Predicting Trust Relations Among Users in a Social Network: The Role of Influence, Cohesion and Valence

Nikhita Vedula and Srinivasan Parthasarathy
Department of Computer Science and Engineering
Ohio State University
{vedula, srini}@cse.ohio-state.edu

Valerie L. Shalin
Department of Psychology and Kno.e.sis
Wright State University
valerie.shalin@wright.edu

ABSTRACT

Trust is a key concept in social networks, reflecting credibility and reliability for a multitude of participants and online data. Nevertheless, the majority of such networks lack explicit trust feedback. This motivates a mechanism to predict and manage trust relations automatically. We extract in an unsupervised manner local trust relationships between pairs of users from social networks derived from Twitter. We take into account factors of measurable influence between users, the impact of the structural topology of users in the network and the valence (sentiment) associated with the language-based information shared by network members. We evaluate our user trust rankings over other members of the network against a metric of ground truth for both social media data and a non-social media dataset, and analyze how the inclusion of valence lends robustness and stability to our model of trust. Knowledge of trustworthy citizens in social networks is quite advantageous in accurately assessing the credibility of the information they provide via social media, for the purpose of emergency response and recovery efforts during a disaster or a catastrophic event.

1. INTRODUCTION

Trust is a vital social construct. The economist Kenneth Arrow defines trust as "a lubricant of the social system" [19]. The trust construct draws attention from multiple areas of research, including sociology, psychology, management, economics and political science, and more recently, computer science. Castelfranchi and Falcone [9] analyze trust in the context of multi-agent systems. They characterize the components that comprise trust, explain the relation between trust and the act of delegation, and describe how trust relates to prior experience. With the rise of Web 2.0 technologies, trust emerges as a key concept in social network and social media analysis, reflecting credibility and reliability for the multitude of participants and available online data.

Advances in information technology dramatically increase reliance on wireless technologies, including social networking tools (e.g. Facebook and Twitter) and give rise to the notion of *citizen sensing*. Citizen sensing (conversations among users, and more broadly, organizations) appears in CNN iReports, Twitter posts in disaster response, Facebook based stress maps, and Instagram tags, for example for asthma triggers. Clearly, the organizational effectiveness and utility of such sensing requires more than the presence of technology (see Rochlin [27]). Leverage depends on the filtering, integration, communication and distribution of *trustworthy* information. Emergency response provides an ideal domain for examining the interaction between organizational forms with such technology, enhanced by information filtering, integration and distribution. Governmental agencies distribute recommendations and guidance to promote citizen response. Such governmental messaging can have an enormous influence on the eventual societal impact (cost, recovery) of a disaster. Victims and their neighbors share timely information (e.g., flood level, road blockages) and offer resources (e.g., vehicles, food, and supplies) on social media. Citizen reports can lead to the prioritization of relief efforts ranging from critical infrastructure repair to saving lives in areas most affected by damage.

Currency and reach make social media sites such as Twitter an attractive resource for capturing public activity during emergencies. However, message recipients must trust the source to provide reliable information. Finding a trust ranking of users in social networks can promote reliance on trustworthy citizens to improve disaster response and recovery efforts [33]. A fundamental challenge to the response agency is rapidly vetting and separating the noise from the informative signal. A Twitter stream with only keyword/tag based filtering generates noise (irrelevant distracting information) in relation to signal, due to inaccurate information from unreliable sources (misinformation) or ambiguities from reliable sources (i.e., channel noise). Among the many elements to the "trust" equation, we primarily focus here on trust among users (as opposed to trust in *facts*, e.g. road closures). We only consider information that can be obtained or inferred from a network, and do not take into account any detail regarding background or previous history of entities. This allows language based social-sensed data extracted from appropriate users on social networks to be combined with or enhance data and/or predictions from physical sensors during an emergency response situation.

We predict local trust relationships between pairs of users

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

in social networks, using both structural properties of the network and information content posted by members of these networks. Though user trust relationships are extracted in an unsupervised manner, evaluation occurs in a supervised setting against ground truth obtained by leveraging other resources of trust rankings. We subject the underlying model to various stress tests to see how robust the model is with respect to background assumptions and the data itself. We identify which aspects of the data are less trustworthy than others, i.e. more likely to be falsified with additional data or change in domain assumptions. We also identify the trustworthiness of the entities themselves belonging to the network. A key contribution of our work is a robust method that takes into account the factors of influence (as a proxy for how much one user trusts another), structural identity (or homophily) within a network (related to the concept of social identity theory), and valence (shared sentiment related to the concept being sensed). In particular we show that taking into account valence lends itself to a robust, and stable model of trust in such an analysis.

2. METHOD

Informally, we seek to infer a model of trust among users within a network and to also identify highly trusted users or organizations within a social network. Specifically we focus on *Twitter* as the social media network of choice [1].

Problem Statement: Formally, we assume the network of interest is represented as a bipartite graph $G = (U \cup T, E)$, where U and T are two disjoint sets representing the set of members or users in the network and the set of clusters of tweets (topics) respectively; and E is the set of edges or interactions between the users and tweets. Tweets cluster based on their textual and contextual similarity; two tweets talking about the same or similar topics are placed in the same cluster. We assume that topic clusters (T) are pre-defined from the domain or result from standard topic model algorithms [4, 36]. The trust prediction algorithm aims to predict a set of edges T_e weighted by trust that link individual users. Users can be rank ordered by their trustworthiness (based on number of users that trust them and weights on trusted edges incident to them).

Our solution desiderata include the twin goals of efficacy and efficiency: Efficacy in effectively identifying and ranking trustworthy users and efficiency in being able to compute these in emergent situations (e.g. during or shortly after a disaster). We discuss below three elements, informed by social theory, that we believe play a role in developing trust between two users in a social network, and we measure these elements to determine a trust metric among users.

2.1 Influence

Leading sociologists note that trust is integral to the concept of social influence. A messenger more easily influences or persuades a recipient to do something or react in a certain manner if that recipient trusts the original messenger. While trust itself is difficult to measure outside of specific paradigms (for instance Berg's trust game [3], which addresses this problem from the perspective of behavioral economics), several researchers have tackled the problem of detecting *influential* users within a social network. Here we explore influence as a proxy measure for pairwise trust relationships. We expect a strong correlation between the trust of a user x on a user y and the observed influence of user y on user x since intuitively, users tend to trust the users who influence them to retweet a particular tweet on a particular topic. Influence can often be measured as a function of the structure of the network (e.g. page-rank style [25]), the dynamics of the network interaction [7], the frequency with which the users' tweets are retweeted (while accounting for the passivity of users and based on prior content history of the users' tweets [28]) or by taking into account the local neighborhood of the tweeter via viewpoint analysis [2].

