Identifying Misaligned Inter-Group Links and Communities

SRAYAN DATTA, University of Michigan
CHANDA PHELAN, University of Michigan
EYTAN ADAR, University of Michigan

Many social media systems *explicitly* connect individuals (e.g., Facebook or Twitter); as a result, they are the targets of most research on social *networks*. However, many systems do not emphasize or support explicit linking between people (e.g., Wikipedia or Reddit), and even fewer explicitly link communities. Instead, network analysis is performed through inference on *implicit* connections, such as co-authorship or text similarity. Depending on how inference is done and what data drove it, different networks may emerge. While correlated structures often indicate stability, in this work we demonstrate that differences, or *misalignment*, between inferred networks also capture interesting behavioral patterns. For example, high-text but low-author similarity often reveals communities "at war" with each other over an issue or high-author but low-text similarity can suggest community fragmentation. Because we are able to model edge direction, we also find that *asymmetry* in degree (in-versus-out) co-occurs with marginalized identities (subreddits related to women, people of color, LGBTQ, etc.). In this work, we provide algorithms that can identify misaligned links, network structures and communities. We then apply these techniques to Reddit to demonstrate how these algorithms can be used to decipher inter-group dynamics in social media.

CCS Concepts: • **Human-centered computing** → **Social media**; *Social networks*; *Social network analysis*;

Additional Key Words and Phrases: Inter-group similarity; Reddit; social network analysis; misaligned links

ACM Reference format:
Srayan Datta, Chanda Phelan, and Eytan Adar. 2017. Identifying Misaligned Inter-Group Links and Communities. *Proc. ACM Hum.-Comput. Interact.* 1, 2, Article 37 (November 2017), 23 pages.
https://doi.org/10.1145/3134672

37

1 INTRODUCTION

Network modeling of online social systems is a common approach for the study of social behavior. Where explicit, the links between individuals measure friendship, shared interests, or other relationships (family, followers, fans, etc.). When the online system does not have features that explicitly support or encourage linking, we rely on inferred connections. For example, we may infer that two people are "linked" if they post on the same discussion forum or that two communities are linked if they are similar based on text. Inference is *sometimes* necessary in the case of *person-to-person* links and *often* in the case of *community-to-community* links, where explicit links are rare. For example, subreddits (communities on Reddit) tend not to make explicit connections between each other. Yet, they are connected in many ways. Pairs of subreddits may share topics, share authors, share moderators, link to similar content in web, and so on. While indirect [62], similarities based on

Authors' contact information – Srayan Datta: srayand@umich.edu; Chanda Phelan: cdphelan@umich.edu; Eytan Adar: eadar@umich.edu. University of Michigan, 105 South State Street, Ann Arbor, MI 48109.

2573-0142/2017/11-ART37 $15.00
https://doi.org/10.1145/3134672

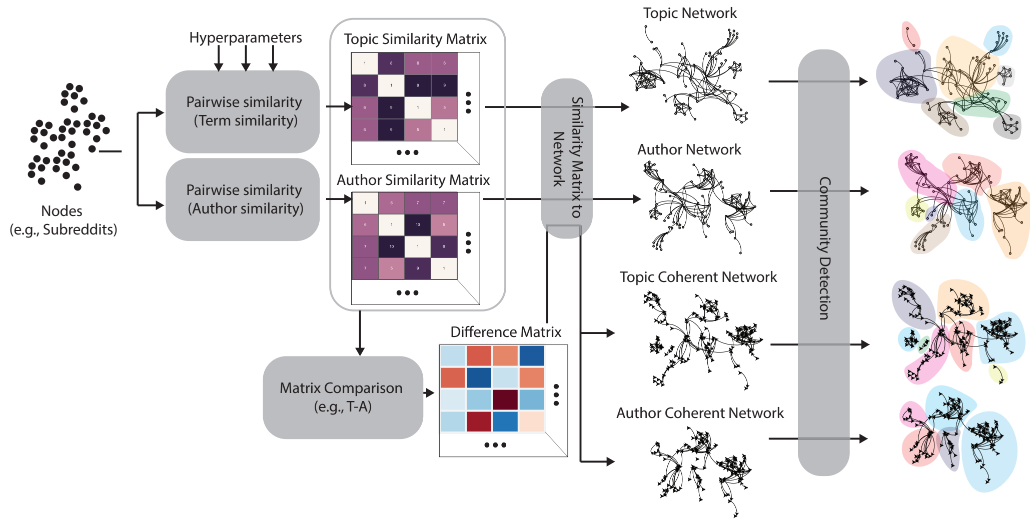

Fig. 1. Network inference pipeline. Topic and author networks refer to topic and author similarity networks respectively.

these features correlate with—and predict—connections. These connections reflect various social processes and can help model both the current state of the social system and the process by which the relationships emerged.

Choices about which similarity measure(s) and inference algorithm to use (not to mention the hyperparameters of the algorithm, such as normalization and thresholding) must be made carefully, as these choices will influence which links are predicted and how they are to be interpreted. The top of Figure 1 depicts a conventional analysis pipeline: similarity measures are applied to a disconnected network to generate a pairwise similarity matrix, and then an inference algorithm determines which values should be considered links and produces a network. On these inferred networks, downstream analysis such as community detection can be performed (e.g., clustering subreddits into larger communities).

In many situations, different similarity measures are likely to be highly correlated. High author similarity between communities often means topic similarity will also be high. Conversely, low author similarity means we should expect topic similarity to be low, as well. When we see this agreement, it often signals a "good link." Here we treat evidence as additive: if both text and authorship agree, the subreddits should be connected. Many inference algorithms rely on variants of this similarity comparison to infer connections. However, as we demonstrate in the context of Reddit, such correlation can be weak and the many edges that violate this expectation result from behavior and design and may lead to very different outputs.

In this work, we argue that *disagreements* between inferences are often as informative as agreements. We define a measure to compare inferred similarity matrices that identifies "misaligned" links (Figure 1, bottom). For example, two subreddits may share many authors but discuss entirely different topics; we call these types of links *author-coherent links*. When two subreddits have high text similarity but low author overlap, we call these *topic-coherent links*. To account for the influence of unequal and diverse levels in popularity of different subreddits we develop a score (double-z score, or z^2-score). This score can be used to create directed networks that capture "misaligned" links both locally (in the context of a specific subreddit) and globally.

Our measure of misalignment acts to operationalize, more generally, various structures of interest for social media researchers. For example, social media researchers have targeted phenomena such as "communities at war" [2, 35, 36, 39], community fragmentation (i.e. multiple linked sub-communities instead of a single large community) [19, 25, 64], isolated or niche-interest community links [17], and "strongly linked" communities [25]. Many of these studies have required domain knowledge that is hard to generalize or automate, especially when using common link inference methods. That is, it is difficult to find *multiple* community pairs/groups that have a certain structure (e.g., "communities at war") or to score or rank these found structures for further analysis. Reasons for this include: (1) a single inference algorithm (e.g., text *or* author) does not provide enough "signal" to capture these relationships, (2) algorithms that use multiple inferences (e.g., text *and* author) make naive assumptions about the agreement–or alignment–between the inferred networks, and (3) many algorithms suffer from the presence of a few highly popular communities which tend to be present in a majority of detected links and hide "unexpected" connections. Instead, we demonstrate that *misalignment* between inferred networks can be more generally measured and normalized, and that this measure can be used to find phenomena of interest.

Concretely, we are able to find repeated patterns in these misaligned links. For example, high topic coherence may unearth subreddits that are "at war" with each other (e.g., those with opposing political viewpoints) or have hierarchical relationships (e.g., a niche video game may have a separate community from a more generic gamer subreddit). We also find that subreddits with different ratios of incoming and outgoing links are often out of the mainstream or marginalized. By comparing networks derived through standard similarity measures (e.g., author and text) to our z^2-derived measure, we are able to characterize different types of subreddit-to-subreddit relationships.

Our contributions in this paper are twofold. First, we demonstrate a methodology for comparing two inference workflows to identify misaligned links and communities. We introduce a score to compare networks derived from different similarity metrics that can be used to detect properties that are missed when considering only single inference techniques, or those that are additive. Second, we apply these techniques to Reddit to identify subreddit-to-subreddit relationships. We identify key structures (i.e. topic-coherent, author-coherent, and satellite structures). Our analysis classifies how pairs of subreddits interact, how specific subreddits are situated in the broader context of the Reddit ecosystem, and proposes mechanisms by which networks and higher level communities are formed.

2 RELATED WORK

2.1 Reddit

Previous work on Reddit is diverse but in large part has focused on a single subreddit or a small, manually-selected set of subreddits, often as case studies to analyze behavior in a specific context. For example, Kiene et al. examined how the subreddit *nosleep* dealt with a sudden increase in readership [28]. Potts and Harrison studied *FindBostonBombers*, a botched attempt to crowdsource finding the Boston Marathon Bombers [52]. Leavitt et al. used a very different news event—Hurricane Sandy, a natural disaster—to study how news content was produced and curated in real time [32], and to examine how Reddit's user interface affected the production and curation process [33]. Studies have also examined the effects of Reddit's interface design. For example, Gilbert found that social loafing damaged the site's ability to highlight quality content [22]. Others have examined the role of bots [38], throwaway accounts [31] in Reddit's design, and moderator disruptions in calls for policy change [9, 42].

