, the power law exponent approaches 0.0. This occurs because as the  $\alpha$  cutoff used for extracting the backbone network approaches 1.0, 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

([Watts & Strogatz, 1998](#)). 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 ( $L_{sw}$ ) that scales logarithmically with the number of nodes ( $N$ ):

$$L_{sw} \approx \log_{10}(N). \quad (1)$$

[Figures 3C](#) and [3D](#) depict 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 ( $\alpha$  cutoff = 0.05) follows [Eq. \(1\)](#), with  $L_{sw} = \log_{10}(2,347) = 3.37 \approx 3.71$  from [Fig. 3D](#). 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 ( $S_G$ ) proposed in [Humphries & Gurney \(2008\)](#):

$$S_G = \frac{C_G/C_{\text{rand}}}{L_G/L_{\text{rand}}} \quad (2)$$

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  $S_G > 1$ , then the network is classified as a small-world network. For the backbone reddit network, we calculated  $S_G = 14.2$  ( $P < 0.001$ ), 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 [Fig. 3F](#), the backbone network exhibits a consistently high modularity score with an  $\alpha$  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 [Fig. 3E](#), the number of identified communities (i.e., clusters) remains relatively low until the  $\alpha$  cutoff is reduced to  $\leq 0.9$ . As the  $\alpha$  cutoff is reduced, the number of identified communities generally decreases, which coincides with the loss of nodes as  $\alpha$  decreases. Thus, the backbone reddit interest network has  $\approx 30$  core communities, and another  $\approx 30$  weakly linked communities that are lost as a more stringent  $\alpha$  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 could provide a *dynamic*

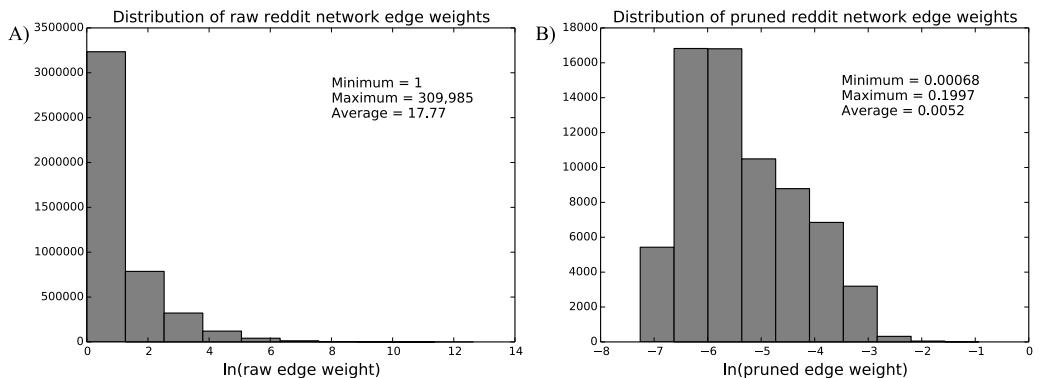
view of interests on the social network, owing to the fact that they are constructed from actual user behavior in the social network. Future applications of this method may also facilitate navigation of other popular social network platforms such as Facebook and Twitter.

Furthermore, such an interest map could allow social media users to self-organize into more specific interest forums, thus reducing potential preferential attachment to large, general interest forums and alleviating the issues that arise in overcrowded social network forums ([Gilbert, 2013](#)). Given previous work that suggests network properties such as small-worldness and even modularity can result solely from network growth processes ([Hintze & Adami, 2010](#)), 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 [Figs. 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 ([Ahn et al., 2007](#); [Mislove et al., 2007](#)). Uniquely, reddit potentially enforces a scale-free network structure on its users by automatically subscribing all new users to the same set of subreddits ([Anonymous, 2011](#)). 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 ([Olson, 2013b](#)).

Interestingly, our findings corroborate earlier analyses of the Del.icio.us tag network focusing on collaborative tagging systems ([Golder & Huberman, 2006](#); [Halpin, Robu & Shepherd, 2007](#)). We show here that reddit's subreddit degree distribution also follows a power law, as predicted for collaborative tagging systems. Further, whereas [Halpin, Robu & Shepherd \(2007\)](#) suggested that the long tail of infrequently-used tags could likely be ignored, we demonstrated that entire interest meta-communities exist in that long tail that would not otherwise be discovered, including the sports, programming, and LGBT meta-communities. Finally, while [Halpin, Robu & Shepherd \(2007\)](#) 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 ([Banerjee et al., 2009](#)). 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