While we might leverage any of the above approaches, we selected a scalable approach [37] based on a simple linear algebraic kernel and iterative sparse vector multiplication. This algorithm also accounts for user passivity (i.e. the likelihood of a user reacting to a messenger). Influence is computed using the Influence-Passivity(IP) algorithm described in [28]. The resulting HITS-style algorithm [20] calculates a global influence and passivity score for each user in the following manner.

We first construct a weighted, directed, unipartite graph $H = (U, E, W)$ consisting of all the users in our bipartite graph, joined by edges E and edge weights W. Edge (i, j) exists between user i and user j if user j retweeted a tweet or URL posted by user i at least once. Weight w_{ij} on edge $e = (i, j)$ represents the ratio of influence that i exerts on j to the total influence i attempted to exert on j. It is expressed as $w_e = \frac{S_{ij}}{Q_{ij}}$ where Q_i is the number of posts or URLs that i mentioned and S_{ij} is the number of URLs mentioned by i and retweeted by j.

The influence function $I_i : N \rightarrow [0, 1]$ that represents node i's influence on the network is calculated as:

$$Infl_i = \sum_{j:(i,j)\epsilon E} u_{ij} Passiv_j$$
$$Passiv_i = \sum_{j:(j,i)\epsilon E} v_{ji} Infl_j$$

Here u_{ij} represents the amount of influence user j accepted from user i normalized by the total influence j accepted from all users in the network. v_{ji} represents the influence that user i rejected from j normalized by the total influence rejected from j by all users in the network.

2.2 Social Cohesion and Valence

Social cohesiveness appears in fields such as Sociology, Psychology and Public Health. We follow the common definition from psychology[26]: cohesiveness relates to the members of a group who share emotional and behavioral characteristics with one another and the group as a whole (see Lott and Lott [22] on group cohesiveness as a function of interpersonal attraction). In the context of our problem and inspired by the strong correlation between user similarity and trust as established in [13] and [38], we incorporate two types of pairwise similarity measures between users in the trust calculation. The first, coarse-grained measure computes the similarity between a pair of users by only taking into account the structural topology of the bipartite network. A simple approach based on Jaccard similarity is sufficient, such that two users are highly similar if they are connected via a similar set of topic clusters. A scalable variant to estimate this similarity uses the notion of minwise independent hashing or more generally locality sensitive hashing [12, 6, 30, 5].

We associate with each user x a *vector* V_x of tweet-cluster ids to which x has a directed edge. The Jaccard similarity between users x and y is then:

$$Jacc(x,y) = \frac{V_x \cap V_y}{V_x \cup V_y}$$

However, this metric largely reflects structural cohesion. It does not account for the popularity of certain topics, nor (as suggested by Lott and Lott [22]) does it account for interpersonal emotional agreement among users. To overcome this limitation we propose a more nuanced measure that takes into account the relative *popularity* of the tweet clusters or topics, the valence or the sentiment of a user with respect to a particular topic and the agreement among them. User sentiment on a particular topic as expressed by the content of their posts can reflect trust between a pair of users. A user x likely has greater trust in a user y who shares x's sentiment or opinion on a particular event or topic.

We define a *shared* tweet cluster or topic between two users in U as any tweet cluster in T to which both users have a directed edge, and for which both users have the same sentiment associated (positive, negative or neutral). We distinguish between shared tweets according to the relative popularity of the central topic, which we associate with the in-degree of each shared tweet cluster. Intuitively, the in-degree of a tweet cluster or a topic (the number of users talking about it) increases, the extent to which we can infer about the relative similarity of two users connected to that topic by a directed edge reduces. Two users who post on a rare topic are more likely to have an affinity towards each other and therefore, a greater trust relationship relative to a pair of users who both post on an popular topic. Thus, we want a highly connected topic common to two users to increase the similarity between them by less than a less connected shared topic. The following function ensures this.

$$P(t) = ce^{-\frac{indeg(t)^p}{2}}$$

$$sim(x,y) = \frac{\sum_{i=1}^{T} P(t)}{T}$$

Here T represents the number of shared topics common to users x and y, $deg(t)$ represents the in-degree of the common topic t, and p $(0 < p < 1)$ is a constant parameter denoting topic popularity. The value of p closer to 1 signifies that a topic is considered popular at a comparatively smaller in-degree, say $indeg(t) = 10$ for $p = 0.8$. p closer to 0 signifies that a topic needs a large number of incoming user-edges to be popular. The value of this parameter depends on the dataset. $P(t)$ represents the impact of a topic common to a pair of users on their similarity. The values of $P(t)$ are normalized to lie in the range [0,1]. $P(t)$ for a common topic t (and by consequence its effect on the similarity between the appropriate pair of users) decays exponentially to a low value (close to 0), as the number of incoming user-edges to t increases. Figure 1 illustrates this behavior.

An alternative method to account for topic popularity is based on the idea of *degree discounting*, as defined in [29]. It is based on the following insights:

1. When two users x and y both point to a topic t and share the same valence or sentiment about t, the contribution of t to the similarity between x and y is inversely related to the in-degree of t.

2. The out-link similarity between two users x and y is inversely related to the out-degrees of x and y.

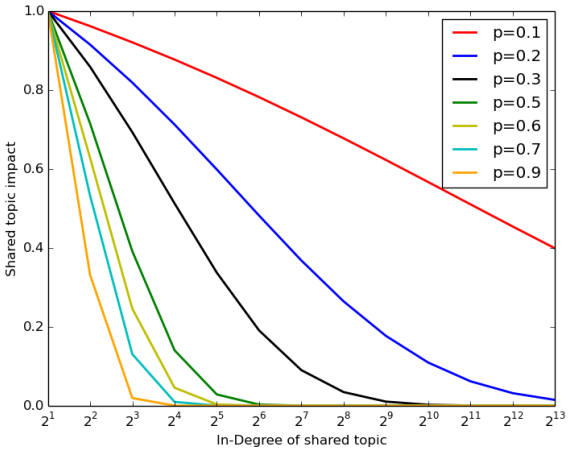

Figure 1: Popularity behavior of shared topics: p closer to 1 signifies that a topic is considered popular at a comparatively smaller in-degree, while p closer to 0 requires a topic to have a large number of incoming user-edges to be considered popular. The impact $P(t)$ of a shared topic on pairwise user similarity reduces as the shared topic's popularity increases.