Reddit is also a popular medium for analyzing language on a particular topic—e.g. smoking cessation [61] or mental health [5, 6, 16, 17, 27]—or studying specific types of user interaction, such as social feedback in weight loss communities [15], seeking support for sexual abuse [4], strategies for persuasive arguments [59, 63] or dogmatism in user comments [18]. Reddit data has also been used to train a model that identifies abusive comments [10] and understand users' moral values using word choice [11].

Less research has focused on the structure of Reddit's network itself [49]. Facing the opaqueness of Reddit's structure, which has little explicit structure beyond subreddits, researchers have attempted to classify subreddits using a variety of methods and metrics. Zhang et al. characterized user behavior *within* subreddits by using comment text to map subreddit topics onto four quadrants: generic-consistent, generic-dynamic, distinctive-consistent, and distinctive-dynamic [65]. Relevant to our study, Hamilton et al. characterized a small number of manually collected subreddits according to the loyalty of their users, finding differences in how much time end-users devote exclusively to a particular subreddit [23]. For example, they found that sports subreddits tended to have loyal users while default subreddits did not. While behaviors within subreddits have clear implications to inter-subreddit behavior, these studies did not extend to analyze linking.

Targeted studies have tried to identify the relationships between subreddits. Hessel et al. focused on highly related communities, identified according to their affixes (e.g. *atheism* and *trueatheism*, or *food* and *foodhacks*) [25]. These pairings often indicated a splintering, either as a result of conflict between users or to afford more specialized discussion. However, these instances only represent a small portion of the Reddit network. Reddit's default subreddits and openness to cross-posting presents an additional challenge, as subreddit networks based on cross-posting are quite dense and require additional filtering. Olson and Neal [49] used author similarity to create a network, then used a backbone extraction algorithm [57] to prune the least important connections. Their analysis of a 2013 dataset found 59 communities with a small-world, scale-free network structure. This power-law distribution was partly attributed to Reddit's UX design, in which new users are subscribed to default subreddits [49].

More recently, Martin [41] also made use of author similarity, in this case for applying topic modelling using an adapted latent semantic analysis. The method indirectly identified topic similarity, as well. For example, the author "subtracted" *politics* from *The_Donald* (a subreddit for Donald Trump supporters) to infer which topics *The_Donald*'s authors contributed most when not talking about politics.

A related study to our own by Hessel et al. [24] combined multiple metrics, using a comparison of author and term similarity to identify obscured interests of users by identifying links according to high user similarity and low term similarity. Using this method, the authors identified several interesting examples, such as the relationship between *LadiesofScience* and *FancyFollicles* (about primarily multicolor hair) and *craftit* (a crafting subreddit). The authors based their analysis only on a limited sample text post submissions (maximum 5000 posts per subreddit), rather than comments or submissions in other media formats (common in many of the popular subreddits). In our work, we extend the idea of finding high-author/low-text coherent subreddits by also identifying other misaligned variants.

2.2 Politics and Social Media

The idea of high topic coherence (high text, but low author, similarity) occurs implicitly in the study of political discourse in social networks. Though they discuss similar issues, authors rarely cross-post, leading to fragmentation. Adamic et al. [2] very clearly demonstrated the lack of cross-links between Democratic and Republican bloggers during the 2004 U.S. election. Within more recent social media contexts, Lotan [39] studied Facebook, Twitter and Instagram user networks discussing the topic of the strife at Gaza strip and showed fragmentation within the context of a specific topic. In Twitter, Liu et al. [35] found that users who often mention each other but don't follow each other are "at war." In our work, we demonstrate how warring sub-communities in Reddit can be detected.

Studies comparing text and network structure have also focused on political discourse. For example, Livne et al. [37] studied interactions between political candidates on Twitter during the U.S. 2010 midterm election using both network structure and tweeted. The works notes differences in the strength of correlation between network similarity and language similarity depending on political party. However, the work did not discuss the interaction between the measures. Conver et al. [13] discussed the difference between the mention and retweet network while describing political polarization in Twitter.

2.3 Link Inference and Community Detection

In the community detection literature, work on signed graphs (e.g., Gao et al. [20]) captures richer relationships, including antagonistic and supportive. In many social networks such signs are not explicit, nor easily inferred. We demonstrate below that misalignment can often be used to identify the valence of the link. Though link-prediction [21] has often focused on purely structural features to infer edges (e.g., a partial network already exists), there are examples of combining metadata and other features to infer these links [1, 34]. However, these techniques focus on combining evidence rather than isolating conflicting 'opinions' between features.

In our work, we use different kinds of similarity metrics between subreddits to create subreddit networks. Our work is partially motivated by work of Hric [26] and Peel et al. [50] which showed that different metadata creates different clusters in a network, which we find out to be very true for subreddit similarity networks. We use a similarity graph approach [40] to create networks from similarity matrices.

3 DATASET

We selected Reddit (www.reddit.com) due to its popularity and structure. Reddit acts as both aggregator of a diversity of content and as a discussion board. We obtained 10.5 years of Reddit data (posts, authors, comments, etc.) ranging from January of 2006 to June of 2016 [1]. We focus our analysis on the month of June 2016, the most recent month at the time of our retrieval. While we find 74,951 subreddits with at least one comment for this period, the distribution is long tail and 22.1% of these subreddits saw only one comment posted. We define a subreddit as "active" if it had more than 500 comments made by more than 100 unique authors in June 2016. Roughly, 500 comments corresponds to the 92.6 percentile in subreddit comment counts and 100 unique authors correspond to 90.45 percentile in subreddit unique author counts. We find 5,193 subreddits that met this criteria. Further filtering out subreddits with "over 18" flags (largely pornographic material), we were left with 4,924 subreddits. Overall, 62.3 million comments (122.7 million sentences) made by 10.6 million unique users were included in our analysis.

Even within this subset, subreddits' activity levels approximate a long-tail distribution. The median number of comments per subreddit was 2083, and the median unique authors was 545. The most active subreddit, *AskReddit*, had about 4.6 million comments made by about 568k unique authors. In contrast, *nashville*—a subreddit well above the median level of commenting activity—had 8573 comments made by 1492 unique authors. This disparity is partly a consequence of Reddit's design. New Reddit users are automatically subscribed to a changing set of "default" subreddits. In our June 2016 dataset, these 56 default subreddits (1.1%) all had more than 2 million subscribers each; no other subreddit had more than 1 million subscribers. These 56 subreddits account for 23.6% of comments.

[1]The dataset was compiled by Reddit user *Stuck_In_the_Matrix*, available at files.pushshift.io/reddit/)

4 METHOD

The standard analysis pipeline for transforming disconnected entities into a network is illustrated in Figure 1. It involves using a similarity metric to create a pairwise similarity matrix. This matrix, often normalized and thresholded, is treated as an adjacency matrix from which a network is constructed. Further analysis, such as community detection, can then be executed on this network. In our work, we assume that multiple such pipelines can exist in parallel and that comparing both intermediate and downstream data structures (e.g., similarity matrices, networks, or communities) can lead to interesting findings.

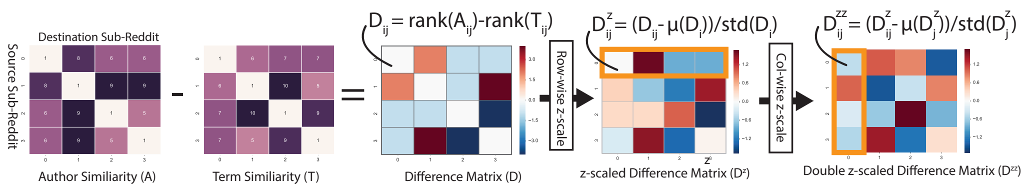

Fig. 2. Computing the z^2-score.

4.1 Similarity Metrics

In our analysis, we selected two common similarity measures: *text similarity* of comments and *author overlap*.

4.1.1 Textual similarity–T_{sim}. Text similarity was calculated by using the angle between term vectors describing each subreddit. Specifically, we applied the standard cosine similarity on a "bag-of-words" model that had been weighted through term frequency-inverse document frequency (TF-IDF) [44]. We applied standard NLP cleaning to all sentences in in a subreddit (stopword and punctuation removal, lowercasing, url removal) and phrase extraction. Sentences of two or fewer words remaining were ignored. Common multi-word tokens (phrases) were detected through a standard algorithm [44]. Specifically, one- to four-grams (all one- to four-word phrases) were extracted from 10% of the text, which we consider training data (roughly 12M sentences). Common "grams" (measured by the number of times the phrase appears relative to the individual terms) were retained. For each subreddit, the count of a particular term (tf_i) was normalized by the maximum frequency for all terms in that subreddit. The IDF frequencies utilized the number of subreddits that contain that term df_i. For calculating document frequency we used all subreddits from June, not only the core 4,924, which removes bias and partially controls for larger values from larger subreddits. The final feature vector for each subreddit contains the TF-IDF score for each term (a term is a single word or a multi-word phrase detected by our phrase detection algorithm) that appears in the corresponding subreddit.

There are many algorithms other than TF-IDF that can be used for measuring textual similarity between two subreddits. These algorithms include word embedding [45] and topic modelling [7, 58]. We opted for a simpler text similarity measure, TF-IDF, as both word embedding and topic modelling approaches are difficult to tune, and are significantly more costly in terms of space and time. This is a concern as our dataset contains 122.7 million sentences. TF-IDF is still a very good measure for measuring text similarity and widely used in information retrieval research.