**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.

as we did in Fig. 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 Fig. 4 for more detail). reddit provides an open source API for anyone to freely mine data from the web site ([Anonymous, 2015c](#)), 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 Fig. 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 ([Serrano, Boguñá & Vespignani, 2009](#)) that has previously been used to reduce bipartite projections ([Ahn et al., 2011](#)). 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  $\alpha$ , 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 binary 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  $\alpha$  parameter allows the backbone network to include more (e.g., when  $\alpha$  is larger) or fewer (e.g., when  $\alpha$  is smaller) such pathways. Figure 3 summarizes the topological properties of backbones extracted using a range of  $\alpha$  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. Future work on extracting the reddit backbone may benefit from exploring the use of more computationally complex algorithms designed specifically for bipartite projections, including the fixed-degree (Zweig & Kaufmann, 2011) and stochastic-degree (Neal, 2014) sequence models.

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 (Hagberg, Schult & Swart, 2008) to compute all network statistics except for the power law exponent, which was computed using the PARETOFIT module in Stata (Jenkins & Van Kerm, 2015). In the backbone graph, we focus only on the largest connected component. We detected network communities using Blondel *et al.* (2008), 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 (Fortunato, 2010). 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 (Bastian, Heymann & Jacomy, 2009). Meta-communities identified by the community detection algorithm, e.g., “sports” and “programming,” were manually annotated using domain knowledge to identify a proper annotation.

## ACKNOWLEDGEMENTS

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.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

The authors declare there was no funding for this work.

## Competing Interests

The authors declare there are no competing interests.

## Author Contributions

- Randal S. Olson conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, performed the computation work, reviewed drafts of the paper.
- Zachary P. Neal conceived and designed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

## Data Deposition

The following information was supplied regarding the deposition of related data:

FigShare: [http://figshare.com/articles/reddit\\_user\\_posting\\_behavior/874101](http://figshare.com/articles/reddit_user_posting_behavior/874101).

## REFERENCES

- Ahn YY, Ahnert SE, Bagrow JP, Barabási AL. 2011. Flavor network and the principles of food pairing. *Scientific Reports* 1 DOI 10.1038/srep00196.
- Ahn YY, Han S, Kwak H, Moon S, Jeong H. 2007. Analysis of topological characteristics of huge online social networking services. In: *Proceedings of the 16th international World Wide Web conference (WWW2007)*. New York: ACM, 835–844. Available at <http://www2007.org/papers/paper676.pdf>.
- Albert R, Barabási AL. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74:47–97 DOI 10.1103/RevModPhys.74.47.
- Albert R, Jeong H, Barabási AL. 1999. Internet: diameter of the world-wide web. *Nature* 401:130–131 DOI 10.1038/43601.
- Anonymous. 2011. Saying goodbye to an old friend and revising the default subreddits. Available at <http://blog.reddit.com/2011/10/saying-goodbye-to-old-friend-and.html> (accessed 4 May 2015).
- Anonymous. 2015a. What is reddit? Available at [http://www.reddit.com/wiki/faq#wiki\\_what\\_is\\_reddit.3F](http://www.reddit.com/wiki/faq#wiki_what_is_reddit.3F) (accessed 4 May 2015).
- Anonymous. 2015b. reddit Alexa ranking. Available at <http://www.alexa.com/siteinfo/reddit.com> (accessed 4 May 2015).
- Anonymous. 2015c. reddit API documentation. Available at <http://www.reddit.com/dev/api> (accessed 4 May 2015).
- Banerjee N, Chakraborty D, Dasgupta K, Mittal S, Joshi A, Nagar S, Rai A, Madan S. 2009. User interests in social media sites: an exploration with micro-blogs. In: *Proceedings of the 18th ACM conference on information and knowledge management*. New York: ACM, 1823–1826 DOI 10.1145/1645953.1646240.
- Barabási AL, Albert R. 1999. Emergence of scaling in random networks. *Science* 286:509–512 DOI 10.1126/science.286.5439.509.
- Barabási AL, Albert R, Jeong H. 2000. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A* 281:69–77 DOI 10.1016/S0378-4371(00)00018-2.
- Bastian M, Heymann S, Jacomy M. 2009. Gephi: an open source software for exploring and manipulating networks. In: *Proc ICWSM 2009, ICWSM'09*. San Jose: The AAAI Press, 361–362.

