Towards Community Discovery in Signed Collaborative Interaction Networks

Petko Bogdanov, Nicholas D Larusso, Ambuj Singh
University of California, Santa Barbara, Computer Science Department
{petko, nlarusso, ambuj}@cs.ucsb.edu

***Abstract*—We propose a framework for discovery of collaborative community structure in Wiki-based knowledge repositories based on raw-content generation analysis. We leverage topic modelling in order to capture agreement and opposition of contributors and analyze these multi-modal relations to map communities in the contributor base. The key steps of our approach include (i) modeling of pairwise variable-strength contributor interactions that can be both positive and negative, (ii) synthesis of a global network incorporating all pairwise interactions, and (iii) detection and analysis of community structure encoded in such networks.**

The global community discovery algorithm we propose outperforms existing alternatives in identifying coherent clusters according to objective optimality criteria. Analysis of the discovered community structure reveals coalitions of common-interest editors who back each other in promoting some topics and collectively oppose other coalitions or single authors. We couple contributor interactions with content evolution and reveal the global picture of opposing themes within the self-regulated community base for both controversial and featured articles in Wikipedia.

I. Introduction

Internet users have access to an abundance of collaboration-enabled applications. The span of target activities ranges from collective image annotation and processing (Flickr, Picasa, Facebook) to text content generation (Wikipedia, Google Docs) and even social software development (github). Social media and social knowledge systems are among those with the highest impact in terms of popularity among end users. Following its launch in 2001, Wikipedia has become the most visited not-for-profit free-content encyclopedia, attracting contributors and users who spend five minutes a day on average editing and accessing information [1]. Activity rate of such magnitude generates an abundance of user interaction data, providing exciting new opportunities for understanding the multifaceted community structure of contributors and the topics they collectively promote.

The multitude of systems that keep track of online collaboration activity has provided new venues for the study of social interactions and behavior in collaborative and self-regulating environments [5]. This is evident by the large number of publications over the past decade, dedicated to mining and analysis of social networks [15]. Special care is necessary in interpreting user interactions in popular social networking systems, since the same type of link may have different meaning and strength [12]. Most of the research attention has been dedicated to studying networks of positive links of friendship and approval that reveal the architecture of social ecosystems only partially. Taking into account the full gamut of positive and negative links allows for in-depth understanding of network organization, both at the local and global level, and may lead to more realistic models of social interaction.

Our goal in this work is to identify coalitions of content contributors who share views on certain topics and collectively oppose other coalitions. Being able to map such communities is important for ensuring the content integrity and objectiveness. For example, if two groups with opposing views are actively editing an encyclopedic article, they can naturally regulate each other, thus producing unbiased content with little administrative overhead. In addition, by leveraging such a community-based model, one can automatically tag new content as (non)controversial, based on involved contributors.

On the other hand, extreme polarity of opinions may also have negative effects by causing *edit wars*, manifested as sequences of insertions and deletions of the same content. Such phenomena add overhead to the process of content generation as administrative resources must be allocated to arbitrate controversies. Prior understanding of opposing communities and their views may assist automated conflict resolution, minimizing the overhead for administrators.

In addition to enhancing the utility of collaboration systems, analysis of content evolution coupled with natural coalitions within the editor community can reveal latent social roles, adopted by community members [16]. Although neutral point of view and respectful interactions among editors are two of the *pillars* [1] of the Wikipedia community, controversial topics often fuel heated debates between users with opposing points of view. We are interested in discovering the stabilizing communities as well as those that work to move articles away from neutrality. Another interesting aspect of studying the functions of editor communities lies in understanding the span of hot topics and how they map to coalitions of contributors. Specifically, for our study of the Wikipedia data, we pose a series of questions such as: Are there differences in the interaction topology of contributors in stable and controversial articles? Are there coalitions that always take a neutral standpoint on conflict issues and others that collectively push the boundaries of objectiveness? Which topics within articles are controversial and who are the contributors who take opposing sides on them?

Addressing the aforementioned questions requires (i) a

[1]http://www.alexa.com/siteinfo/wikipedia.org

suitable model of contributor interaction based on editing behavior, (ii) a method to synthesize global contributor networks, and (iii) algorithms for community detection within the latter. In this work, we take the first steps towards providing a framework for constructing an interaction network from the set of historical revisions of a text resource. The editing activity model we assume is based on two hypotheses of content revision incentives: (1) when revising text, a user approves of all previously contributed ideas that she/he leaves unchanged and (2) users are drawn to edit specific content dependent only on its current state produced by the author of the preceding revision.

We analyze the revision history of articles in Wikipedia to showcase our framework for synthesis and mining interaction networks and present an analysis of our findings. The methods, however, are general and can be applied to other collaboration systems by considering suitable criteria to detect the mode of pairwise contributor interactions. We target contributor interactions arising from the process of content generation itself and synthesize an author collaboration networks, corresponding to Wikipedia articles. In contrast to traditional networks studied in social science, the networks we infer encode a gradient of interaction strengths from both the positive and negative spectrum. We adopt desirable criteria for optimal partitioning of nodes in such networks and introduce a method that achieves this goal, while outperforming previous algorithms for community discovery in signed graphs.