We define the degree-discounted out-link similarity sim_d between two users x and y. D_o is the diagonal matrix of out-degrees of users and D_i is the diagonal matrix of in-degrees of topics in the bipartite graph. A is the adjacency matrix of the bipartite graph. t represents a shared topic or tweet cluster between users x and y, so that both x and y are connected to t in the bipartite graph and *share* the sentiment or valence regarding t i.e. positive, negative or neutral.

$$sim_d(x,y) = \frac{1}{\sqrt{D_o(x,x)}\sqrt{D_o(y,y)}} \sum_t \frac{A(x,t)A(y,t)}{\sqrt{D_i(t,t)}}$$

2.3 Putting It All Together

After preprocessing each Twitter dataset, we extracted its prominent topics using a Non-Negative Matrix Factorization [36] based topic modeling approach. We grouped the tweets based on their central topic by analyzing the textual content of each tweet and constructed a bipartite network from each dataset. We then analyzed the overall sentiment of each user with respect to all tweet clusters (topics) they are connected to in the network by matching the cluster lexical content with an opinion lexicon of positive and negative sentiment oriented words, obtained from [21]. The topic popularity parameter p for tweet cluster nodes reflected a dataset specific threshold for the number of incoming edges.

To combine the measures, we note that the influence values and similarity measures follow a Gaussian distribution. Thus, we can normalize them using *z-scores*. We therefore obtain a regularized formula for calculating trust as follows:

$$Trust(x,y) = \alpha Infl(y,x) + \beta Jacc(x,y) + \gamma sim(x,y)$$
or,
$$Trust(x,y) = \alpha Infl(y,x) + \beta Jacc(x,y) + \gamma sim_d(x,y)$$

such that $\alpha + \beta + \gamma = 1$.

α, β and γ are regularization parameters representing the factor weights and require experimental tuning.

3. EVALUATION

Our experimental evaluation examines: i) Which among the three factors (influence, cohesion, valence) are most important to estimate trust relationships among users in a social network and ii) Method robustness with and without valence.

3.1 Datasets and Ground Truth

We employed Twitter data relating to three domains during the year 2011: a political dataset (related to the 2011 Anti-Corruption Movement in India), a disaster dataset (related to the July 2011 Terrorist Attack in Mumbai, India) and a product dataset (tweets related to Phones and Tablets). Ruan et al [26] collected the data using a Twitter Streaming API-based crawler for a search of appropriate keywords. We also tested our approach on a non-social media dataset, namely the Film DVD dataset collected by Guo et al [17] in December 2013 by crawling 17 categories of film DVDs from the dvd.ciao.co.uk website. This provides ground truth information as a list of pairs of users who trust each other. Due to the absence of textual information for this dataset, we used the provided genre of each movie as its topic. Additionally, we used user provided ratings to estimate the sentiment towards a particular movie. A user rating of 4 or more (out of 5) was taken as the expression of positive sentiment towards the movie, 2 or less was taken as negative sentiment and a rating of 3 was assumed to be neutral. Tables 1 and 2 provide dataset details.

Dataset	Users	Topic based Tweet clusters	Nodes	Edges	Tweets
India Corruption	2104	15	2119	7180	100K
Mumbai Blast	581	10	591	932	10K
Phone and Tablet	9939	15	9954	16265	100K
CiaoDVD	4658	17	4675	18561	-

Table 1: Twitter data and bipartite graph statistics of each dataset

While ground truth for each pairwise trust relationship is near nigh impossible to obtain without a dedicated social survey instrument, we can leverage existing, independent domain knowledge for determining the top-ranked trustworthy users for each dataset. The top users in each dataset who enjoy the highest trust among most of the other users of the network appear in Table 3. Most are well known personalities in fields such as journalism, social work or entertainment and are thus likely to be more trusted over other users. Certain users aren't well known in any fields, yet commanded a high trust value during that period of time. For example in Table 3, the Twitter user *ashwinsid* had voluntarily provided transport to people stranded during the Mumbai blast period (as noticed from his tweet log), which increased his trust level on topics related to the blast though he wasn't as popular as the other highly trusted users. In case of the Phone/Tablet dataset, a number of users enjoying a high trust among people are widely recognized authors of fiction such as Richard C.Hale, R.C. McCracken and P.T. Mayes, because many of their books are available as e-readers.

Dataset	Timeline
India Corruption	Thursday 24th November, 2011 to Tuesday 29th November, 2011: Protests against political corruption https://en.wikipedia.org/wiki/2011_Indian_anti-corruption_movement
Mumbai Blast	Wednesday 13th July, 2011: A series of three coordinated bombings https://en.wikipedia.org/wiki/2011_Mumbai_bombings
Phone and Tablet	During the day of Monday 15th April, 2013
CiaoDVD	Sunday 1st December, 2013 to Tuesday 31st December, 2013: Movie rating and movie review rating based dataset crawled from movie DVDs of varied genres. dvd.ciao.co.uk

Table 2: Timeline displaying the beginning and end dates of each dataset

Dataset	Most Commonly Trusted Users
India Corruption	BreakingNews, PRSLegislative, BBCWorld, shekharkapur, ndtv, rameshsrivats, swaroopch, fakingnews, AnupamPKher, SachinKalbag
Mumbai Blast	ndtv, AnandWrites, KiranKS, Netra, RamCNN, htTweets, ashwinsid, mid_day, dina
Phone and Tablet	Richard_C_Hale, RCMcCracken, AuthorNetwork, Androidheadline, engadgetmobile, TalkAndroid, engadget, ptmayes, androidcentral, AndroidAuth

Table 3: Top trusted users for each dataset

3.2 Factor Impact and Analysis

The three parameters α, β and γ required to compute a pairwise trust value were chosen experimentally based on an exhaustive grid search (which we shall discuss shortly). We leveraged independent domain content to generate a trust-based ranking of the top trusted users and used it as ground truth to validate the algorithm.

Tables 4, 5 and 6 show the tuned parameter values that best matched the ground truth table of top users (see Table 3). Influence appears to be the strongest factor contributing to an accurate list (α values typically range between 0.6 and 0.7) . Valence (conditioned by the popularity of the topics) common to each pair of users has γ values between 0.25 and 0.30 for two of the datasets. Finally, structural cohesion had a non-trivial but muted role with β values up to 0.35 for the Phone and Tablet dataset. Structural cohesion contributes the least for the Corruption and the Blast datasets. But for the Phone/Tablet dataset, the structural cohesion (based on bipartite network topology) contributed more than valence, conditioned by popularity. Next we detail the performance of our method on a trust dataset having ground truth available, and address method sensitivity to the above defined parametric settings.

