World Wide Web (2015) 18:111–137
DOI 10.1007/s11280-013-0244-2

A personalized credibility model for recommending messages in social participatory media environments

Aaditeshwar Seth · Jie Zhang · Robin Cohen

Received: 4 January 2013 / Revised: 29 May 2013 /
Accepted: 8 July 2013 / Published online: 19 July 2013

Abstract We propose a method to determine the credibility of messages that are posted in participatory media (such as blogs), of use in recommender systems designed to provide users with messages that are considered to be the most credible to them. Our approach draws from theories developed in sociology, political science, and information science—we show that the social context of users influences their opinion about the credibility of messages they read, and that this context can be captured by analyzing the social network of users. We use this insight to improve recommendation algorithms for messages created in participatory media environments. Our methodology rests on Bayesian learning, integrating new concepts of context and completeness of messages inspired by the strength of weak ties hypothesis from social network theory. We show that our credibility evaluation model can be used to significantly enhance the performance of collaborative filtering recommendation. Experimental validation is done using datasets obtained from social networking websites used for knowledge sharing. We conclude by clarifying our relationship to the semantic adaptive social web, emphasizing our use of personal evaluations

A. Seth
Department of Computer Science and Engineering, Indian Institute of Technology Delhi,
New Delhi, Delhi 110016, India
e-mail: aseth@cse.iitd.ernet.in

J. Zhang (✉)
School of Computer Engineering, Nanyang Technological University, Nanyang Avenue,
Singapore 639798, Singapore
e-mail: zhangj@ntu.edu.sg

R. Cohen
School of Computer Science, University of Waterloo, 200 University Ave. W,
Waterloo, ON N2L 3G1, Canada
e-mail: rcohen@cs.uwaterloo.ca

of messages and the social network of users, instead of merely automated semantic interpretation of content.

Keywords Social networks · Recommender systems · Credibility modeling · Participatory media

1 Introduction

With the goal to provide personalization on the web, one topic of concern is how to assist users in processing vast collections of participatory media content that exist in this environment. A specific challenge is to design a personalized recommender system that will be able to propose messages of interest to users. For example, *citizen journalism* through participatory media content such as blogs and comments on news articles, has become a popular supplement to mass media [10]. It adds diversity of opinion to news topics, and provides an additional level of localization in news coverage, addressing issues that may have been skipped by national news agencies. However, these collections of participatory media content tend to be huge and dynamic; for example, on the order of 1.6 million blog posts are written each day [34]. Users in these settings are faced with a plethora of messages and existing techniques such as RSS feeds are not personalized, causing users to often sift their way through hundreds of messages each day. In an effort to help users to cope with this vast amount of news, personalized recommender systems that propose news articles of interest to users would be beneficial.

In this paper, we examine one particular concern in the processing of messages, the modeling of a message's credibility, and use this to construct a recommender system which provides to users those messages that are considered to be the most credible. We contend that credibility is an important component to judge the usefulness of participatory media content. This is because of the ease of publishing information on the Internet without any editorial checks by a formal agency: Anybody can publish "incorrect" information, or bad-mouth "correct" information. We depart from this conventional thinking on the polarity of information credibility, however, and instead develop a personalized credibility model suitable for the scenario of participatory media. We show that the social context of users influences their opinion about the credibility of messages they read, where we define *social context* as the embedding of people in the real world, based on their families, professions, incomes, geographical locations, political affiliations, etc. We capture this social context by analyzing the social network of users, and the ratings users give to messages. This results in a personalized delivery of web content to users by making use of structured data that is machine readable, rather than relying only on the semantic analysis of message content.

Our model for evaluating message credibility aims to capture the following principles:

- *D-1*: Different users may judge the credibilities of blogs differently according to their own social context.
- *D-2*: Different users may associate different degrees of credibility to the public opinion or to the beliefs of other groups of users or to their own beliefs.

- *D-3*: The credibility of a blogger is topic specific; an expert in some area may not be an expert in another.
- *D-4*: A highly credible blogger can occasionally make mistakes and give inaccurate information. Analogously, useful blog-entries could be written by a blogger unknown so far.