A unique aspect of our work is that we combine the social network (consisting of edits by users) with the information network (consisting of the latent topics of a page and how these topics evolve). The analysis of such an integrated network opens up new possibilities for analysis of user behavior and interactions.

Our contributions in this work can be summarized as follows:

1) We propose a framework for constructing signed user interaction networks from content generation flow. These networks encode links between users who interact while authoring the same resource and provide a gradient of their level of (dis)agreement. The result is a network with edge weights in the range $[-1, 1]$.
2) We propose and evaluate a method for partitioning signed networks for community discovery, according to a generalized partition optimization criterion. Our method is based on the framework of simulated annealing. We show that our algorithm tends to identify group structure even when the network is noisy and unbalanced. We apply our method for the analysis of Wikipedia interaction communities and show that it can be used to provide valuable conclusions.
3) We provide an in-depth analysis of our findings on the Wikipedia dataset. We analyze and interpret the differences in community structure based on the evolution of both controversial and established content and relate these communities to the ideas they promote. We also investigate the evolution of an article over time in terms of its prevalent topics.

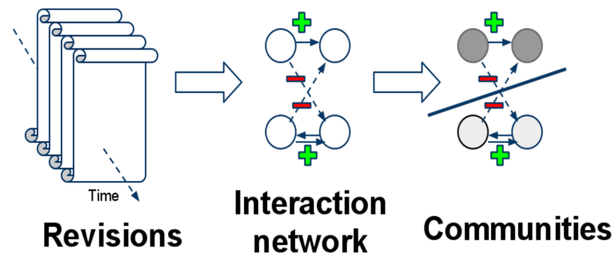


Fig. 1. Road-map of our approach. We infer interaction network from revision history and further detect communities in the latter.

II. Related Work

a) Inferring link strength and sign: Although social networks are typically homogeneous and unweighted, it has been shown that uncovering the true relationship strength can improve prediction models. Xiang et al. [2] introduce a statistical model based on the frequency of user interactions to determine edge weights that correspond to the strength of a social tie. A different approach aims at classifying the types of links between users by utilizing latent space models in order to summarize the raw features of nodes into semantically meaningful ties. One instance of the above family was proposed by Hoff et al. [9] who map users into an unobserved *social space* in which the probability of a link depends on the latent space distance between the adjacent users. While these methods address the issue of uncovering the link strength from interactions, they are tailored to node-similarity measures and thus cannot be trivially adopted to uncover the full spectrum of positive and negative relations.

Recently, signed networks have been considered to represent and analyze the multi-modal dynamics of communities, where links can be both positive and negative [12]. Specific networks of study include the voting network in Wikipedia, where users vote for or against the promotions for administrators [6]; systems that allow the creation of explicit friend or foe relationships such as the technology blog SlashDot [10]; and trust systems such as Epinions in which users collectively evaluate goods and services [8]. While the objective of Leskovec et al. [12] is to predict signed edges, their setup is different from ours in that we use non-relational data to infer edges, while they leverage explicitly predefined $(-1/1)$-edge networks and pursue prediction of new edges.

b) Community Discovery: Detecting communities in networks has traditionally been an active area of research [14]. Much of the previous work focuses on social networks, comprised of sign-homogeneous links [13] in which a partitioning algorithm maximizes the number of within community links.

Recently, two studies have addressed the problem of community detection in networks with both positive and negative links [11], [17]. Kunegis et al [11] generalize the traditional spectral clustering method to handle negative edges. This approach is limited to undirected graphs and thus it is not directly applicable to the directed networks we construct from Wikipedia. The second approach by Yang et al. [17] adopts a truncated random walk idea to compute proximity to a randomly chosen sink node. It proceeds by performing a sweep on the proximity vector to find an optimal partition. The first step of the approach uses solely the positive edges around the randomly chosen sink node, making the approach sensitive to the choice of sink and density of positive edges. Our experiments indicate that this technique is not effective in the presence of unbalanced communities, containing some positive edges between communities. Doreian and Mrvar [7] propose a greedy approach for clustering signed networks in which nodes from different clusters are randomly interchanged. If the global partition function is improved with the swap, then it is kept, otherwise it is thrown out and another swap is proposed. This procedure is then iterated until a local minimum is found.

Local graph partitioning is a different method for analysis of network communities. The goal of this type of analysis is to obtain the cluster containing a query node as opposed to computing a global partitioning of the whole network [14]. Local partitioning algorithms are usually more efficient and allow for node-centric analysis. A recent method by Andersen and colleagues [3] utilizes *Random walk with restarts (RWR)* stationary distributions to obtain a good local cut around a query node for local partitioning. This method is developed for positive-edge networks and it is shown, both theoretically and experimentally, to obtain good cuts in an efficient manner [3]. Due to space constraints, we leave an investigation of local partitioning algorithms for future work.

c) Edit networks from revisions: The work that is most similar to our user interaction model is a recent method by Brandes et al. [5] who construct a network based on article revisions. In this work, positive links result from newly added text and negative links are associated with content *reversion*. Our model is different in that it is based on the topical content of the change as opposed to the difference in size.