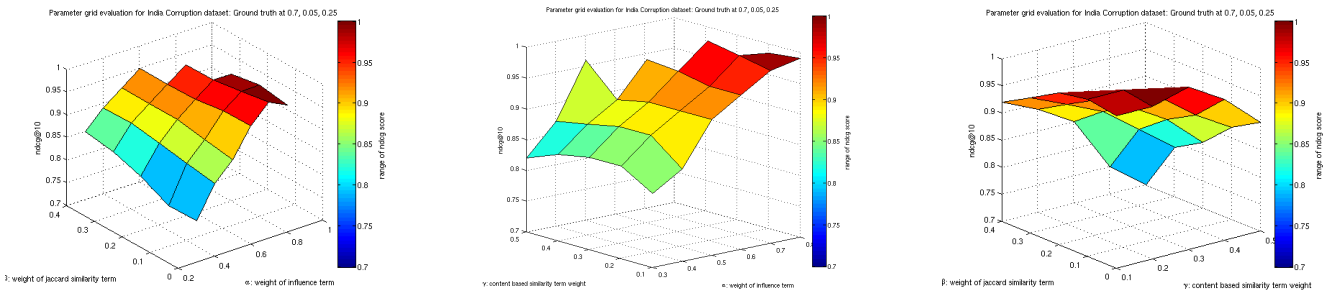

Figure 2: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, without using degree discounted similarity (India Corruption dataset)

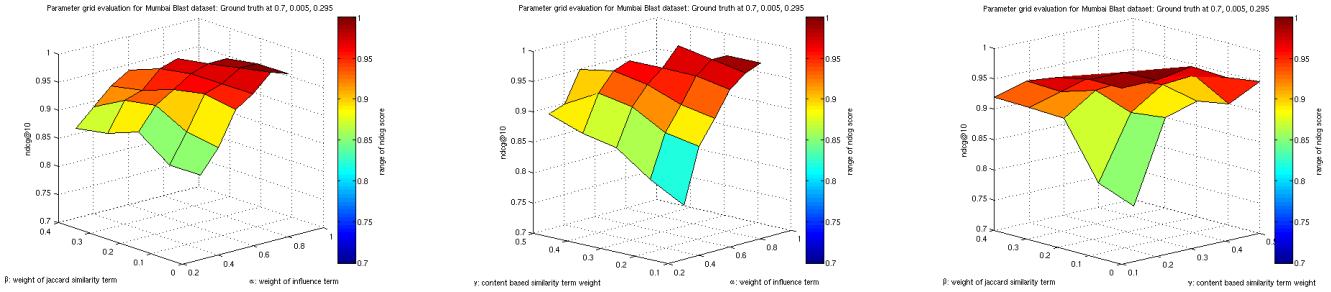

Figure 3: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, without using degree discounted similarity (Mumbai Blast dataset)

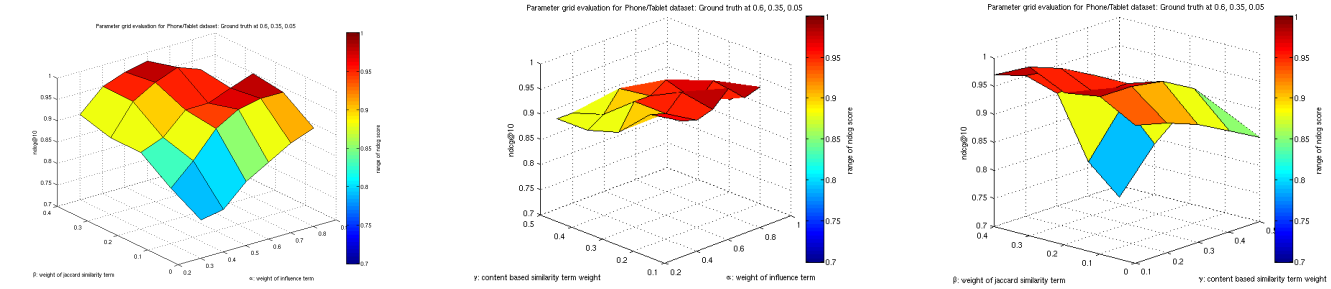

Figure 4: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, without using degree discounted similarity (Phone/Tablet dataset)

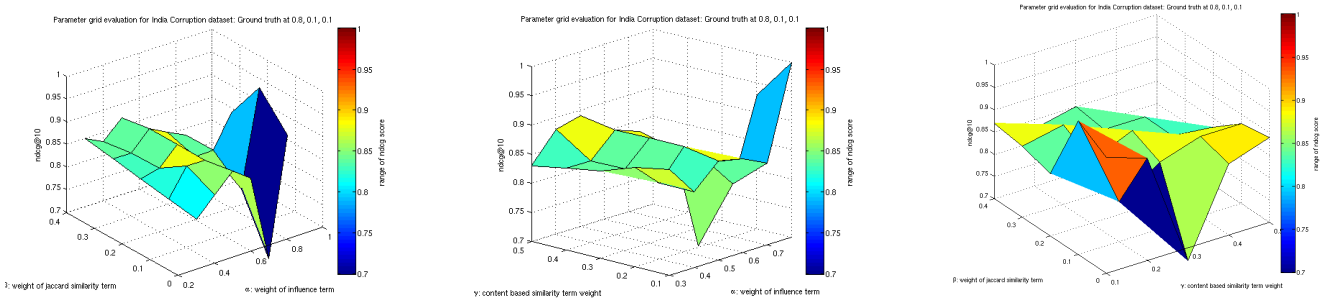

Figure 5: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity (India Corruption dataset)

Dataset	p	α	β	γ
India Corruption	0.3	0.7	0.05	0.25
Mumbai Blast	0.4	0.7	0.005	0.295
Phone and Tablet	0.2	0.6	0.35	0.05

Table 4: Selected parameter values for ground truth, without using degree discounted user similarity

Dataset	α	β	γ
India Corruption	0.8	0.1	0.1
Mumbai Blast	0.65	0.05	0.3
Phone and Tablet	0.65	0.3	0.05

Table 5: Selected parameter values for ground truth, using degree discounted user similarity

Dataset	α	β	γ
India Corruption	0.6	0.05	0.35
Mumbai Blast	0.65	0.05	0.3
Phone and Tablet	0.7	0.1	0.2
CiaoDVD	0.7	0.1	0.2

Table 6: Selected parameter values for ground truth, using degree discounted user similarity and sentiment

3.3 Evaluation on CiaoDVD Film Dataset

We begin with evaluating how the pairwise trust values we have designated as ground truth using a particular set of values for the parameters $\alpha = 0.7$, $\beta = 0.1$ and $\gamma = 0.2$ (from Table 6) compare to the existing ground truth for the CiaoDVD Film dataset, using sentiment-based degree discounted out-link similarity to account for the impact of the popularity of shared topics on trust between a pair of users, as explained earlier. This provides an independent assessment of our method on a dataset with factual ground truth. Because the pairwise trust values are independent, categorical variables, to relate the above mentioned measures of ground truth we first performed a chi-squared test of independence and then calculated the Cramer's V measure of correlation. We obtained a reasonable correlation value of 0.5615. Next, we created a global ranking of users for both our designated ground truth and the actual one, consisting of the users who appeared to be highly trusted by most users of the dataset, and examined the agreement between the two rankings by making use of the metric of Normalized Discounted Cumulative Gain (NDCG). We found an agreement of 0.955 between the 2 rankings, which strongly validates our method and parametric settings.