4.1.2 Author similarity–A_{sim}. To calculate author similarity, we similarly calculated the cosine distance to the weighted (TF-IDF) "bag-of-authors." For each subreddit, author frequency (TF) was calculated as the number of times an author posted in the subreddit, normalized by the maximum number of posts made by a single author in that subreddit. Author IDF was determined by the number of subreddits the corresponding author posted on. As with text, for IDF calculation we considered all subreddits that were active in June. Deleted authors were removed.

4.2 Matrix Generation

Using the two similarity functions described above, we calculated the pairwise similarity of the 4,924 subreddits to generate two symmetric similarity matrices–A and T–for author and text similarity, respectively. A cell, A_{ij} (or T_{ij}) contained the result of the similarity calculation for subreddits i and j.

4.3 Pairwise relationships between subreddits

Once they are constructed, we are able to compare A and T (for example, using Spearman's rank correlation to calculate the correlation between the matrices). The author matrix is sparse, as many subreddits do not share any authors; in contrast, the term matrix contains no 0's as there is invariably some textual overlap. The next steps account for this difference between A and T.

4.3.1 Matrix agreement. To compare the likely links, or "edges," that will be formed from the matrices we set thresholds A_{thresh} and T_{thresh} as filters on the corresponding matrix. If the cell value is above the threshold, the cell was set to 1 (an edge exists); otherwise it is 0. Because the author matrix was already sparse, we set A_{thresh} to 0 so all non-zero cells were retained as edges. For the term matrix, T_{thresh} can be varied; we consider this a tunable hyper-parameter of the analysis pipeline. Once we transformed the matrices into binary form, we simply determined the agreement between them as a measure of similarity.

4.3.2 Binned comparison. As we have two scores for each subreddit pair, a natural analysis would map each pair onto a standard (though likely binned) x-y plot. One could then easily find pairs matching specific constraints; for example, one could find all subreddits pairs with a 90^{th} percentile score for text and less than the 10^{th} for author similarity. Pairs in this set would roughly correspond to topic-coherent pairs (high term but low author similarity).

This approach has a number of problems, however. First among them is that certain subreddits may dominate the pairings in a particular quadrant. For example, a default subreddit will likely have many author-coherent links, as they have authorship overlap with nearly all other subreddits even when they are topically unrelated. Second, we would ideally like a single score to identify misaligned subreddit connections. Neither the rank differences between similarities nor raw difference produce a satisfactory answer.

4.3.3 Double z-score (z^2). To create our misalignment metric, the z^2-score, we went through a four-step process of calculating and standardizing the differences between the author and term similarity matrices.

To generate a single score comparing the similarity matrix, we might expect to be able to simply calculate a new matrix D where each cell $D_{ij} = norm(A_{ij} - T_{ij})$, meaning each cell in the difference matrix would correspond to the *difference* in the values for that cell in the original similarity matrices. However, because the data distributions for author and term similarity are very different, we chose to calculate the difference matrix using *rank differences* instead of simply subtracting the raw similarity scores.

Thus, the first step was to create *ranked* similarity matrices, where the raw similarity scores for a given source subreddit and each of the 4,923 remaining destination subreddits are ranked against each other. In the original matrices, rows and columns were equivalent, as the raw similarity scores are symmetric. In these new ranked similarity matrices, this is no longer true: for any particular subreddit pair, it is very unlikely that the similarity relationship will be symmetric. For example, a small subreddit is likely to have high author overlap with a large, popular subreddit, simply because of its size; however, this overlap accounts for only a small proportion of the large subreddit's authorship, so the link returning from the large subreddit to the small one is likely to be ranked much lower.

The second step was to create a single rank-difference matrix D by subtracting the two rank-similarity matrices (the center matrix in Figure 2). In this asymmetric matrix, the rows represent the source subreddits; the columns represent the destination subreddits.

Because of the subreddits' diversity, the distributions of rank differences in each row are very different. Therefore, in the third step, we standardize the scores in each row by calculating the z-score (represented by the fourth matrix in Figure 2). Represented as an equation: $D_{ij} = norm(rank(A_{ij}) - rank(T_{ij}))$. Recall that the z-score (or standard score) K^z for a set of values K is calculated by subtracting the mean of K, μ_K, from each value k_i in K and dividing by the standard deviation of K, std_K. Thus, $K_i^z = (K_i - \mu_K)/std_K$. The z-score normalized values will be mean-centered on 0 and will capture the number of standard deviations the value is from the mean. In the matrix context, mean and standard deviation can be calculated per row or per column. Therefore, to calculate the single z-score transformed matrix, D^z, we determined the mean and standard deviation of each row D_i of the difference matrix D. Specifically, for any cell D_{ij}^z we computed $(D_{ij} - \mu(D_i))/std(D_i)$. The values in this matrix tell us the difference between author and term ranks for the source and destination subreddit, standardized by the distribution of source similarities.

This has not yet solved the problem of some subreddits simply being similar to all others, however. As described earlier, very large subreddits have this problem because of their size, but it can be caused by other subreddit quirks as well. *CatsStandingUp*, a popular image subreddit, is one example. When comparing its single z-score distribution to that of a second subreddit—say, *pokemongo*—*CatsStandingUp* has high positive z-score in D^z (see Figure 3). This is misleading, partly because *CatsStandingUp* has high author similarity with many subreddits, but also because it has unusually low text similarity with most other subreddits: the only word allowed in the comments is the word "cat." Commenting rules such as these can artificially inflate or deflate single z-scores.

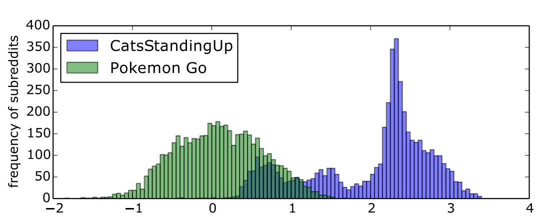

Fig. 3. Comparison between z-score distribution in rank difference list of all other subreddits for two subreddits: *CatsStandingUp* and *pokemongo*

To address this, we take our fourth and final step: taking the z-score again, this time *column-wise*. This produces D^{zz}: the double z-score (z^2-score) difference matrix (the rightmost matrix in Figure 2). For any cell D_{ij}^{zz} we compute $(D_{ij}^z - \mu(D_j^z))/std(D_j^z)$. Subreddits which have high positive z^2-score have high author coherence: higher author similarity than would be expected, given the term similarity. A high negative z^2-score indicates high topic coherence: higher term similarity than would be expected, given the author similarity. Where term and author similarities are about as expected (i.e. aligned), the z^2-score is close to 0.

This final matrix of z^2-scores, D^{zz}, is asymmetric and this matrix can be used to build directed and weighted networks.

4.4 Subreddit networks

Given our similarity matrices (author and term) and the z^2-score difference matrix, we are able to produce various network representations. While correlated, the networks have different semantics, each with a different application.

For the **Author similarity network**, an easy way to determine which edges should be created would be to use a threshold on author similarity and to create an edge between subreddits if the threshold is exceeded. However, we found that arbitrary thresholds result in a very dense central component with with many disconnected subreddits outside it. This is in part due to default subreddits with high similarity to all subreddits. We instead used an alternative approach popular for producing networks from pairwise similarity values [40]. In this method, we took the top 1% of similarity values for *each* subreddit to create edges. Each threshold is thus unique to the subreddit and ensures a connected graph. To further filter out very weak edges, we included only edges which are in the top 5^{th} percentile of similarity *globally*. The trade-off is that while this does not ensure a completely connected graph, most nodes are connected and edges are largely reliable in community-detection applications. Other approaches for generating networks (e.g., [57]) emphasize other features, such as preserving power-law degree distributions in the final network. The analysis we describe below is applicable when generating networks using alternative strategies.

To produce **Term similarity networks** we applied the same technique as above. As with the author network, we did not get a fully connected graph, but we did have a giant component containing majority of the nodes. As we used the same initial nodes (subreddits), we had a direct node-mapping between the author and term networks, though the edges were different.

Using the difference matrix D_{zz} we introduce the notion of **Misaligned networks** i.e. networks where edges are defined by misaligned links. Recall that when there is agreement between the author and term similarity for a pair of subreddits, the difference matrix cell will contain a value close to 0. These cells are common, and expected. Extremely high or extremely low values, on the other hand, are what we term "misaligned." Using the difference matrix, we produced two networks: the *author-coherent network*, containing only edges with a z^2-score of 3.0 or more, and the *topic-coherent network*, containing only edges with a z^2-score of -3.0 or less. The first contains only edges when the z^2-score is 3.0 or more. Because the matrices are asymmetric, the produced graphs are directed.

Given these networks, we were able to apply standard metrics such as the clustering coefficient (density of closed triangle compared compared to connected triplets).