III. Interactions from revision history

The objective of the first step in our method is to extract contrasting or aligned viewpoints of editors based on the states of the text before and after a revision. We use the inherent causal relation between consecutive edits of the same text to infer interactions. Our basic assumption is that subsequent revisions are triggered by their predecessors due to either incompatible opinions or conversely due to a reaction of approval and desire to extend the line of reasoning. The assumption of causal relationship between subsequent revisions was also adopted in a recent study by Brandes and colleagues [5], although interactions in the latter are measured only quantitatively, based on amount of changed text, while our model takes into account the content of the change.

Evidence for the causal relationship of consequent revisions can be drawn from the co-localization within the article of the modified text of consecutive edits. Roughly 50% of the pairs of consecutive revisions of a typical article modify the same paragraph. This number is slightly higher for controversial articles, such as *Anarchism* and *The Gaza Strip* and slightly lower for featured (stable) articles such as *Microsoft*. The inter-arrival time for such "reactive" pairs of edits is significantly lower than on average. The rest of the revision pairs constitute adding new content in the form of new paragraphs, deletion

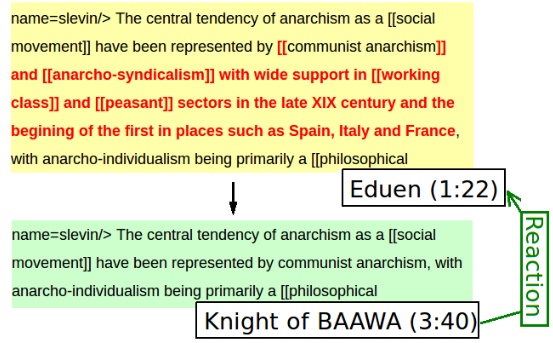


Fig. 2. A sample revision of the Wikipedia *Anarchism* article. Editor *Knight of BAAWA* removes the the previous contribution of *Eduen*. The removed text is in bold font in *Eduen*'s version.

of content without subsequent insertion or editing the same paragraph in later revisions. These statistics corroborate the hypothesis that a significant part of the collaborative content generation process is driven by reaction to co-authors' revisions as opposed to independent interaction with the text, disregarding other contributors' actions.

We expect that the proper minimal content block, within which to track interactions, will vary with the type of analyzed collaboration system and the process of collaboration. For example, consider the objective of extracting interactions between software developers using a collaborative software development system such as *Github*. The proper unit of interaction for such task will be a program class or function as these will represent the logical units within which changes are being made. In our analysis of text content in Wikipedia, paragraphs provide natural boundaries for separate ideas. As we discussed above, paragraphs also constitute the resolution at which interactions are manifested in the revision process of Wikipedia. Thus, we chose the paragraph as the main unit of content when extracting contributor interactions.

The example presented in Fig. 2 corresponds to an actual revision of a paragraph in the article *Anarchism*. The revision of contributor *Knight of BAAWA* removes the definition on *"anarcho-syndicalism"*, previously added by *Eduen*, from the Anarchism definition. This edit constitutes a reaction of *Knight of BAAWA* to *Eduen*'s view as to what content should describe Anarchism.

The above observations reveal two major challenges that we need to address in order to define an interaction between *Knight of BAAWA* and *Eduen* based on the content difference of their consecutive revisions. The first challenge is how to represent revisions in order to capture general topics such as "Western Europe syndicalism" and "working class". The second challenge lies in quantifying the level of agreement between consecutive revisions and transferring agreement of text to agreement between authors. We adopt *latent topic modeling* to address the first challenge and then measure the directions of topic shift to quantify their level of agreement. These two steps are explained in detail next.

d) Topic models: To extract meaningful features from the raw text of an article revision, we employ semantic modelling based on *Latent Dirichlet Allocation (LDA)* [4].

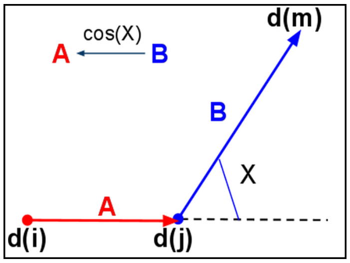


Fig. 3. Scoring the interaction between users B and A based on two consecutive revisions moving the same paragraph between states $d(i) \rightarrow d(j)$ and $d(j) \rightarrow d(m)$ respectively. The interaction score is computed as the cosine of the angle between the two edit vectors.

This allows us to make comparisons between user edits in the more manageable space of latent topics rather than a high dimensional term frequency space.

Choosing the number of latent topics in an LDA model is not a trivial task for a general text corpus. If the number of topics is too low there is possibility of mixing latent topics that would otherwise be well separated. Setting the number of topics too high may lead to redundancy in the representation. For our corpus of Wikipedia edits, we set the number of topics empirically by looking at the level of correlation of the top terms for different instantiations of the model topics. We aimed to work with topics that map to significant sub-areas within the articles, while keeping the overall correlation between the topics relatively low. For all of our experiments on the articles of *Anarchism* and *Microsoft*, we set the number of topics to 30.