3.4 Parametric Tuning and Impact of Valence

We next turn to social media data (the focus of this work), to compare our results with domain knowledge designated ground truth. As we only computed pairwise trust values per pair of users, we generated a global ranking of users in the network, consisting of the users who appeared high in the trust ranking of most users of the dataset. We then evaluated the ranking obtained from each parametric setting against the ground truth. Figures 2, 3 and 4 illustrate the NDCG scores of these varied rank orders against the ground truth ranking for each dataset, without using degree discounted similarity, and Figures 5, 6 and 7 illustrate the same using degree discounted out-link similarity to account for the impact of the popularity of shared topics on trust between a pair of users. Figures 8, 9, 10 and 11 show the impact of sentiment on the NDCG score for the 3 Twitter datasets and the CiaoDVD dataset. Each figure contains 3 surface heat plots, showing the variation in the NDCG score against values of a pair of trust parameters. The first series of plots displays the NDCG score on the z-axis at different values of the pair of parameters α and β on the x and y axes respectively. The second series of plots shows the NDCG score at different values of the pair α and γ on the x and y axes, and the third series of plots displays the NDCG score on the z-axis at different values of the pair of parameters γ and β on the x and y axes.

A high weight on the influence term (> 0.5) followed by a second highest weight on the topic popularity impact term (in the range $0.1 - 0.3$) gives the highest NDCG score for the Corruption, Blast and CiaoDVD datasets, and lowest weight to the structural similarity term (typically < 0.1). For the Phone/Tablet dataset, the structural similarity contributes slightly more to the trust value than the impact of shared topic popularity.

The addition of sentiment or valence to the trust computation is significant. Without sentiment or valence, the NDCG scores range from $0.7 - 1$, whereas their inclusion leads to a rise in the NDCG scores to the range of $0.95 - 1$. Thus, the incorporation of sentiment causes the inference of trust to become robust to changes in the parametric settings.

Finally, for our Twitter datasets, we explored the correlation between our trust function and an independent metric derived from a user's Twitter profile, namely whether a user's Twitter account is 'verified' or not. Since only the authentic accounts of prominent Twitter users enjoy a 'verified' status, and our algorithm detects many such users as highly trustworthy, we expect to see a high correlation. Table 7 shows the results of a chi-squared test and a subsequent Cramer's V metric.

Dataset	Correlation between our trust function and 'verified' status of a user
India Corruption	0.6
Mumbai Blast	0.26
Phone and Tablet	0.15

Table 7: Selected parameter values for ground truth, using degree discounted user similarity and sentiment

For the politically oriented Corruption dataset, we note good agreement between trustworthy users (most of whom are key public figures) and verification the status of their Twitter accounts. However for the Blast and Phone and Tablet datasets we obtain a low correlation between these two metrics. Hence, although many users identified as highly trustworthy by our algorithm are quite active and popular on Twitter judging by the number of retweets, mentions and followers they possess, they lack a verified status on Twitter.

3.5 Discussion

We sought to answer the following two questions: i) Which among the three factors (influence, cohesion, valence) are most important to estimate trust relationships among users in a social network? ii) How robust is the method to estimate trust with and without valence?

As expected, influence serves a crucial purpose in the trust equation (at least in as much as detecting the top trustwor-

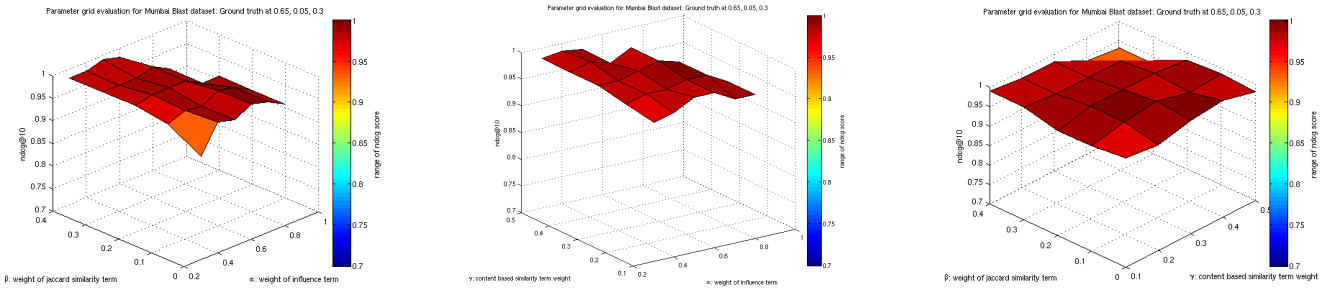

Figure 6: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity (Mumbai Blast dataset)

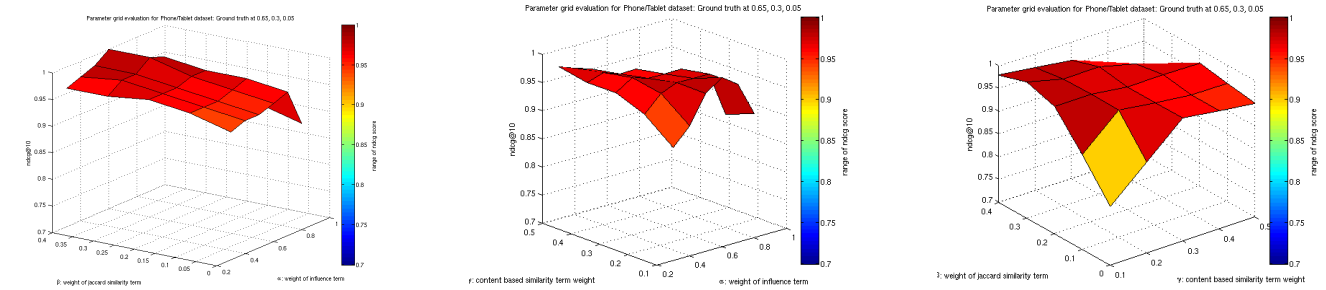

Figure 7: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity (Phone/Tablet dataset)