4.4.1 *Community detection and modularity.*

There are many different community detection algorithms that can be used to detect communities in the undirected similarity networks and the directed misaligned networks. Some of the common algorithms for undirected networks are FastGreedy [12, 47], InfoMap [56], Label Propagation [54], Louvain or Multilevel [43], Spinglass [55] and Walktrap [51]. These algorithms make use of a varied range of underlying techniques for detecting communities. In recent comparisons [3, 30, 53], both Louvain and Infomap are shown to perform well. Louvain or multilevel algorithm [8, 43] is based on modularity maximization for community detection. Recall that modularity is a measure of cohesiveness of a network [48]. Louvain follows a hierarchical approach by first finding small, cohesive communities and then iteratively collapsing them in a hierarchical fashion. Infomap uses a different criteria. Specifically, it is an algorithm based on information theory that assumes that a random walk within a community is likely to stay within that community, as there are more intra-community edges than inter-community edges.

For the author and term similarity networks, we applied all six aforementioned algorithms and compared their results. Random walk based algorithms such as InfoMap and Walktrap produces a large number of very small communities, whereas Label Propagation and FastGreedy produces one very large community which encompasses most of the network. Louvain produces reasonably-sized (not too small or large) communities compared to other algorithms for these graphs and is scalable for larger graphs. For this reason we use Louvain for community detection in similarity networks.

However, Louvain can only be used on undirected networks. Therefore, for the directed misaligned networks we used Infomap [56]. InfoMap can be applied to both directed and undirected networks, which is a rare property among many community detection algorithms.

4.4.2 Measuring difference in the detected communities. To compare the different communities produced by the different networks, we used a metric called Normalized Mutual Information (NMI) [60], which produces a score between 0 and 1 depending on how similar two "partitionings" are. Generally, NMI is used to evaluate two different community partitions given by two different algorithms on the same network. Here, we used NMI to compare community partitions of different networks which shared the same set of nodes. Specifically, we used NMI to give us a single numerical representation of the difference between the communities in the author- and term-similarity networks.

We also calculated the μ-score, which is the proportion of edges that goes outside the community compared to all edges that are touching the community. We can extend this for communities derived from multiple networks. For example, we took all pairs of subreddits in a community and calculated average pairwise author and term similarity. For a "better" community these values should be higher. Especially for communities in the author similarity network, the average pairwise author similarity should be higher than that of the term similarity network; for term similarity network, the average pairwise term similarity should be higher. This gives us an external evaluation (i.e. not dependent on network properties) of the goodness of these communities.

5 RESULTS

The z^2-score can give us three types of information: information about individual subreddits, pairs of subreddits, and subreddit networks. In this section, we report descriptive statistics of the matrices, and then illustrate for each information type the characteristics and relationships that can be analyzed using the z^2-score, using examples from the Reddit data.

5.1 Matrices

5.1.1 Similarity matrices. An analysis of the similarity matrices produces the expected results, providing a first validation of the z^2-score. As expected, a Spearman's rank correlation found a weak positive correlation between author and term similarity, $r(12120424) = 0.266$, $p < 0.001$. The relationship between author and term similarity is visualized in Figure 4. We note the many subreddit pairs with both high author and high term similarity (represented by the relatively brighter bins in the upper right of the heatmap) and the high values along the diagonal.

When thresholding the two matrices for creating a simple network representation (i.e., a value above some threshold means an edge should exist) we find the overlap between matrices to range from 43% to 61%. We compute this by thresholding the author similarity matrix at above 0 and varying the term similarity threshold from 0.05 to 0.95. Edge agreement (based on thresholding) peaks at a threshold of 0.20.

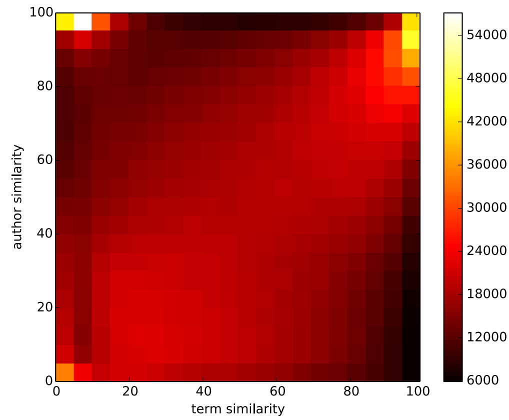

Fig. 4. Author similarity vs. term similarity for all pairs of subreddits that have non-zero author similarity score. Subreddit pairs are binned in a 20x20 grid according to percentile value. Brighter colors indicate more subreddit pairs in a 2-dimensional bin.

Given the fairly linear fit between author and term similarity we analyzed pairs by ranking outliers. Specifically, we ran a regression analysis on author similarity score and term similarity score and used Cook's distance [14] to identify outliers with undue influence on the regression line. Taking the 1% of subreddit pairs with the highest Cook's distance, we found that some defaults–*AskReddit*, *bestof*, *gifs*, *LifeProTips*–appeared in an unusually high number of pairs. This is consistent with expectations, as Reddit's default memberships induce high author overlap between the defaults and other subreddits.

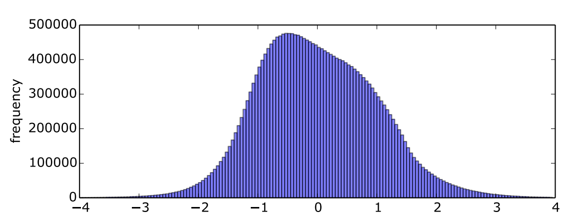

Fig. 5. Distribution of z^2-scores for all pairs of subreddits.

5.1.2 Misaligned matrices. We use the z^2-score difference values to find misaligned links both in terms of *absolute* and *relative* misalignment. We define absolute misalignment as z^2-scores of 3.0 or greater (author coherent) or less than -3.0 (topic coherent), meaning that it is more than 3 standard deviations away from the mean (as z-scores for a majority of subreddits are normally distributed). The distribution of z^2-scores for all subreddit pairs is plotted in Figure 5. On average, each subreddit has 18 outgoing author-coherent links and 9 outgoing topic-coherent links. However, there is considerable variation in this value, and in cases where a subreddit has few or no links with z^2-scores high enough to qualify as absolute misalignment, it is still valuable to look at the most misaligned links for that subreddit. We call these types of links *relatively* misaligned links.

5.2 Characterizing subreddit pairs

The z^2-score affords analysis of four types of subreddit pairs: strongly similar links (high term similarity and high author similarity), strongly dissimilar links (low term and low author similarity), misaligned author-coherent links and misaligned topic-coherent links. Each of these also have several subtypes. In this section, we report the results of the analysis of the misaligned links.

5.2.1 Author-coherent links. The first type of misaligned link is one with high author similarity but low term similarity (a high positive z^2-score). In many cases, this reveals a latent shared interest between the two subreddits. We call these *author-coherent links*. Author-coherent links appear in two types of relationships; we describe both below.

Hierarchical links: Author-coherent links can indicate that the subreddits are part of a hierarchy. One example is the subreddit *pokemon*, which had surprisingly high author similarity with a number of subreddits devoted to other videogames available on the Nintendo 3DS platform (*MyCastleFE*, *EtrianOdyssey*, *monsterhunterclan*, etc.), with z^2-scores ranging from 2.55 to 3.22. Author overlap can be attributed to Nintendo 3DS gamers talking about different games. These subreddits represent niche interests within the same broad topic (Nintendo 3DS games).

In the case of the Nintendo 3DS subreddits, the hierarchy is natural. Others have a hierarchical structure that is enforced through subreddit rules or norms. For example, the subreddit pair *mturk* and *HITsWorthTurkingFor* had a misalignment score of 3.42. Both had a high author overlap, but we expected a similarly high term overlap given the topics (both are forums for Amazon's Mechanical Turk crowd-sourcing service).

One strategy for differentiating between natural and enforced hierarchical links is to perform a log-odds analysis [46] of the text in the two subreddits (identifying which terms were more probabilistically likely to appear in one of the subreddits or the other). This makes it possible to identify important differences in terms. Words most likely to appear in *mturk* were related to a general discussion of MTurk (e.g. "work," "pay," "mturk"). The top phrases in *HITsWorthTurkingFor* are from a bot using the same template repeatedly (e.g. "bot action performed automatically"), with very little discussion otherwise. This shows an enforced hierarchical pair, where *mturk* is the general-interest subreddit and *HITsWorthTurkingFor* is a niche subreddit for posting work with primarily bot activity in the comments.

Community fragmentation: Author-coherent links are sometimes an indication of community fragmentation: groups that we would expect to share one community are in fact spread across several communities. For example, *USMilitarySO*, a subreddit for the significant others of U.S. military members, shared high author coherence with subreddits about budget makeup (*drugstoreMUA*, *MakeupRehab*), pregnancy and motherhood (*CautiousBB*, *clothdiaps*), and local subreddits for cities with large military presence (*jacksonville*, *MotoLA*). This is an indication that though many of *USMilitarySO*'s members also post on, for example, *CautiousBB*, they discuss pregnancy primarily on *CautiousBB* and not on *USMilitarySO*. In other words, the community is fragmented across multiple different subreddits.

This interpretation can be validated by "adding" the text of the two subreddits together and calculating term cosine similarity between the combined text (in our example, *CautiousBB +USMilitarySO*) and all other subreddits. Subreddits that have similar levels of cosine similarity with *CautiousBB* and *USMilitarySO* individually, and high cosine similarity with *CautiousBB +USMilitarySO*, are subreddits represent the topic overlap between the original two. *TheGirlSurvivalGuide* is an example of this: it has a cosine similarity of 0.67 and 0.63 with *USMilitarySO* and *CautiousBB*, respectively, and 0.76 for *CautiousBB +USMilitarySO*.