We compute term frequencies for the paragraphs of each revision and use those to train the LDA model. In LDA terminology, the extracted paragraphs map to "documents", and the set of all documents comprise the training "corpus". A latent topic is represented as a distribution over terms. Assuming that L is the set of latent topics, we obtain a latent-topics representation of a paragraph by multiplying its term frequency vector $W \times 1$ from the left by the $W \times |L|$ topics matrix, resulting in a $|L| \times 1$ dimensional topic representation of the paragraph denoted $d^{|L|}$. We then use these topic vector $d^{|L|}$ to obtain interactions between contributors.

e) Positive and negative interactions from revisions: We determine the mode of interaction (positive or negative) based on the evolution of the latent topics within the content, undergoing consecutive revisions. Back to our example in Figure 2, in order to detect and assign weight to the interaction between *Knight of BAAWA* and *Eduen*, we compute the shift of topics *Knight of BAAWA*'s revision produces and compare it with the effect of *Eduen*'s revision. If the first revision also decreased the presence of topics like *Syndicalism* and *Western Europe*, then we assign a positive interaction to the directed pair. If, on the other hand, the preceding edit strengthens different topics then the follow-up interaction is negative. The magnitude of interactions is determined by the amount of shift between topics.

We next define formally the procedure of interaction computation. Every state of a text unit is represented as a topic vector $d^{(|L|)}$ in the latent space. We compute the topic-wise shift as the vector difference between two consecutive text states:

$$e(i,j) = d(i) - d(j), \tag{1}$$

where e is termed an edit vector and i and $j, i < j$, are revision indices ordered chronologically. Note, that the states i and j do not need to be immediately adjacent revisions. For a given paragraph state $d(i)$, there is a unique next state in the first subsequent revision j in which the same paragraph has been changed. The next state j can be in the next revision, which we observe in roughly half of the revisions in a typical Wikipedia article, but can occur several revisions later as well.

Given two consecutive edit vectors $e(i,j)$ and $e(j,m)$, we measure their agreement as the cosine of the angle between them:

$$s(m,j) = \cos(e(i,j), e(j,m)) = \frac{e(i,j) \cdot e(j,m)}{||e(i,j)||\ ||e(j,m)||}. \tag{2}$$

This measure is widely adopted in the text retrieval community and is referred to as *Cosine similarity*. The score $s(m,j)$ can take values in the range $[-1,1]$, where -1 corresponds to reverting the previous edit (i.e. complete disagreement) and 1 corresponds to reinforcing exactly the same set of topics as the previous user (i.e. maximal agreement). Scores in-between the extreme cases can be generated in multiple ways and thus this measure allows us to capture the *degree* of approval between two users over a set of latent ideas.

We use the score $s(m,j)$ as a measure of the interaction between the users responsible for the corresponding revisions. The scoring process is illustrated pictorially in Figure 3. A text entity is moved from state $d(i)$ to state $d(j)$ by user A, and then to state $d(m)$ by user B. As a result, we derive an interaction from user B to A encoding the feedback of A's edit. Note that interaction links are directed, thus for each interaction we only capture an approval score from the second user to the first (from B to A, not from A to B) which results in a directed network in the general case.

IV. From Pairwise Interaction Detection to Global Interaction Networks

In the previous section, we proposed a method to detect signed edges by scoring local interactions between pairs of nodes. Next, we analyze properties of the interactions derived from the revisions of two Wikipedia articles: *Gaza Strip* and *Microsoft* and their corresponding interaction networks. We focus on these particular pages as one of them (Microsoft) is declared a featured (high-quality) article by the Wikipedia community, while the other (Gaza Strip) is highly controversial. In this section we elucidate what a typical contributor interaction network looks like and show that it shares features with well-studied social networks. We study the existence of "heavy" contributors (corresponding to network hubs), and the distribution of scores between users with multiple interactions. General statistics for the networks are listed in Table I.

Figure 4 shows the distribution of interactions that we measure in the Microsoft and Gaza Strip revision histories. The number of exact "reverts" (score of -1) exceeds all other types of interaction. This indicates that the content which remains in

TABLE I
USER INTERACTION NETWORK STATISTICS FOR WIKIPEDIA ARTICLES

Network	Nodes	- Edges	+ Edges	Total
Gaza Strip	774	1,292	596	1,888
Microsoft	2,196	24,953	5,252	30,205

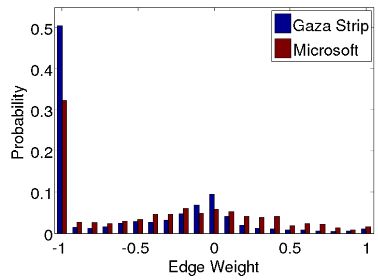
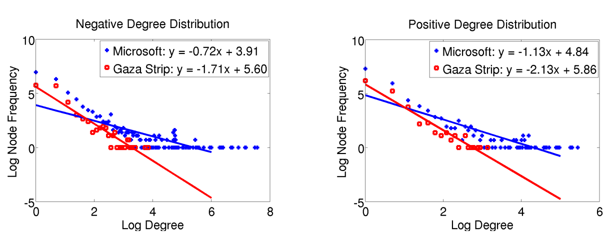


Fig. 4. Edge weight distribution in the Microsoft and Gaza Strip networks.

the article undergoes multiple rounds of insertion-discussion-deletion and re-insertion. We are able to verify such back-and-forth chain patterns qualitatively by analyzing user comments, available for some of the revisions of the articles.