- Benevenuto F, Rodrigues T, Cha M, Almeida V.** 2012. Characterizing user navigation and interactions in online social networks. *Information Sciences* **195**:1–24 DOI [10.1016/j.ins.2011.12.009](https://doi.org/10.1016/j.ins.2011.12.009).
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E.** 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **P10008**.
- Boguna M, Krioukov D, Claffy KC.** 2008. Navigability of complex networks. *Nature Physics* **5**:74–80 DOI [10.1038/nphys1130](https://doi.org/10.1038/nphys1130).
- Buntain C, Golbeck J.** 2014. Identifying social roles in reddit using network structure. In: *Proceedings of the 23rd international world wide web conference*, 615–620.
- Chang HC.** 2010. A new perspective on Twitter hashtag use: diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology* **47**:1–4.
- Fortunato S.** 2010. Community detection in graphs. *Physics Reports* **486**:75–174 DOI [10.1016/j.physrep.2009.11.002](https://doi.org/10.1016/j.physrep.2009.11.002).
- Gilbert E.** 2013. Widespread underprovision on reddit. In: *Proc CSCW 2013, CSCW'13*. New York: ACM, 803–808 DOI [10.1145/2441776.2441866](https://doi.org/10.1145/2441776.2441866).
- Golder SA, Huberman BA.** 2006. Usage patterns of collaborative tagging systems. *Journal of Information Science* **32**:198–208 DOI [10.1177/0165551506062337](https://doi.org/10.1177/0165551506062337).
- Grolmusz V.** 2012. A note on the pagerank of undirected graphs. ArXiv preprint. [arXiv:1205.1960](https://arxiv.org/abs/1205.1960).
- Hagberg AA, Schult DA, Swart PJ.** 2008. Exploring network structure, dynamics, and function using NetworkX. In: *Proc SciPy 2008*. Pasadena, CA, USA, 11–15.
- Halpin H, Robu V, Shepherd H.** 2007. The complex dynamics of collaborative tagging. In: *Proc WWW 2007, WWW'07*. New York: ACM, 211–220.
- Hintze A, Adami C.** 2010. Modularity and anti-modularity in networks with arbitrary degree distribution. *Biology Direct* **5**:32 DOI [10.1186/1745-6150-5-32](https://doi.org/10.1186/1745-6150-5-32).
- Humphries MD, Gurney K.** 2008. Network “small-world-ness”: a quantitative method for determining canonical network equivalence. *PLoS ONE* **3**:e0002051 DOI [10.1371/journal.pone.0002051](https://doi.org/10.1371/journal.pone.0002051).
- Jenkins S, Van Kerm P.** 2015. PARETOFIT: stata module to fit a Type 1 Pareto distribution. Available at <http://EconPapers.repec.org/RePEc:boc:bocode:s456832> (accessed 4 May 2015).
- Leavitt A, Clark JA.** 2014. Upvoting Hurricane Sandy: event-based news production processes on a social news site. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, 1495–1504.
- Merritt E.** 2012. An analysis of the discourse of internet trolling: a case study of reddit.com. Ph.D. thesis.
- Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B.** 2007. Measurement and analysis of online social networks. In: *Proc IMC 2007, IMC'07*. New York: ACM, 29–42.
- Neal Z.** 2014. The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors. *Social Networks* **39**:84–97 DOI [10.1016/j.socnet.2014.06.001](https://doi.org/10.1016/j.socnet.2014.06.001).
- Olson RS.** 2013a. redditviz, the interactive reddit interest map. Available at <http://rhiiever.github.io/redditviz/clustered/> (accessed 4 May 2015).
- Olson RS.** 2013b. reddit user posting behavior (mid-2013). Available at <http://dx.doi.org/10.6084/m9.figshare.874101> (accessed 4 May 2015).
- Page L, Brin S, Motwani R, Winograd T.** 1999. The pagerank citation ranking: bringing order to the web. Technical Report 1999-66. Stanford InfoLab.
- Rainie L, Wellman B.** 2012. *Networked: the new social operating system*. Cambridge: The MIT Press.

- Sanderson B, Rigby M.** 2013. We've reddit, have you? What librarians can learn from a site full of memes. *College & Research Libraries News* **74**:518–521.
- Serrano M, Boguñá M, Vespignani A.** 2009. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* **106**:6483–6488 DOI [10.1073/pnas.0808904106](https://doi.org/10.1073/pnas.0808904106).
- Strand JL.** 2011. Facebook: trademarks, fan pages, and community pages. *Intellectual Property & Technology Law Journal* **23**:10–13.
- Watts DJ, Strogatz SH.** 1998. Collective dynamics of small-world networks. *Nature* **393**:440–442 DOI [10.1038/30918](https://doi.org/10.1038/30918).
- Zweig KA, Kaufmann M A.** 2011. systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining* **1**:187–218 DOI [10.1007/s13278-011-0021-0](https://doi.org/10.1007/s13278-011-0021-0).