In contrast, subreddits that have high cosine similarity with the combined *CautiousBB +USMilitarySO*, but high cosine similarity with only one of the subreddits when taken individually, represent the fragments of the community. Therefore, while *pregnant* has high cosine similarity (0.80) with *CautiousBB* +*USMilitarySO*, it has a much higher similarity with *CautiousBB* than with *USMilitarySO* (0.84 versus 0.53). This validates our initial interpretation that users move to subreddits like *CautiousBB* to discuss pregnancy and do not discuss it frequently on *USMilitarySO*.

In this way, it is possible to identify author-coherent links that indicate community fragmentation without relying on interpretation or domain knowledge of the subreddits in question, and differentiate them from hierarchical author-coherent links.

5.2.2 Topic-coherent links. A second type of misaligned subreddit pairs are those with very negative z^2-scores. Here we find high term similarity and low author similarity, which we call *topic-coherent links*.

Communities at war: In some cases, topic-coherent links connect communities that have opposing opinions on the same topic. One example is *TrollXChromosomes* (a feminist subreddit) $\longrightarrow$ *MGTOW* (an anti-woman subreddit), which has a z^2-score of -3.05. Another example is *askscience* (form for discussion of science-related topics) and *theworldisflat*, where participants look for scientific evidence that the world is flat (z^2 of -3.01). In both of examples, the subreddits use similar language but relatively few authors in the one subreddit also post in the other.

One effect of the z^2-score is that when opposing communities cross-post, the score is close to 0. For example, one natural place to look for topic-coherent links is in political subreddits. By June 2016, both Hillary Clinton and Donald Trump had gained the presumptive nomination for their respective parties in the 2016 U.S. presidential election. However, the z^2-scores for the candidates' largest subreddits indicate relatively low topic cohesion. This is because they ranked highly in both term and author similarity. *The_Donald* $\longrightarrow$ *hillaryclinton* had a z^2-score near zero, -0.25, indicating the text and author similarities were about what was expected. The z^2-score for *hillaryclinton* $\longrightarrow$ *The_Donald* (the mirrored link) was higher, at -1.90, but in both cases they were in each other's top decile in both author and text similarity.

Topic-coherent fragmentation: Other topic-coherent links connect communities that are not necessarily antagonistic. A number of subreddits about different programming languages have high topic coherence, e.g. *javascript* and *matlab*. A number of programming subreddits—*java*, *javahelp*, *programming*—have high topic coherence with subreddits for students taking computer science courses, such as *OSUOnlineCS* and *cs50* (a Harvard University programming class). Interestingly, some of these links are approximately symmetric, often not the case for these misaligned links: *javahelp* $\longrightarrow$ *cs50* has a z^2-score of -3.76, while *cs50* $\longrightarrow$ *javahelp* has a z^2-score of -3.02. With mutually low author overlap, this example suggests that the students of *cs50* are not utilizing Reddit as a resource for programming help; however, the z^2-score alone cannot differentiate between antagonistic and non-antagonistic topic coherence.

5.3 Characterizing individual subreddits

In the previous section, we discussed examples primarily in terms of relative misalignment. If we focus only on the most misaligned links overall—those greater than 3.0 or less than -3.0—we see that some subreddits have many more outgoing misaligned links than others, or many more incoming misaligned links. For example, *USMilitarySO* has 55 outgoing links outside of the -3.0 to 3.0 range, while *The_Donald* has only one.

Intuitively, we might expect a subreddit with *many outgoing* links with high author coherence (high mean z^2) to mean that the authors in that subreddit also post in many other parts of Reddit about unrelated topics. *Few incoming* high author coherence links, that might mean that Reddit's "mainstream" is not interested in the subreddit. For example, a mother might post on many subreddits, but mostly talk about motherhood on subreddits devoted to the topic, such as *Mommit*. Similarly, having many outgoing topic-coherent links with few incoming topic-coherent links could indicate that the subreddit serves a niche audience for a general topic that is part of the Reddit mainstream, such as a subreddit for trading skins and other equipment within a specific popular video game (*csgotrade* is one example of this). Conversely, we would expect subreddits with many incoming links and not a lot of outgoing links to be gathering places on Reddit. If there are both author- and topic-coherent links incoming, we might be able to characterize these subreddits as Reddit's "mainstream": places where many different authors with many different interests gather to discuss topics of universal interest. By the same token, a subreddit with many more outgoing than incoming links that are both author- and topic-coherent might represent a community that exists on the margins of the Reddit mainstream: not isolated, but not fully accepted, either.

5.3.1 Mainstream vs. marginalized subreddits. To investigate this, we looked at the proportion of outgoing versus incoming links for each subreddit. This proportion of outgoing versus incoming links is similar to what is called the *hub* and *authority* scores [29], where hubs have many outgoing links and authorities have many incoming links. For this analysis, we combined author- and topic-coherent links (i.e., edges were considered if they had a z^2 above 3 or below -3). This made it possible to filter outliers like *CatsStandingUp*, which has an artificially high number of outgoing author-coherent links because its text is limited to the single word "cat." No comprehensive categorization of subreddits exists, so we focused on a small sample of the subreddits at the extremes: those with many more outgoing links, and with many more incoming links. Specifically, we selected subreddits which had above the median number of total links and were in the highest 5% of subreddits with more outgoing than incoming links, and vice versa. We then assigned each subreddit to one of several broad categories according to topic.

This expected pattern is reflected in the data. Subreddits displaying authority behavior—many more incoming than outgoing links—tend to have content that is appealing to a general Reddit audience. Subreddits displaying hub behavior—many more outgoing than incoming links—tend to cater to identities that are marginalized from the Reddit mainstream.

We categorized 122 subreddits with the highest proportion of incoming to outgoing links. Of these, 30 (24.6%) were default subreddits, which tend to appeal to general audiences. Another 22 (18.0%) were image boards such as *nononononoyes* and *reactiongifs*. Much of the remaining content covered topics of technology, gaming, and comics. In total, these subreddits—we refer to them broadly as "internet culture"—comprise 63.1% of these 122 subreddits. Of the remaining subreddits, 5 were far-right political forums, 1 was anti-capitalist, and 3 were advice forums targeted at a male audience; the others did not fall into any particular category.

The 121 subreddits with the highest proportion of outgoing to incoming links had a different composition. Of these, only 28 (23.1%) were internet culture subreddits, and tended to have a more specific focus (e.g. *xmen* instead of *comics*). In the remaining subreddits, 28 (23.1%) targeted specific identities that are less well-represented in Reddit's readership: women, people of color, people over 30, and local subreddits for specific cities. Two far-right and four far-left forums were also in this group. The remaining subreddits did not fall into any particular category.

To validate our analysis, we calculated hub and authority scores using the HITS algorithm [29] in a subreddit network with the most misaligned links as directed edges. The score is commonly used in network analysis and identifies central nodes based on both the number of incoming links (authorities) and outgoing links (hubs). The correlations were fairly strong for both: the authority score and number of incoming links had a correlation of $r(2388) = 0.71$, $p < 0.001$, and the hub score and number of outgoing links had a correlation of $r(2388) = 0.44$, $p < 0.001$. In addition, a few selected case studies from throughout the proportion distribution were also consistent. Since Reddit contributors are more likely to be male than female, we would expect subreddits targeted toward men to have more incoming than outgoing links, with the opposite being true for subreddits targeted toward women. *AskMen* has 197 incoming links and 0 outgoing links (an authority); in contrast, *AskWomen*, a subreddit with an almost identical number of subscribers, has 1 incoming link and 21 outgoing links (a hub). Further, in a selection of subreddits targeted toward a male or female audience (six each), three of the male-targeted subreddits were hubs, while for female-targeted subreddits only one qualified as a hub, the default subreddit *TwoXChromosomes*. Six LGBTQ subreddits were more evenly split: three hubs, three authorities. However, subreddits targeted specifically toward trans and genderqueer folk were much more likely to be hubs; of the 10 subreddits tested, eight had more outgoing than incoming links.

5.3.2 Satellite subreddits. Most misaligned pairs have a high author similarity rank and low term similarity rank, or vice versa. That is not always the case, however, as the z^2-score is not simply a measure of rank difference. Some links have a high z^2-score even though both their author and term similarity ranks are both very high, i.e. in the top 10% for the origin subreddit for both similarity types. These links are interesting because they identify relatively close relationships for subreddits that are otherwise unusually isolated in authorship or shared terms. We call these *satellite pair links*.

One example of this type of subreddit is *Divorce*, an advice and support forum with relatively high term similarity but very low author similarity with the rest of Reddit. It has a median raw term similarity score of 0.33, compared to an overall median of 0.22 for all subreddit links, and yet shares an author in common with only 27.4% of other subreddits, compared to 60.7% for all subreddits. Consequently, the pair *Divorce* $\rightarrow$ *datingoverthirty* has surprisingly high author coherence (z^2-score of 3.11) even though the raw authorship similarity score is low. This suggests that the authors in these subreddits are usually isolated from the rest of the network compared to other subreddits that discuss similar topics.