At the same time, it is interesting to note that on average there is a higher probability of disagreement in Gaza Strip than in Microsoft. This alludes to the higher controversy associated with information related to the Gaza Strip as well as to the existence of strong opposing points of view of the corresponding editors. This is further confirmed by the experimental section presenting our community-based analysis.

To gain insight into the existence of contributors responsible for the majority of the article content, we analyze the degree distributions in the resulting interaction networks. Both networks fit exponential degree distributions of their positive (Figure 5(a)) and negative (Figure 5(b)) edges. The degree distribution for Microsoft has a steeper slope than that for the Gaza Strip. There are a handful of users in the Microsoft network responsible for the better part of the interactions. This reconciles with the fact that Microsoft is a featured article monitored by several users while the Gaza Strip has a much larger and more diverse set of active contributors.



(a) (b)

Fig. 5. Edge degree distributions for the Microsoft and Gaza Strip for (a) negative and (b) positive edges.

V. COMMUNITY DISCOVERY

Our goal is to discover ideology-coherent coalitions within the contributor base. These are groups of contributors who tend to promote some topic at the expense of others. We also seek to interpret and characterize the commonalities within coalitions and understand the opposing views between coalitions. We use the terms "coalition" and "community" interchangeably in the remainder of the presentation.

Coalition discovery in signed networks is a natural generalization of the problem of graph clustering for traditional positive-edge graphs. Coherent coalitions are expected to exhibit agreement among coalition members and disagreement with other coalitions. While there exist a plethora of methods targeted to positive-edge networks [14], there are only a few that can handle both positive and negative edges. Moreover the existing methods are either defined solely for undirected graphs [11], or are based on local greedy search [7] and heuristics [17] which work well for well-behaved networks but ineffective for unbalanced ones. Next, we propose and describe a general coalition discovery algorithm that works well in unbalanced (directed or undirected) signed networks.

We denote the graph corresponding to a signed network as $G(V, P, N, w)$, where V is the set of vertices, P is the set of positive edges, N is the set of negative edges and w is a weighting function that maps every edge to a real value $w : (i, j) \to \mathbb{R}$.

Global coalition discovery is the problem of partitioning the nodes V into k groups, such that both negative edges within group $N_{(i,j),i,j\in A}$ are minimized and the positive edges across groups $P_{(i,j),i\in A,j\notin A}$ are minimized. Note, that this optimization criterion was adopted in [17] and is a generalization of the desirable property of good clusters in positive-edge graphs, augmented by the requirement that negative edges are placed between coalitions. Formally, the generalization of the unnormalized cut [14] minimization for signed networks is defined as:

$$\min_{\mathbb{C}} \sum_{C\in\mathbb{C}} (1-\gamma) \sum_{\substack{i\in C\\ j\notin C}} w(i,j)P_{i,j} - \gamma \sum_{i,j\in C} w(i,j)N_{i,j}, \quad (3)$$

where $\gamma \in [0, 1]$ controls the balance of the penalty difference between putting a positive edge across and a negative within a coalition. In the extreme case ($\gamma = 1$), we disregard negative edges and are left with the standard formulation of the graph clustering criteria in which the goal is to maximize the within community edges. If $\gamma = 0$, i.e. focusing solely on negative edges, the problem is equivalent to the *Max-cut* problem. Both of the above extreme formulations are NP-complete problems. For our experiments, we set $\gamma = 0.5$, but this parameter is application specific and should be chosen according to the task at hand.

Our method of global coalition mapping is centered around a general framework for combinatorial optimization tailored to the specifics of our problem at hand. The concept of *Simulated annealing (SA)* has traditionally been applied to a number of combinatorial optimization problems. The basic idea behind SA is to define a statespace over all potential solutions for a specific problem (eg. all possible partitions of the nodes in a network) and perform a random walk over this space. The set of neighbors of a given state s includes any state s' that can be reached through a unit change to s. The transition probabilities are controlled by the state evaluation criteria and the so called *annealing schedule*. Initially, all transitions are uniformly distributed, and over time, preference is given to transitions that move to more optimal states. Since the solution space of typical combinatorial optimization problems is vast, controlling the transition probabilities allows a fine control of the trade-off between exploring the space and a strict hill-climbing optimization.