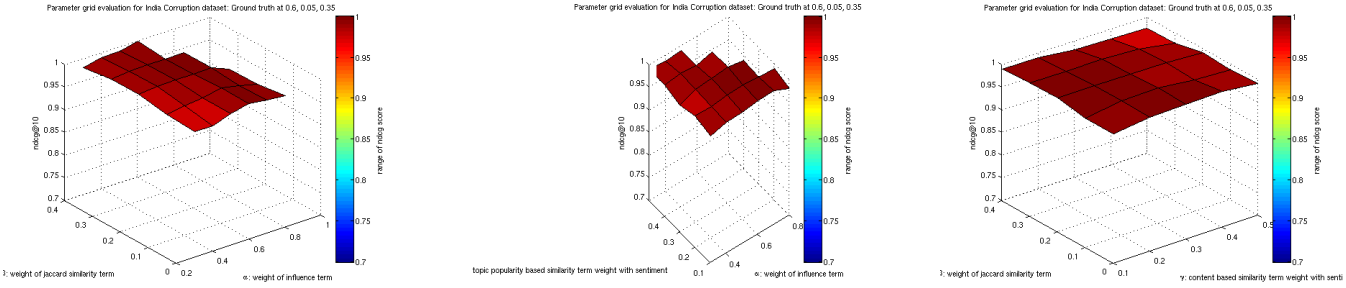

Figure 8: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity and sentiment (India Corruption dataset)

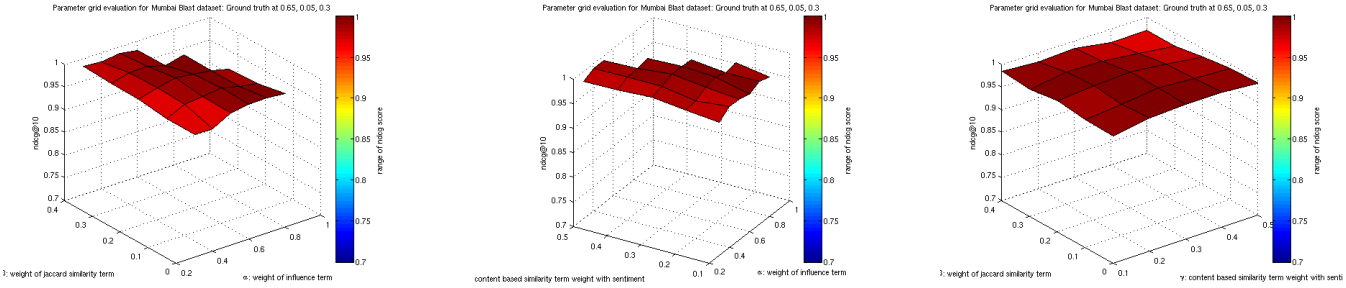

Figure 9: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity and sentiment (Mumbai Blast dataset)

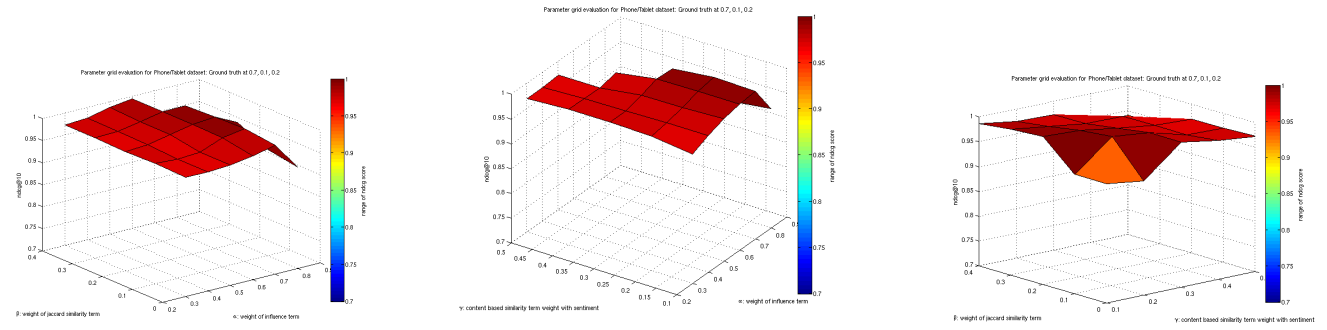

Figure 10: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity and sentiment (Phone/Tablet dataset)

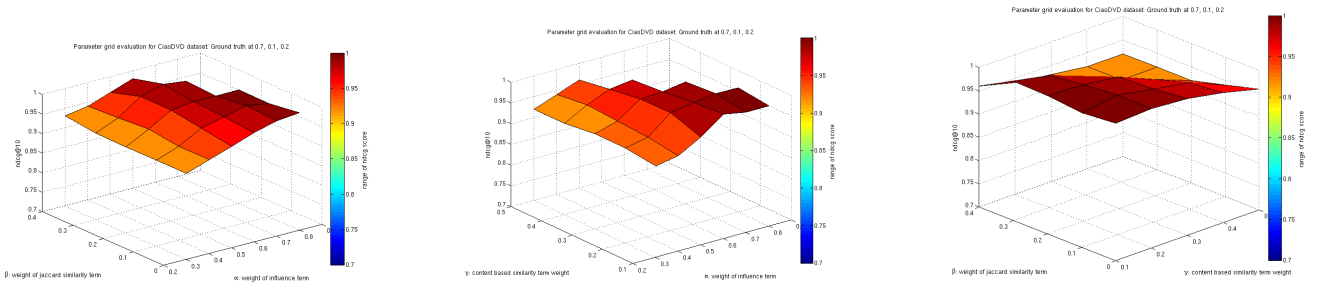

Figure 11: Normalized Discounted Cumulative Gain of various rank orders for the grid of parameter values against ground truth, using degree discounted user similarity and sentiment (CiaoDVD dataset)

thy users). Valence (conditioned by popularity) plays an essential role as well. Structural cohesion, while still prominent has less of an impact than the other two factors. This may be related to the type of event, since structural cohesion did play a more important role for the dataset related to phones, tablets and e-readers. Structural cohesion certainly plays less of a role in the case of ephemeral (Mumbai attacks) and political movement events (Anti-Corruption in India).

Regarding the Phone/Tablet dataset, users who were reasonably high up on the list of trustworthy users (top 500-1000) represented relatively tighter knit communities, as authors of fiction books or technical blogs and information sources such as $Engadget, Phandroid, TechnoBuffalo$ and $FreeKindleReads$. Outside of bloggers, the authors detected as top trustworthy sources fell into two categories. First, those who are very popular on Twitter (for example, R.C McCracken, Richard C.Hale and Luke Romyn) and second those who have more nuanced writing styles (for example, Ed Drury and Ken Lindsey). Both sets of authors were captured as they seemed to be quite similar in terms of their edge connections to tweet clusters and their emotions regarding the topics on which the tweet clusters were based. In contrast, for the Anti-Corruption as well as the Terrorist Attack datasets, the highly trusted users (both renowned on Twitter and otherwise) belonged to a more diverse community of people, including journalists and news agencies ($mid\_day, ndtv, RamCNN$), entertainment industry personalities ($shekharkapur, AnupamPKher$), political and social figures ($PRSLegislative, rameshsrivats$), marketing personnel ($maheshmurthy, GautamRamdurai$), entrepreneurs ($NaveenBachwani$) and non-Indian Twitter users ($lpolgreen, AJEnglish, acarvin, skinnylatte$).