Subreddits that are unusually isolated in authorship but have multiple outgoing links of high author coherence tend to share common characteristics. Like *Divorce*, most of these subreddits are advice and support subreddits for relationships, mental health issues, or other medical problems. Of the 50 subreddits that have five or more of these high-author satellite pair links, 44 fall under this category. Examples include *selfharm*, *TryingForABaby*, *Fibromyalgia*, and *rapecounseling*. These 50 subreddits have a median text similarity considerably higher than the median for all subreddits (0.31 versus 0.22), but have lower median shared authorship than Reddit at large (median 47.5% versus 60.7%). Further, the majority of these links (55.6%) are to other isolated advice and support subreddits.

Less common are subreddits that are unusually isolated in term similarity with surprisingly high topic coherence. Only 139 of these pairs exist, compared to 2258 for authorship satellite pairs, which make these subreddits more difficult to characterize.

5.4 Characterizing subreddit networks

5.4.1 Author similarity network. Community detection produced 8 large communities (Table 1) in the author similarity network that include all but 6 of the 4,924 nodes (those remaining were sorted into communities of less than three nodes).

As shown in the table, author communities tend to have a relatively high μ-score (median = 0.54) and a relatively low clustering coefficient (median = 0.20), indicating that community structure is fairly poor. This partly a function of the size of the communities. As measured by the clustering coefficient, cohesiveness improves in the smaller communities. Pairwise raw similarity scores are another way to evaluate the community structure. While at first glance, the median raw pairwise author similarities seem very low, this is because these scores are very low overall, with an overall

Top 3 subreddits	Size	μ-score	cc	Aut. score	Term score
funny todayilearned pics	1128	0.44	0.16	0	0.21
AskReddit worldnews announcements	988	0.52	0.19	0	0.18
IAmA videos Music	887	0.35	0.16	0.004	0.17
gaming leagueoflegends Games	740	0.49	0.17	0	0.19
space technology europe	595	0.55	0.24	0	0.21
sports BlackPeopleTwitter nfl	360	0.57	0.22	0	0.34
Fitness trees Drugs	160	0.62	0.44	0	0.32
Art ArtisanVideos montageparodies	60	0.56	0.74	0.001	0.23

Table 1. Top communities in the author similarity network.

Top 3 subreddits	Size	μ-score	cc	Auth. score	Term score
gaming sports leagueoflegends	845	0.24	0.27	0	0.31
IAmA blog tifu	728	0.44	0.25	0	0.22
AskReddit funny todayilearned	621	0.60	0.18	0	0.21
worldnews announcements news	516	0.39	0.30	0	0.39
LifeProTips Futurology technology	503	0.53	0.31	0.001	0.45
pics travel beer	289	0.42	0.32	0	0.59
hiphopheads Guitar tipofmytongue	155	0.12	0.44	0.001	0.43
nonononoyes cars bicycling	143	0.13	0.71	0.002	0.46
hearthstone magicTCG CompetitiveHS	45	0.01	0.72	0	0.44
gardening Aquariums Fishing	30	0.08	0.66	0.001	0.39

Table 2. Top 10 communities in the term similarity network. The community's three largest subreddits are listed, along with the size and a number of metrics that measure the quality of the community structure: μ-score, clustering coefficient (CC), and median raw pairwise similarity scores for both authors and terms.

median of 5.24×10^{-6}. The author communities have a median pairwise author similarity that is about 10 times higher than the overall median, an indication that the identified communities are reflecting a true community structure, even if the μ-score and clustering coefficient indicate that it is not strong.

As with the author similarity *matrix*, it is likely that the muddy structure of the author similarity *network* is a reflection of Reddit's design. Because all new users begin with a similar set of default subreddits from which they explore other parts of Reddit, those defaults have author connections to many other subreddits. This would explain why all of the identified communities in the author similarity network include at least one default subreddit, and the largest communities include multiple defaults. Even so, there is a coherent theme in many of these communities: *gaming*, *leagueoflegends*, and *Games*, as an example, or *trees* and *Drugs* (*trees* is a subreddit for users of cannabis).

5.4.2 Term similarity network. Compared to the author similarity network, the term similarity network produced many more communities (Table 2): 26 in total, with more variation in size (3 to 845 subreddits). The term similarity network also produced many more communities of one to two nodes, totaling 19.8% of subreddits.

The term similarity network had low μ-scores and high clustering coefficients: communities in the term network have a median μ-score of 0.007 and median clustering coefficient of 0.71. This indicates a good community structure, particularly in comparison to the author similarity network.

As with the author communities, cohesiveness improves as the community shrinks. The μ-score also tends to get lower as the communities get smaller, an indication that the community structure is improving. Looking at the pairwise term similarity scores, the term communities' median score of 0.30 is higher than the overall median of 0.22.

As with the communities in the author network, many of the term communities include defaults—unsurprising, given that the defaults are chosen to have widespread appeal—but the influence is not as strong so not all communities include defaults. As would be expected, the themes of each community are clearer than in the author communities, and we can clearly make out different gaming, news, technology, music, and hobby-related communities.

5.4.3 Comparing similarity networks. We find the NMI score between the term and author derived networks to be fairly low at 0.284. In this particular pairing the differences are likely due to the existence of default subreddits. Communities in author similarity network are larger and mostly centered around these defaults, whereas communities in term similarity network vary greatly in size and are topically more cohesive. However, we note that the largest communities in both networks share many common nodes.

5.4.4 Misaligned networks. We computed two misaligned networks using the z^2-scores, one with high author-coherent links ($z^2 \geq 3$) and one with high topic-coherent links ($z^2 \leq 3$). Both are composed of one giant component (containing 2670 and 2023 nodes for the author- and topic-coherent networks, respectively), and many small connected components.

After running InfoMap community detection algorithm, we found the clearest structure in the topic-coherent network. Disregarding communities with fewer than 5 subreddits, the algorithm found one large community of 1940 subreddits and many 127 smaller communities ranging in size from 5 to 143. These communities usually have very low clustering coefficient (median 0.0) and very high *mu*-score (median 0.7), but have median pairwise author and term similarity scores that are much higher than the communities in the similarity networks (3.3×10^{-4} the median author score and 0.31 median term score in the similarity networks). Table 3 shows some examples of these topic-coherent communities. We observe that they consist of subreddits related by broad general category (e.g., "sports") with different sub-topics that often have very little to no overlap in participation (*nfl*, *soccer*, *nba*).

Top 3 subreddits	**Size**	**Broad category**
listentothis tipofmytongue electronicmusic	143	Music and music/video recommendation
hearthstone magicTCG boardgames	91	Board games and card video games
Python perfectloops dailyprogrammer	81	Programming
Diablo diablo3 Smite	68	Online multiplayer gaming
Gunners LiverpoolFC Seahawks	48	Sports fan clubs
EatCheapAndHealthy keto fitmeals	46	Food and health
nfl soccer nba	45	Sports
compsci jobs cscareerquestions	44	Education and career
bicycling motorcycles MTB	39	Motorcycles and bicycles

Table 3. Some interesting misaligned communities in the topic-coherent network. The community's three largest subreddits are listed, along with its size and broad categories.

5.5 Summary

The z^2-score method makes it possible to characterize communities and the relationships between them in a dense network with little explicit structure. By using multiple measures of similarity—author overlap and term similarity—we are able to identify relationships that would be invisible if one used a single measure, and use those relationships to analyze elements in the network that are of interest to researchers. Below, we summarize how the z^2-score can be used to analyze different elements in the network.

5.5.1 *Individual subreddits.*

- **Mainstream vs. marginalized subreddits**: By calculating the proportion of the total number of incoming and outgoing links with extreme z^2-scores (similar to *hub* and *authority* scores in network analysis) one can differentiate between communities that are part of the network's mainstream and communities that participate in the mainstream but are pushed to the margins. In this way, one can also learn about the network as a whole: what is mainstream, what is marginalized, and what is isolated.
- **Satellite subreddits**: Satellite subreddits have authors that are more isolated from the rest of the network than would be expected, given the subreddit topic, or vice versa; in our analysis, these communities tended to cater toward vulnerable users (e.g. *selfharm*). These are distinct from subreddits that are just isolated, which can be found simply by using the similarity matrices to find subreddits that have low term and low author similarity. Satellite subreddits can be identified using a particular subtype of author- and topic-coherent links we call *satellite pair links*, in which two subreddits have *both* very high author and very high term similarity but still have a high z^2 because both subreddits are otherwise unusually isolated from other subreddits.

5.5.2 *Subreddit pairs.*

- **Author-coherent links**: These relationships indicate that a community (in this analysis, a subreddit) shares more authors with another community than would be expected, given their level of textual similarity. These links can identify two different phenomena: *hierarchical links* and *community fragmentation*. Hierarchical links occur between subreddits that represent niche interests within the same broad topic; these can occur naturally, or can occur as a consequence of subreddit rules. Community fragmentation occurs when a single group of users that we would expect to share one community instead spread across multiple communities to discuss different topics. These links are interesting because they can reveal a shared interest in topics that do not always seem related at first glance.
- **Topic-coherent links**: In these relationships, a community has more textual similarity with another community than would be expected, given their author overlap. This usually means that two different groups of people are discussing the same topic, but not talking to each other. These can take several different forms. Some are links between *communities "at war"* and not speaking to each other; conversely, links with low z^2-scores between two antagonistic communities indicate that they are cross-posting on each others' forums. Topic-coherent links are not always an indication of antagonism, however; in *topic-coherent fragmentation*, the linked communities may have ambivalent or neutral opinions of the other.