We adapt the general SA framework to our partitioning problem over real-value signed networks by defining the state space as the set of all possible partitions $\mathbb{C}$. A transition from $\mathbb{C} \to \mathbb{C}'$ is performed by randomly sampling a new partition label for a randomly selected node in the network. For the current state $\mathbb{C}$, and the proposed state $\mathbb{C}'$, the acceptance probability of moving from $\mathbb{C}$ to $\mathbb{C}'$ is defined in (3):

$$p^{(t)}(s'|s) = \begin{cases} e^{\frac{(E(\mathbb{C}')-E(\mathbb{C}))}{T(t)}} & \text{if } E(\mathbb{C}') < E(\mathbb{C}) \\ 1 & \text{o.w.} \end{cases} \quad (4)$$

where $T(t)$ is the annealing temperature for the t^{th} iteration of the algorithm and $E(\cdot)$ is the evaluation of the optimization criterion (3), also referred to as the *energy* of the current solution in the simulated annealing literature.

The annealing schedule is an important factor in the overall performance of the SA algorithm. If the temperature drops too rapidly to 0, the walk in the solution space is kept insufficiently narrow and is equivalent to a simple hill-climbing optimization approach. On the other hand, if the temperature remains at a high value for too long, then the algorithm does not make sufficient progress towards the optimal solution. For our experiments we chose a simple annealing schedule $T(i) = T(i-1) - \beta T(i-1)$ based on empirical evaluation. We found this annealing schedule to produce consistent convergence on controlled synthetic datasets. The obtained energy levels for our interaction networks are stable under this schedule as well. We set $\beta = 2e10^{-4}$ and an initial temperature of $T(0) = 100$; these values are used throughout the experimental section unless otherwise noted.

As mentioned, the annealing schedule is indeed an important parameter for our algorithm; however, we have found the algorithm is not particularly sensitive to the values of β and $T(0)$ for our datasets. We are able to achieve stable optimization results by assigning any *reasonable* values to these parameters. These observations are consistent with previous applications of SA.

In the literature on partitioning positive-value networks, there exist other criteria that optimize the balance of a partitioning in terms of nodes and in terms of edges. These criteria can be incorporated in SA at the level of local transition evaluation. Our analysis is focused solely on the unnormalized optimization criteria in eq. (4).

VI. Experimental Evaluation

All real world data evaluations are performed using the contents of the full revision history XML dump from January 2010, available at the Wikimedia download site[2]. We chose to use the pages of "Gaza Strip" and "Microsoft" for case studies in this work (statistics of these networks is provided in Table I). These pages were chosen to showcase networks with varying degrees of controversy. While the *Gaza Strip* article contains a fair amount of opinionated content, the page on *Microsoft* is recognized as a featured article in Wikipedia, i.e. it is considered objective and high-quality by the community.

[2]http://dumps.wikimedia.org/enwiki/20100130/

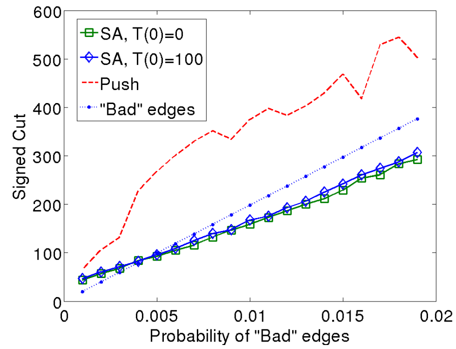


Fig. 6. Optimality of the signed cut for two clusters and increasing negative-within and positive-between edges

A. Comparison of coalition discovery algorithms

We evaluate the efficiency of our global discovery algorithm, referred to as SA, in comparison with the algorithm proposed by Yang and colleagues [17], referred to as *Push*. Push proceeds in two phases. First, it choses a random sink node and computes the probability of reaching all nodes from it while doing random walks of length up to l. For this first step it uses only the positive edges and computes the above quantities using an iterative algorithm. Second it sorts all nodes based on the random walk vector in the first step and finds an optimal cut by performing a sweep on this ordering while trying to ensure that partitions contain more positive edges than the inter-partition ones. *Push* is highly dependent on good connectivity through positive edges and according to our experiments is sensitive to the initial choice of the sink node. In addition, the criterion for splitting in the original paper is too optimistic, requiring a maximal value of 2 in cut detection vector. This value is never achieved in the synthetic and real networks we analyze. Hence, we augment *Push* to consider suboptimal partition values and sample randomly several sink nodes retaining the best clustering outcome.

Figure 6 presents an experimental comparison between *Push* and *SA*. For this experiment, we synthetically generate 2-cluster networks with 100 nodes in each. We generate random undirected edges according to four probabilistic parameters: p_i^+, p_i^-, p_o^+ and p_o^- denoting the probabilities of in-cluster positive, in-cluster negative, inter-cluster positive and inter-cluster negative edge probabilities. The probabilities are defined in terms of the number of all possible edge for the chosen nodes size. We adjust the generation parameters to match the corresponding values measured in clusterings of our real-world interaction networks. We vary the the probability of "Bad" edges (i.e. p_i^- and p_o^+) simultaneously. For every instantiation of the parameters, we generate 50 networks and average the results of the competing algorithms. We measure the value of the signed cut as defined by equation 3, where smaller values correspond to better partitioning.