With regards to the question of robustness, valence agreement (sentiment/opinion) plays an important role in increasing the robustness of the method under various parametric settings. We plan to enhance our evaluation with a detailed social survey instrument that polls individual users separately regarding their trust relationships. Such ground truth can provide a broader evaluation on not just the top trustworthy users but also on pairwise relationships across a broad sample of network users.

Finally, our approach not only captures trusted users who are well known and popular but also effectively identifies emergent trustworthy users who are neither well known nor previoulsy popular, during and immediately after an emergency situation (e.g. $ashwinsid$ in the Mumbai blast dataset). This is particularly relevant for our ongoing efforts in identifying newly trustworth citizen sensors both during and immediately after a natural disaster strikes.

4. RELATED WORK

An existing literature addresses the challenge of identifying or predicting the credible aspects of data and entities, under a social model of trust.

Inferring and Encoding Trust: Under some pre-defined notion of trust or conditional trust, the challenge is to encode this notion of trust in some numeric form (say, weights). A simple strategy to evaluate here involves learning the encoding (via a regression or classification) from a labeled set of specific interactions and their associated values obtained

from a domain expert, and then subsequently use the learned model to categorize automatically the trust values associated with other relationships within the network. In the absence of a well established notion of trust, how might this be inferred from signals within the data?

The literatures in both biological and social network analysis rely on the local topology within the network to assess the confidence associated with a specific relationship [10]. However, topological signal alone may be insufficient. For example, Palen et al [33] imply that content and context, and metrics such as re-tweeting by others reflect a notion of trust between the user and her immediate followers. Some of these signals can be directly recovered from the data (e.g. location-specific tags from Twitter, volunteered GIS information etc.) but others must be infered through suitable content analysis and other methods [24]. We hypothesize that the idea of expertise or influence can serve as an prominent parameter of trust. Influence can often be estimated from the structure of the global network, as in Twitter-Rank [34] and Trust-Rank [35] or local network via viewpoint-based methods [2]. Another technique to compute influence simultaneously accounts for the passivity of users in the network [28]. We use of this measure of influence in our own work.

Propogating Trust: One may not have enough data or domain knowledge to assess a meaningful trust value for all of the entities and relationships in the network. Propogation of trust values is required for a completely specified confidence (trust) weighted network. A common method to infer trust relationships in a social graph propagates trust along a path connecting a pair of users using a set of rules, as described by Guha et al in [16]. Golbeck and Hendler [14] described some of the challenges with propogating trust (including conflict resolution), and observed a dependence on the underlying model of trust. Propogation under non-symmetric notions of trust (person A may trust person B but not vice-versa) or non-transitive notions of trust (person A may trust person B and person B may trust person C but this may not say anything about the relationsship between A and C) may lead to trust weighted networks that diverge from those that assume trust symmetry and transitivity.

Several efforts in the literature incorporate trust into online systems and more particularly, into social networks. The important EigenTrust algorithm for determining peer trustworthiness advocates a page-rank style approach to propogating and computing trust [18]. Golbeck and Hendler have adopted an application specific model in the context of the TrustMail application [14]. Additionally, in an algorithm called TidalTrust, Golbeck et al [15] compute predictive movie ratings based on the ratings of trusted people in the network. They propagate trust ratings along a path between a source and a sink node in Friend-Of-A-Friend based social networks using a simple Breadth-First search. The MoleTrust algorithm [23] is another peer-to-peer trust algorithm similar to TidalTrust, but with some variations in the search technique and propagation rules.

Gatterbauer and Suciu discuss three models of trust propogation in the context of probabilistic and uncertain databases, including skeptic, eclectic and optimist models [11] and discuss various measures for conflict resolution. Estimation of trust has also been modeled as a path probablity inference problem. Dubois et al [8] developed an algorithm to infer trust and distrust between users by combining a probabilistic inference algorithm based on random graphs with a modified spring-embedding algorithm, and [32] describes a comparative analysis of Bayesian approaches that use Dirichlet and Beta probability density functions to estimate binary and multidimensional trust in machine networks. This work does not explicitly propagate trust across a network. Instead, it evaluates the trust associated with each member of the network individually.

Another crucial, distinguishing characteristic of trust algorithms, as mentioned in [40] concerns whether they propose a *global* or a *local* trust metric. Global trust metrics compute a single trust value for every node in the network by accounting for information from all nodes and trust edges connecting them. Local trust metrics account for the personal opinions and bias of individual users and compute a different trust value for each pair of users involved in a trust relation. Such local trust metrics contribute to content personalization and recommendation and reputation systems, like the one proposed in [39]. Our work aims to develop a local trust metric.

Finally, the notion of structural similarity between users surely shapes their trust relationships. *Social Identity Theory* [31] substantiates this claim. Tajfel defines the concept of social identity as "the individual's knowledge that he belongs to certain social groups together with some emotional and value significance to him of this group membership". Group membership here is based on a notion of "shared self-identification" or shared interests, and not "cohesive interpersonal relationships". However, such shared identity among members of a group or a team on the basis of the group structure can in turn lead to cohesiveness, uniformity and the motivation to sustain the reputation of their associated identity, which can consequently increase the feeling of mutual trust and affinity among the group members. We incorporate this concept into our trust calculation by increasing the pairwise trust score between users if they belong to a "group", i.e. if they share one or more common topics. Golbeck in [13] also supports idea of a positive correlation between user profile similarity and personalized trust values.

5. CONCLUSION

We presented an algorithm that considers measurable user influence, structural cohesion and valence (sentiment of the users towards the topic in question) as components of evaluating pairwise trust relationships among users in a social network to infer highly trustworthy users, and have evaluated it on three real world social network datasets and a non-social network dataset. Though influence has been found to be the principal factor contributing to recognizing trustworthy users in social media, the presence of valence or sentiment while calculating trust adds significantly to the stability of the decision surface. It develops a trust based ranking of users that is in very good agreement (NDCG score of $> 95\%$) with the ground truth. We are at present investigating how such an analysis can be very useful in emergency response applications, by providing trustworthy sources of citizen sensed data. We are also enhancing our evaluation to examine broad spectrum effects of our trust inference procedure and in particular the propagation of trust when limited ground truth is available as input to our method.