5.5.3 Subreddit networks. The z^2-score can also be used to characterize a network as a whole. Unlike the results for individual subreddits and subreddit pairs, where we reported only the misaligned results, we analyzed networks constructed from the similarity matrices as well for additional face validity. In the undirected similarity networks, both author and term networks show definite community structure in Reddit, although the communities in author similarity network are very much focused on default subreddits as an artifact of Reddit design (all new users are automatically subscribed to default subreddits). For the directed misaligned networks that used z^2-scores, the author-coherent network did not show a pronounced network structure when a directed community detection algorithm is run, but topic-coherent network produced small communities based on a common, broad interest with the individual subreddits different enough from each other that they have few users in common.

6 DISCUSSION

In this work we demonstrate how the z^2-score based methodology can be used to find misaligned links and communities. The z^2-score makes it possible to find particular relationships that have been identified as interesting by previous research but have been difficult to find and characterize systematically. Hierarchical communities and community fragmentation, both of which can be identified using the z^2-score, are important in understanding how a network and its communities develop over time. In addition, the z^2-score may have other practical applications, such as identifying vulnerable users who may be at risk of harm. For example, high-author satellite subreddits are populated by users who are unusually isolated from the rest of Reddit; as this tends to correspond with subreddits such as *SuicideWatch* and *selfharm*, these users may be in particular need of support.

6.1 Limitations

The z^2-scores alone can not distinguish between all relationship types; in many cases, human input and domain knowledge is required to interpret the z^2-score results. For example, we can not distinguish "communities at war" from other topic-coherent links. Human input or additional NLP techniques (e.g., sentiment analysis) are necessary. Though rare we also find that if there is significant cross-posting between warring subreddits, then both author and term similarity between the subreddits are high. Thus, z^2-score alone can not detect these pairs. For example, *The_Donald* and *hillaryclinton* are warring subreddits, but many authors from *The_Donald* posted in *hillaryclinton* in the month of June. This resulted in high author similarity between both. We also need to employ additional measures to identify misaligned links that are produced by subreddit moderation like *mturk* $\rightarrow$ *HITsWorthTurkingFor*. We demonstrated some techniques that can automatically distinguish between relationship types–for example, the subreddit "addition" that makes it possible to differentiate between hierarchical and fragmented communities–but some level of human interpretation was still necessary in many of our reported results.

We also need to keep in mind that z^2-score makes use of pairwise similarity values that necessities access to metadata of all social entities in the process. Getting access to metadata is difficult in many cases. For example, we do not have access to moderator list of all public subreddits. This limits the usefulness of z^2-scores when using common moderators as a similarity measure in Reddit.

6.2 Future work

6.2.1 Expanding text analysis. Additional quantitative methods such as log-odds ratios or sentiment analysis could shore up validity in future analyses. For example, sentiment analysis might be able to differentiate between communities at war and topic-coherent fragmentation. Developing additional techniques that can automatically differentiate between different types of relationships will increase the method's validity and reduce the need for subjective interpretation. It would also be useful in z^2-score analyses that look for additional hidden structures in the network. For example, Reddit has a small network of non-English-language subreddits. Unsurprisingly, these tend to be more isolated from the rest of Reddit. However, links of high author coherence pointing to these subreddits from English-language subreddits can show points of crossover, e.g. between English and non-English soccer subreddits.

6.2.2 Additional similarity measures. Useful insights can be gained from using other kinds of similarity measures between subreddits, such as a moderator network where edges between subreddits indicate that they share at least one common moderator. In a preliminary analysis, we found a community of Internet meme subreddits (*dankmemes*) and subreddits of popular animes in meme culture (*KillLaKill*, *cowboybebop*). This is not apparent when observing only author and term similarity networks.

6.2.3 Application in other social media. The concept of z^2-scores can readily be applied to other social media and social networking websites like Twitter or Facebook. For example, differences tweet hashtag use and @-mentions can reveal nuances in communication in Twitter. Our pipeline does not restrict the user from using any kind of similarity measure between two entities. Moreover, choice of similarity measurement algorithms or community detection algorithms can be fine-tuned. We believe with appropriate choice of similarity measures and algorithms z^2-scores can be used to detect "communities in war", community fragmentation or isolated groups in other social media. However, as discussed before, when we have incomplete data, z^2-scores have limited usefulness as they depend on pairwise similarity values.

7 CONCLUSION

In this paper, we describe a method for inferring network structure using different similarity metrics for social media data. Rather than focusing on agreement between different scores, we identify the importance of differences in capturing uncommon structures and behaviors. We provide a method for comparing the pairwise similarity matrices and a normalization (the z^2-score) that identifies 'misaligned' connections. Both extremely high and extremely low values of z^2 can be used to produce 'misaligned networks' that display topical or author coherence. We apply these methods to the study of subreddits and demonstrate that they are able to identify (and help in classifying) different types of behavioral patterns as well as artifacts of UX design. We believe that our technique can be applied in other scenarios where network inference is employed in the study of social media.

ACKNOWLEDGEMENTS

This work was partially supported by the NSF under grant IIS-1421438.

REFERENCES

[1] Lada A Adamic and Eytan Adar. 2003. Friends and neighbors on the Web. *Social Networks* 25, 3 (2003), 211 – 230. https://doi.org/10.1016/S0378-8733(03)00009-1

[2] Lada A. Adamic and Natalie Glance. 2005. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD '05)*. ACM, New York, NY, USA, 36–43. https://doi.org/10.1145/1134271.1134277

[3] Rodrigo Aldecoa and Ignacio MarÃŋn. 2013. Exploring the limits of community detection strategies in complex networks. *CoRR* abs/1306.4149 (2013). http://dblp.uni-trier.de/db/journals/corr/corr1306.html#AldecoaM13

[4] Nazanin Andalibi, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding Social Media Disclosures of Sexual Abuse Through the Lenses of Support Seeking and Anonymity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3906–3918. https://doi.org/10.1145/2858036.2858096

[5] Shrey Bagroy, Ponnurangam Kumaraguru, and Munmun De Choudhury. 2017. A Social Media Based Index of Mental Well-Being in College Campuses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1634–1646. https://doi.org/10.1145/3025453.3025909

[6] Sairam Balani and Munmun De Choudhury. 2015. Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*. ACM, New York, NY, USA, 1373–1378. https://doi.org/10.1145/2702613.2732733

[7] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

[8] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (2008), P10008. http://stacks.iop.org/1742-5468/2008/i=10/a=P10008

[9] Alissa Centivany and Bobby Glushko. 2016. "Popcorn Tastes Good": Participatory Policymaking and Reddit's. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1126–1137. https://doi.org/10.1145/2858036.2858516

[10] Eshwar Chandrasekharan, Mattia Samory, Anirudh Srinivasan, and Eric Gilbert. 2017. The Bag of Communities: Identifying Abusive Behavior Online with Preexisting Internet Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3175–3187. https://doi.org/10.1145/3025453.3026018

[11] Jilin Chen, Gary Hsieh, Jalal U. Mahmud, and Jeffrey Nichols. 2014. Understanding Individuals' Personal Values from Social Media Word Use. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '14)*. ACM, New York, NY, USA, 405–414. https://doi.org/10.1145/2531602.2531608

[12] Aaron Clauset, M. E. J. Newman, , and Cristopher Moore. 2004. Finding community structure in very large networks. *Physical Review E* (2004), 1– 6. https://doi.org/10.1103/PhysRevE.70.066111

[13] Michael Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2011. Political Polarization on Twitter. In *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2847

[14] R. Dennis Cook. 1977. Detection of Influential Observation in Linear Regression. *Technometrics* 19, 1 (1977), 15–18. http://www.jstor.org/stable/1268249

[15] Tiago Oliveira Cunha, Ingmar Weber, Hamed Haddadi, and Gisele L. Pappa. 2016. The Effect of Social Feedback in a Reddit Weight Loss Community. *CoRR* abs/1602.07936 (2016). http://arxiv.org/abs/1602.07936

[16] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2098–2110. https://doi.org/10.1145/2858036.2858207

[17] Munmun De Choudhury, Sanket S. Sharma, Tomaz Logar, Wouter Eekhout, and René Clausen Nielsen. 2017. Gender and Cross-Cultural Differences in Social Media Disclosures of Mental Illness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 353–369. https://doi.org/10.1145/2998181.2998220

[18] Ethan Fast and Eric Horvitz. 2016. Identifying Dogmatism in Social Media: Signals and Models. *CoRR* abs/1609.00425 (2016). http://arxiv.org/abs/1609.00425

[19] Deen Freelon, Marc Lynch, and Sean Aday. 2015. Online Fragmentation in Wartime. *The ANNALS of the American Academy of Political and Social Science* 659, 1 (2015), 166–179. https://doi.org/10.1177/0002716214563921 arXiv:http://dx.doi.org/10.1177/0002716214563921

[20] Ming Gao, Ee-Peng Lim, David Lo, and Philips Kokoh Prasetyo. 2016. On detecting maximal quasi antagonistic communities in signed graphs. *Data Mining and Knowledge Discovery* 30, 1 (2016), 99–146. https://doi.org/10.1007/s10618-015-0405-2

[21] Lise Getoor and Christopher P. Diehl. 2005. Link Mining: A Survey. *SIGKDD Explor. Newsl.* 7, 2 (Dec. 2005), 3–12. https://doi.org/10.1145/1117454.1117456

[22] Eric Gilbert. 2013. Widespread Underprovision on Reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 803–808. https://doi.org/10.1145/2441776.2441866

[23] W. L. Hamilton, J. Zhang, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec. 2017. Loyalty in Online Communities. *ArXiv e-prints* (March 2017). arXiv:1703.03386

[24] Jack Hessel, Alexandra Schofield, Lillian Lee, and David M. Mimno. 2015. What do Vegans do in their Spare Time? Latent Interest Detection in Multi-Community Networks. *CoRR* abs/1511.03371 (2015). http://arxiv.org/abs/1511.03371

[25] Jack Hessel, Chenhao Tan, and Lillian Lee. 2016. Science, AskScience, and BadScience: On the Coexistence of Highly Related Communities. *CoRR* abs/1612.07487 (2016).