We draw from research in media studies, information science, political science, and social networks to refine these design principles into specific criteria that can be used to judge the credibility of information. These criteria include, for example, the influence of public opinion, influence of close friends of people, and the extent to which different people may trust their own beliefs. The influence of social networks itself is validated through surveys of users of an online social networking website, orkut.com, and shows that social network information can be used to infer the social context in which users may perceive the credibility of information. We then use these criteria to build and learn a Bayesian network on a personalized basis for each user, to predict which messages the user may find to be credible. Our method makes extensive use of social network information to create the user model, and combines the link structure of social networks of users with information about authorship and ratings of messages by users. We test our method on a dataset obtained from a popular knowledge sharing website, digg.com. Experimental results show that our method outperforms other well-known methods such as Pagerank used to rank Internet web-pages in order of their importance [26], Eigentrust used in peer-to-peer (P2P) file-sharing systems to identify trustworthy peers that upload valid content [19], and the beta-reputation system used in e-commerce to evaluate the trustworthiness of buyers and sellers [40].

Our method has important implications for the design of recommender systems for participatory media content. It serves to predict the probability of a user finding a message to be credible, and can hence be used as a pre- or post-filtering stage with existing recommendation algorithms. In this paper, we show that our method can be adapted to integrate closely with collaborative filtering (CF) [2]; enhancing a CF algorithm with our credibility model can be shown to perform significantly better than the basic CF for a binary-classification of messages (i.e. {*recommend*, *not recommend*}) to a user.

Our research also emphasizes the importance of considering the social network of a user, when making recommendations. In particular, we develop a new approach for leveraging a user's social network, show that social network information can be used to infer the social context of users, and validate it through surveys of users of a social networking website, orkut.com.

Finally, our techniques are ones that are demonstrated to work well in practice. As will be seen, our validation draws on responses from real users using a dataset from digg.com, providing users with messages that are rated highly.

In all, the contributions of our current work can be summarized as follows:

- We draw insights from various studies in different disciplines to determine the credibility judgement criteria and how social context may influence users' judgement of message credibility;
- A Bayesian network based credibility model is constructed to combine the multiple dimensions (criteria) of credibility as well as the social context information, to infer the credibility of messages to users;

- The effectiveness of the proposed credibility model as well as its usefulness in improving recommender systems are validated using a real dataset.

In the sections that follow, we first use insights from sociology to determine the credibility judgement criteria used by people, and how social network information can be used to aid discrimination between these different criteria in Section 2. We then describe the Bayesian network model in detail in Sections 3 and 4, and provide the evaluation of our method in Section 5. We also show how our modeling technique can be adapted to improve the performance of collaborative filtering for recommendation of participatory media content. Finally, we present related work, a discussion, and future work in Sections 6, 7 and 8.

2 Credibility judgement criteria

In this section, we build upon insights about credibility developed in different disciplines. We then use these insights to construct a Bayesian model for each user; the model parameters can be learned using positive and negative ratings given by a user to messages seen by her in the past, and can be used to predict whether the user will find a new unseen message to be credible.

2.1 An example

Websites such as Amazon.com allow people to post book reviews. Consider the following reviews in Examples 1 and 2 given for a book.

Example 1 (*A Professor's Review*) I have been working in the field of signal processing and speech for more than 40 years and, more recently, as a **professor** ... I am **extremely impressed** with the book. The writing style is such that **understanding is maximized** by the clarity of thought and examples provided ...

Example 2 (*A Student's Review*) I can appreciate others who might think that this is a great book ... but I am a **student** using it and I have some very different opinions of it ... Second, while it is certainly a textbook, the author clearly has an understanding of the material that seems to **undermine his ability to explain** it. Though there are mentions of **examples there are, in fact, none** ...

Both the reviews appear to give contradictory opinions: which review should be considered more credible? The contradiction disappears if the social context of the reviewers is considered. The first reviewer is a professor who has a good background in statistics and machine learning, and it is quite possible that the examples given in book would have seemed sufficient to him. However, the second reviewer is a student who probably does not have a rich background in the subject, and hence found the book hard to read. Consider a user who is reading these two different reviews. Determining which one is credible would need to be a subjective assessment; it is unsurprising to find conflicting judgements. Note also that both the reviewers mentioned their role, that is, whether they were a professor or a student. This suggests that credibility judgement by users can also be influenced by the role of the reviewer because, for instance, a user may believe a professor's review to be more credible than that of a student.