Acknowledgements: This work is supported by NSF Award NSF-EAR-1520870 and NSF-DMS-1418265.

6. REFERENCES

[1] Twitter inc. http://www.twitter.com. Accessed: 2016-05-20.

[2] Asur, S., and Parthasarathy, S. A viewpoint-based approach for interaction graph analysis. In *15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009), ACM, pp. 79–88.

[3] Berg, J., Dickhaut, J., and McCabe, K. Trust, reciprocity, and social history. *Games and economic behavior 10*, 1 (1995), 122–142.

[4] Blei, D. M. Probabilistic topic models. *Communications of the ACM 55*, 4 (2012), 77–84.

[5] Chakrabarti, A., and Parthasarathy, S. Sequential hypothesis tests for adaptive locality sensitive hashing. In *24th International Conference on World Wide Web* (2015), International World Wide Web Conferences Steering Committee, pp. 162–172.

[6] Charikar, M. S. Similarity estimation techniques from rounding algorithms. In *34th annual ACM symposium on Theory of computing* (2002), ACM, pp. 380–388.

[7] Domingos, P., and Richardson, M. Mining the network value of customers. In *Seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (2001), ACM, pp. 57–66.

[8] DuBois, T., Golbeck, J., and Srinivasan, A. Predicting trust and distrust in social networks. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)* (2011), IEEE, pp. 418–424.

[9] Falcone, R., and Castelfranchi, C. Social trust: A cognitive approach. In *Trust and deception in virtual societies*. Springer, 2001, pp. 55–90.

[10] Fortunato, S. Community detection in graphs. *Physics reports 486*, 3 (2010), 75–174.

[11] Gatterbauer, W., and Suciu, D. Data conflict resolution using trust mappings. In *2010 ACM SIGMOD International Conference on Management of data* (2010), ACM, pp. 219–230.

[12] Gionis, A., Indyk, P., Motwani, R., et al. Similarity search in high dimensions via hashing. In *VLDB* (1999), vol. 99, pp. 518–529.

[13] Golbeck, J. Trust and nuanced profile similarity in online social networks. *ACM Transactions on the Web (TWEB) 3*, 4 (2009), 12.

[14] Golbeck, J., and Hendler, J. Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology (TOIT) 6*, 4 (2006), 497–529.

[15] Golbeck, J., Hendler, J., et al. Filmtrust: Movie recommendations using trust in web-based social networks. In *IEEE Consumer communications and networking conference* (2006), vol. 96, Citeseer, pp. 282–286.

[16] Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. Propagation of trust and distrust. In *13th international conference on World Wide Web* (2004), ACM, pp. 403–412.

[17] Guo, G., Zhang, J., Thalmann, D., and Yorke-Smith, N. Etaf: An extended trust antecedents framework for trust prediction. In *2014 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (2014), pp. 540–547.

[18] Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. The eigentrust algorithm for reputation management in p2p networks. In *12th international conference on World Wide Web* (2003), ACM, pp. 640–651.

[19] Kenneth, A., et al. The limits of organization. *N–Y.: Norton* (1974).

[20] Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. S. The web as a graph: measurements, models, and methods. In *Computing and combinatorics*. Springer, 1999, pp. 1–17.

[21] Liu, B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies 5*, 1 (2012), 1–167.

[22] Lott, A. J., and Lott, B. E. Group cohesiveness as interpersonal attraction: a review of relationships with antecedent and consequent variables. *Psychological bulletin 64*, 4 (1965), 259.

[23] Massa, P., and Avesani, P. Controversial users demand local trust metrics: An experimental study on epinions. com community. In *National Conference on artificial Intelligence* (2005), vol. 20, Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, p. 121.

[24] Nagarajan, M., Gomadam, K., Sheth, A. P., Ranabahu, A., Mutharaju, R., and Jadhav, A. *Spatio-temporal-thematic analysis of citizen sensor data: Challenges and experiences*. Springer, 2009.

[25] Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: bringing order to the web.

[26] Purohit, H., Ruan, Y., Fuhry, D., Parthasarathy, S., and Sheth, A. P. On understanding the divergence of online social group discussion. *ICWSM 14* (2014), 396–405.

[27] Rochlin, G. I. Mind the gap: The growing distance between institutional and technical capabilities in organizations performing critical operations. In *Intelligence and Security Informatics*. Springer, 2004, pp. 349–358.

[28] Romero, D. M., Galuba, W., Asur, S., and Huberman, B. A. Influence and passivity in social media. In *Machine learning and knowledge discovery in databases*. Springer, 2011, pp. 18–33.

[29] Satuluri, V., and Parthasarathy, S. Symmetrizations for clustering directed graphs. In *14th International Conference on Extending Database Technology* (2011), ACM, pp. 343–354.

[30] Satuluri, V., and Parthasarathy, S. Bayesian locality sensitive hashing for fast similarity search. *VLDB Endowment 5*, 5 (2012), 430–441.

[31] Tajfel, H. *Social identity and intergroup relations*. Cambridge University Press, 2010.

[32] Thirunarayan, K., Anantharam, P., Henson, C., and Sheth, A. Comparative trust management with applications: Bayesian approaches emphasis. *Future Generation Computer Systems 31* (2014), 182–199.

[33] Vieweg, S., Hughes, A. L., Starbird, K., and Palen, L. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *SIGCHI conference on human factors in computing systems* (2010), ACM, pp. 1079–1088.

[34] Weng, J., Lim, E.-P., Jiang, J., and He, Q. Twitterrank: finding topic-sensitive influential twitterers. In *Third ACM international conference on Web search and data mining* (2010), ACM, pp. 261–270.

[35] Wu, B., Goel, V., and Davison, B. D. Topical trustrank: Using topicality to combat web spam. In *15th international conference on World Wide Web* (2006), ACM, pp. 63–72.

[36] Xu, W., Liu, X., and Gong, Y. Document clustering based on non-negative matrix factorization. In *26th annual international ACM SIGIR conference on Research and development in information retrieval* (2003), ACM, pp. 267–273.

[37] Yang, X., Parthasarathy, S., and Sadayappan, P. Fast sparse matrix-vector multiplication on gpus: implications for graph mining. *VLDB Endowment 4*, 4 (2011), 231–242.

[38] Ziegler, C.-N., and Golbeck, J. Investigating interactions of trust and interest similarity. *Decision support systems 43*, 2 (2007), 460–475.

[39] Ziegler, C.-N., and Lausen, G. Analyzing correlation between trust and user similarity in online communities. In *Trust management*. Springer, 2004, pp. 251–265.

[40] Ziegler, C.-N., and Lausen, G. Propagation models for trust and distrust in social networks. *Information Systems Frontiers 7*, 4-5 (2005), 337–358.