SA clearly outperforms the *Push* at all "Bad"-edge regimes. The explanation for this behavior is that there are not enough positive edges within the clusters. As a result the produced ordering in the first step of *Push* is suboptimal and all retrieved cut results are suboptimal as well. We experiment with SA in two operational modes: one starts with a zero temperature (SA, T(0)=0) and thus does not allow for suboptimal solution state transitions (i.e. hill climbing). The second mode is under (trace SA, T(0)=100) uses a cooling schedule lowering the temperature from 100 to 0 thus allowing for suboptimal state transitions when the temperature is high. There is no significant difference in the optimal cut between these two regimes. The reason for this could be that the state space we are considering might be too well-behaved, rendering a mechanism for local minima avoidance, such as the temperature in S, unnecessary.

B. Mining the coalitions in Wikipedia

Next, we apply our global clustering algorithm to the networks synthesized from the Wikipedia content.

Deriving an exact criterion to choose the appropriate number of clusters k is a notoriously hard problem even in the case of positive-edge networks. Instead, we vary k from 2 to 10 and plot the energy horizons for the best cuts at each value of k. The energy measure combines, in addition to edges accounted for in the cut optimization criterion, the weights of positive edges within and negative edges between clusters. The smaller this value the better. We plot the energy horizon in Figure 7 and observe that the network energy improvement quickly tapers after just 3 clusters for both networks. We choose $k = 3$ for further analysis on these networks.

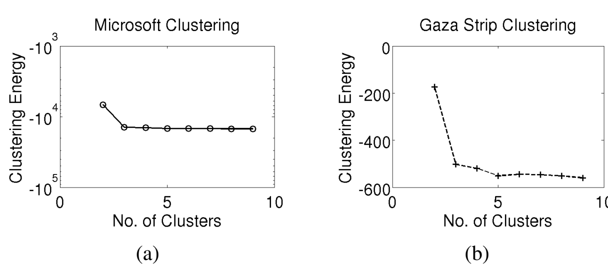


Fig. 7. Energy of clustered network for *Gaza Strip* and *Microsoft*.

We analyze the resulting clustering output by counting the number of positive and negative between cluster edges. This is a rather high level measure, as we do not consider the specific scores on the edges, only the sign, however, we are able to validate that SA is identifying 'good' communities as designated by our clustering criteria. We provide a matrix of between cluster positive and negative edges for Microsoft in table II, the table for the Gaza Strip network is similar and thus we do not show it here. We notice that the frequency of positive edges (the numerator) is significantly higher along the diagonal, as these represent within-cluster edges. These results are promising, signifying that even in real-world noisy networks we are able to uncover coherent groupings of nodes in a signed network.

TABLE II
POSITIVE/NEGATIVE EDGE COUNT BETWEEN CLUSTER FOR THE MICROSOFT NETWORK

	Cluster 1	**Cluster 2**	**Cluster 3**
Cluster 1	**630 / 1,709**	344 / 2,574	369 / 2,348
Cluster 2	1,275 / 5,508	**574 / 783**	576 / 4,406
Cluster 3	594 / 2,986	510 / 4,262	**381 / 377**

Next, we explore the relationship between communities and how they collectively affect the topics of an article. Figure 8 shows how the different communities interact in editing the subtopics for both the Microsoft and Gaza Strip networks. These figures were constructed by averaging the edit topics over all users in each cluster. To account for the variance within a cluster we perform two sample t-tests ($\alpha = 0.1$ significance level) between pairs of clusters for a given topic to determine which clusters are separated enough to claim "opposition".

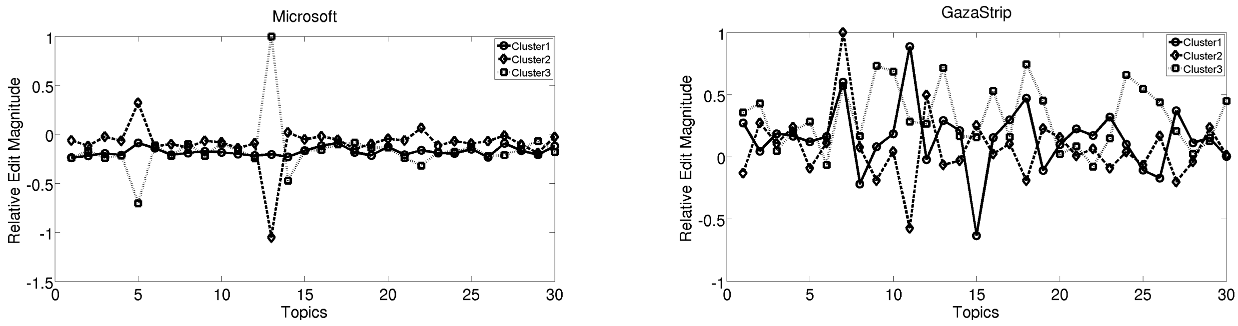
As a global comparison, we see that there are fewer sharp movements in the topics between the different clusters in the Microsoft network. This is expected as the Gaza Strip is a much more controversial article.