Two fundamental principles seem to come out of this example. First, credibility seems to be a multi-dimensional construct with various factors influencing the degree to which a user may find a message to be credible or not. Second, the social context of a user seems to have a significant influence in how the user perceives the message. We describe these in greater next, situating our insights in reference to related work in the area.

2.2 Multi-dimensional construct

Such observations have been resolved by various researchers who model credibility as a multi-dimensional construct [8, 30]. Fogg and Tseng reason about credibility criteria used by people to judge the credibility of computerized devices and software, and identify the following different types of credibility:

- *Experienced*: This is based on *first-hand experience* of a user, and reflects her personal belief about a device or software.
- *Presumed*: This reflects personal biases of a user that give rise to *general assumptions* about certain categories of computing products; for example, presumptions based upon the company which developed the product, the cost of the product, the importance of the function performed by the product, etc.
- *Reputed*: This is based on *third-party reports* about different products.

A model with similar distinctions is developed in [30] to evaluate the trustworthiness of agents in an e-commerce setting. Here, the authors distinguish *witness reputation* (i.e. general public opinion) from *direct reputation* (i.e. opinion from a user's own experience) and include as well *system reputation* (i.e. the reputation from the role of an agent, as buyer, seller or broker). We next consider relevant studies from sociology and political science for additional valuable insights.

2.3 Social context and social networks

With reference to the example cited earlier, it seems that the social context of users being students or professors influence their credibility judgement criteria. How can we infer the social context that users may apply to different messages? We take help from social networks here.

People are embedded in real-world social networks of relationships as friends, acquaintances, family members, etc. The *strength-of-weak-ties* hypothesis in sociology [11] states that such social networks consist of clusters of people with *strong* ties among members of each cluster, and *weak* ties linking people across clusters, shown in Figure 1. Whereas strong ties are typically constituted of close friends, weak ties are constituted of remote acquaintances. The hypothesis claims that weak ties are useful for the diffusion of information and economic mobility, because they connect diverse people with each other. On the other hand, people strongly tied to each other in the same cluster may not be as diverse.

Applying this hypothesis to the example, it may seem that professors would be strongly tied in their own cluster, while students may be strongly tied within other clusters, and weak ties would link the different clusters of students and professors

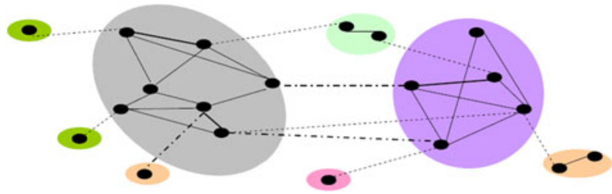


Figure 1 Strong ties (*solid lines*) and weak ties (*dashed lines*)

together. If this is indeed true, then we may be able to use social network information to infer the social context of users, and use that to improve credibility models.

One among many studies indeed corroborates the *strength-of-weak-ties* hypothesis. Baybeck and Huckfeldt [3] traces the changes in political opinion of people before and after the 1996 presidential elections in USA, observed with respect to the social networks of people. It is shown that weak ties (identified as geographically dispersed ties of acquaintances) are primarily responsible for the diffusion of divergent political opinion into localized clusters of people having strong ties between themselves. As indicated by the *strength-of-weak-ties* hypothesis, this reflects that local community clusters of people are often homogeneous in opinion, and these opinions may be different from those of people belonging to other clusters. Furthermore, people have different degrees to which they respect the opinions of those not in their immediate local community cluster. This reflects that the personal characteristics of people also influence the extent to which they would be comfortable in deviating from the beliefs of their immediate local cluster. These observations provide two insights:

- Reputed credibility has at least two sub-types: *cluster credibility* based on the opinions of people in the same cluster or local community, and *public credibility* based on the general opinions of everybody.
- Users have different personal characteristics to weigh the importance of different types of credibilities.