[26] Darko Hric, Richard K. Darst, and Santo Fortunato. 2014. Community detection in networks: Structural communities versus ground truth. *Phys. Rev. E* 90 (Dec 2014), 062805. Issue 6. https://doi.org/10.1103/PhysRevE.90.062805

[27] Ramakanth Kavuluru, María Ramos-Morales, Tara Holaday, Amanda G. Williams, Laura Haye, and Julie Cerel. 2016. Classification of Helpful Comments on Online Suicide Watch Forums. In *Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB '16)*. ACM, New York, NY, USA, 32–40. https://doi.org/10.1145/2975167.2975170

[28] Charles Kiene, Andrés Monroy-Hernández, and Benjamin Mako Hill. 2016. Surviving an "Eternal September": How an Online Community Managed a Surge of Newcomers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1152–1156. https://doi.org/10.1145/2858036.2858356

[29] Jon M. Kleinberg. 1999. Authoritative Sources in a Hyperlinked Environment. *J. ACM* 46, 5 (Sept. 1999), 604–632. https://doi.org/10.1145/324133.324140

[30] Andrea Lancichinetti and Santo Fortunato. 2009. *Community detection algorithms: a comparative analysis.* arXiv e-print 0908.1062. http://arxiv.org/abs/0908.1062 Physical Review E 80, 056117 (2009).

[31] Alex Leavitt. 2015. "This is a Throwaway Account": Temporary Technical Identities and Perceptions of Anonymity in a Massive Online Community. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 317–327. https://doi.org/10.1145/2675133.2675175

[32] Alex Leavitt and Joshua A. Clark. 2014. Upvoting Hurricane Sandy: Event-based News Production Processes on a Social News Site. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1495–1504. https://doi.org/10.1145/2556288.2557140

[33] Alex Leavitt and John J. Robinson. 2017. The Role of Information Visibility in Network Gatekeeping: Information Aggregation on Reddit During Crisis Events. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1246–1261. https://doi.org/10.1145/2998181.2998299

[34] David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology* 58, 7 (2007), 1019–1031. https://doi.org/10.1002/asi.20591

[35] Zhe Liu and Ingmar Weber. 2014. *Is Twitter a Public Sphere for Online Conflicts? A Cross-Ideological and Cross-Hierarchical Look*. Springer International Publishing, Cham, 336–347. https://doi.org/10.1007/978-3-319-13734-6_25

[36] Zhe Liu and Ingmar Weber. 2014. Predicting Ideological Friends and Foes in Twitter Conflicts. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion)*. ACM, New York, NY, USA, 575–576. https://doi.org/10.1145/2567948.2576964

[37] Avishay Livne, Matthews P. Simmons, W. Abraham Gong, Eytan Adar, and Lada A. Adamic. 2011. The party is over here: structure and content in the 2010 election. In *ICWSM*. Association for the Advancement of Artificial Intelligence.

[38] Kiel Long, John Vines, Selina Sutton, Phillip Brooker, Tom Feltwell, Ben Kirman, Julie Barnett, and Shaun Lawson. 2017. "Could You Define That in Bot Terms"?: Requesting, Creating and Using Bots on Reddit. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3488–3500. https://doi.org/10.1145/3025453.3025830

[39] Gilad Lotan. 2014. Israel, Gaza, War & Data. (2014). https://medium.com/i-data/israel-gaza-war-data-a54969aeb23e

[40] Ulrike Luxburg. 2007. A Tutorial on Spectral Clustering. *Statistics and Computing* 17, 4 (Dec. 2007), 395–416. https://doi.org/10.1007/s11222-007-9033-z

[41] Trevor Martin. 2017. Dissecting Trumps most rabid online following. (2017). http://goo.gl/rVA510

[42] J. Nathan Matias. 2016. Going Dark: Social Factors in Collective Action Against Platform Operators in the Reddit Blackout. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1138–1151. https://doi.org/10.1145/2858036.2858391

[43] Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. 2011. Generalized Louvain method for community detection in large networks.. In *ISDA*. IEEE, 88–93. http://dblp.uni-trier.de/db/conf/isda/isda2011.html

[44] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781 (2013). http://arxiv.org/abs/1301.3781

[45] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13)*. Curran Associates Inc., USA, 3111–3119. http://dl.acm.org/citation.cfm?id=2999792.2999959

[46] B. L. Monroe, M. P. Colaresi, and K. M. Quinn. 2008. Fightin'Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* 16, 4 (2008), 372.

[47] M.E.J. Newman. 2003. Fast algorithm for detecting community structure in networks. *Physical Review E* 69 (September 2003). http://arxiv.org/abs/cond-mat/0309508

[48] M E Newman. 2006. Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103, 23 (June 2006), 8577–8582. https://doi.org/10.1073/pnas.0601602103

[49] Randal S. Olson and Zachary P. Neal. 2013. Navigating the massive world of reddit: Using backbone networks to map user interests in social media. *CoRR* abs/1312.3387 (2013).

[50] Leto Peel, Daniel B. Larremore, and Aaron Clauset. 2016. The ground truth about metadata and community detection in networks. (2016). arXiv:arXiv:1608.05878

[51] Pascal Pons and Matthieu Latapy. 2004. Computing communities in large networks using random walks. *J. of Graph Alg. and App. bf* 10 (2004), 284–293.

[52] Liza Potts and Angela Harrison. 2013. Interfaces As Rhetorical Constructions: Reddit and 4Chan During the Boston Marathon Bombings. In *Proceedings of the 31st ACM International Conference on Design of Communication (SIGDOC '13)*. ACM, New York, NY, USA, 143–150. https://doi.org/10.1145/2507065.2507079

[53] Arnau Prat-Pérez, David Dominguez-Sal, and Josep-Lluis Larriba-Pey. 2014. High Quality, Scalable and Parallel Community Detection for Large Real Graphs. In *Proceedings of the 23rd International Conference on World Wide Web (WWW '14)*. ACM, New York, NY, USA, 225–236. https://doi.org/10.1145/2566486.2568010

[54] Usha N. Raghavan, Reka Albert, and Soundar Kumara. 2007. Near linear time algorithm to detect community structures in large-scale networks. (Sept. 2007). arXiv:0709.2938 http://arxiv.org/abs/0709.2938

[55] Joerg Reichardt and Stefan Bornholdt. 2006. Statistical Mechanics of Community Detection. *Physical Review E* 74 (2006), 016110. http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0603718

[56] Martin Rosvall and Carl T. Bergstrom. 2008. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences* 105, 4 (2008), 1118–1123. https://doi.org/10.1073/pnas.0706851105 arXiv:http://www.pnas.org/content/105/4/1118.full.pdf+html

[57] M. Angeles Serrano, Marian Boguna, and Alessandro Vespignani. 2009. Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* 106, 16 (2009), 6483–6488. https://doi.org/10.1073/pnas.0808904106 arXiv:http://www.pnas.org/content/106/16/6483.full.pdf

[58] M. Steyvers and T. Griffiths. 2007. *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum, Chapter Probabilistic topic models.

[59] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of the 25th International Conference on World Wide Web (WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 613–624. https://doi.org/10.1145/2872427.2883081

[60] Lei Tang and Huan Liu. 2010. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery* 2, 1 (2010), 1–137.

[61] Greg Wadley, Wally Smith, Bernd Ploderer, Jon Pearce, Sarah Webber, Mark Whooley, and Ron Borland. 2014. What People Talk About when They Talk About Quitting. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design (OzCHI '14)*. ACM, New York, NY, USA, 388–391. https://doi.org/10.1145/2686612.2686671

[62] Eugene J Webb, Donald Thomas Campbell, Richard D Schwartz, and Lee Sechrest. 1966. *Unobtrusive measures: Nonreactive research in the social sciences*. Vol. 111. Rand McNally Chicago.

[63] Zhongyu Wei12, Yang Liu, and Yi Li. 2016. Is This Post Persuasive? Ranking Argumentative Comments in the Online Forum. In *The 54th Annual Meeting of the Association for Computational Linguistics*. 195.

[64] W.W. Zachary. 1977. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33 (1977), 452–473.

[65] J. Zhang, W. L. Hamilton, C. Danescu-Niculescu-Mizil, D. Jurafsky, and J. Leskovec. 2017. Community Identity and User Engagement in a Multi-Community Landscape. *ArXiv e-prints* (May 2017). arXiv:1705.09665

Received April 2017; revised July 2017; accepted November 2017