From Figure 8(a) and Table III, we see that the Microsoft network appears to have at least two communities which have opposing views. This is shown very clearly by the topics related to vandalism (5) and software products (13) in which cluster 2 and 3 push in opposite directions. Based on this opposition, we can infer that cluster 2 contains a group of vandals which are inserting anti-Microsoft text in place of content on Microsoft's software products. Additionally, users in cluster 3 are consistently cleaning up the vandalism in the article. There appears to be a divergence between these two clusters on other topics as well, though to a lesser extent. For instance, topic 14, which has negative connotation towards Microsoft, and topic 22 (Apple) is again promoted by cluster 2 and demoted by cluster 3.

Figure 8(b) and table IV shows the interactions between the different communities over the topics on the Gaza Strip page. Again, we see evidence of clusters editing in opposition to one another. Cluster 2 clearly shows an opposition to cluster 1 on topics dealing with terrorism (11) and opposition to the occupation of Gaza (12). Given the relationship between these two topics, we may infer that the users in cluster 2 are in support of the Palestinian people. That is, they are removing content portraying Palestinians as terrorists and adding content in clear opposition to the occupation of Gaza. Additionally, cluster 3 promotes topics on conflict resolution (9) and the geography of the Gaza Strip (25) in opposition to both cluster 1 and 2. While these communities are more difficult to evaluate holistically than in the case of the Microsoft network, a division of opinions on topic subsets can still be identified.

VII. CONCLUSION

We provide an end-to-end framework for analysis of community relations based on collaboration activities. We define a



(a) (b)

Fig. 8. Average topic shifts per community for the networks of Microsoft (a) and the Gaza Strip (b).

TABLE III
SELECTED TOPICS EXTRACTED FROM THE MICROSOFT NETWORK.

	Topic Label	**Top Words**
2	Software	linux, source, open, programs, run, desktop
5	Vandalism	ref, microsuck, ro$oft, winblows, offi
13	Software Products	microsoft, windows, office, news, installed
14	Anti-Microsoft	microzoff, windblows, image, overly, logo
22	Apple	apple, logo, slogan, started, intel, mac

TABLE IV
SELECTED TOPICS EXTRACTED FROM THE GAZA STRIP NETWORK.

	Topic Label	**Top Words**
9	Conflict Resolution	plan, agreements, withdrawal, settlers
11	Terrorism	terrorist, established, funds, renounce, violence
12	Occupation opposition	occupation, human, rights, pna (Palestinian National Authority), news
13	Palestinian State	sovereignty, internationally, status, recognized
15	External Access	access, population, foreigners, entry, tax
25	Geography	territories, west, east, borders, southwest, portion
29	Human Rights	united, human, rights, nations, law, humanitarian

novel interaction model based on content generation behavior. We are the first to propose a method for interaction network construction according to contributor-level topic promotion. The collaboration networks we derive are expressive enough to encoding the full gamut of human interaction modes: from absolute approval to complete rejection.

We employ the constructed networks to reveal coalitions within the contributor base, tied by strong positive links and opposed to each other via strong negative relations. We construct and evaluate a community discovery algorithm for signed directed networks based on simulated annealing. We then apply the algorithms to two article related contributor sets and analyze the behavioral cohesiveness of the obtained communities. Our analysis provides evidence of coalitions with opposing points of view in both networks.

Our study uniquely combines information and social aspects of Wikipedia in order to elucidate the implicit collaborative structure of this unique content generation ecosystem.

REFERENCES

[1] Five pillars of wikipedia. http://en.wikipedia.org/wiki/Five_pillars_of_Wikipedia.

[2] Modeling Relationship Strength in Online Social Networks. In *WWW*, pages 1–8, 2009.

[3] R. Andersen, S. Diego, L. Jolla, F. Chung, K. Lang, and S. Clara. Using PageRank to Locally Partition a Graph. *Science*, (Focs):1–23, 2006.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022, May 2003.

[5] U. Brandes, P. Kenis, J. Lerner, and D. van Raaij. Network analysis of collaboration structure in Wikipedia. *WWW*, page 731, 2009.

[6] M. Burke and R. Kraut. Mopping up: modeling wikipedia promotion decisions. In *CSCW '08: Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 27–36, New York, NY, USA, 2008. ACM.

[7] P. Doreian and A. Mrvar. A partitioning approach to structural balance. *Social networks*, 18(2):149–168, 1996.

[8] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th conference on World Wide Web - WWW '04*, page 403, New York, New York, USA, 2004. ACM Press.

[9] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, 2002.

[10] J. Kunegis, A. Lommatzsch, and C. Bauckhage. The Slashdot Zoo : Mining a Social Network with Negative Edges. In *WWW*, 2009.

[11] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. De Luca, and S. Albayrak. Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization. In *SDM*, 2010.

[12] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. *WWW*, pages 641–650, 2010.

[13] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Aug 2003.

[14] S. Schaeffer. Graph clustering. *Computer Science Review*, 1(1):27–64, Aug. 2007.

[15] J. Scott. Social Network Analysis. *Sociology The Journal Of The British Sociological Association*, 2010.

[16] H. T. Welser, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith. Finding social roles in Wikipedia. In *American Sociological Association*, pages 1–11, 2006.

[17] B. Yang, W. Cheung, and J. Liu. Community Mining from Signed Social Networks. *TKDE*, 19(10):1333–1348, Oct. 2007.