The first insight suggests refining *reputed* credibility to also consider reports from those in the same cluster. The second insight is reinforced by studies in information science [28], which argue that users have different preferences for different types of credibilities discussed so far. In the next few sections, we use these insights to develop and operationalize a multi-dimensional personalized credibility model for participatory media.

3 Bayesian credibility model

3.1 Knowledge assumptions

Suppose that we wish to predict whether a message m_k about a topic t and written by user u_j, will be considered credible by user u_i. We assume that we have the following prior knowledge:

- We consider a scenario where all messages about topic t written in the past are labeled with the author of each message. In addition, a message may have also been assigned ratings by various recipient users, whenever users would have

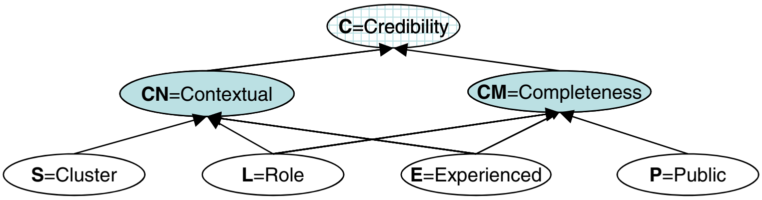


Figure 2 Credibility model

read the message, based on the credibility of the message for the recipient. The credibility ratings of messages are also assumed to be available.[1] Here, the credibility rating of a message is binary, either 1 or 0, indicating that the message is credible or not, respectively.

- Users may declare a subset of other users as their "friends". We refer to an explicitly declared relationship between two users as a *link* between them, and assume to have knowledge of the social network graph formed by all users and the links between pairs of users.
- Users may also declare topics of interest to them. We use this information, and the social network graph, to derive the *topic specific social network graph* for topic t, as the induced subgraph of the overall social network graph consisting only of those users and edges between users who are interested in topic t.
- For each topic specific social network graph, community identification algorithms such as [9, 36, 37] can identify dense clusters of users and links. We use the definition of *strong* and *weak* ties proposed by [11], and refer to *strong* ties as links between users in the same cluster, and *weak* ties as links between users in different clusters. We use V_{it} to denote the local cluster of users strongly tied to user u_i with respect to topic t.

These assumptions are reasonable in the case of websites such as digg.com, which allow users to construct social networks by declaring some users as their friends. Information about message authorship and ratings given by users to messages is also available. We will show that we can use this knowledge to quantify different types of credibilities for each message with respect to each user. Then, based on ratings given by a particular user to older messages, we can use a Bayesian model to learn preferences of the user towards these different kinds of credibilities of messages. Finally, we can use this learned model to predict whether or not the new message m_k will be considered credible by user u_i.

3.2 Bayesian network

The different types of credibilities that we choose to model include cluster credibility, public credibility, experienced credibility, and role based credibility. We also propose a Bayesian network to combine them into a single credibility score. The model is shown in Figure 2 with two hidden variables as shaded ovals. Directed edges indicate

[1]We also assume that we are beyond the cold-start stage so that the set of older messages have all received some ratings, and all users have provided at least some ratings.

a dependency from the originating variable to the target variable. Unshaded ovals represent evidence variables. The partially shaded oval for message credibility is a variable denoting the rating given by the user, and is available as an evidence variable during the training phase only. The goal is to infer this variable for a new message, given the evidence variables and the parameters of the learned model. The hidden variables help make the model more tractable to learn, and also capture the insight of being able to use social networks to infer the social context of users. We define context, completeness and the different types of credibilities as follows:

- *Context* relates to the ease of understanding of the message, based on how well the message content explains the relationship of the message to its recipient. *Simplification* of the meaning of the message [5], can be considered as an outcome of the amount of *context* in the message. That is, messages that are more contextual for users, will be more simple for them to understand.
- *Completeness* denotes the depth and breadth of topics covered in the message. The *scope* of the message, or the *opinion diversity* expressed in the message [5], can be considered as outcomes of the degree of *completeness* of the message. That is, messages that are more complete will carry more diverse opinions or more mention of relationships with other issues.
- s_{ikt} = *cluster credibility*: A sub-type of reputed credibility discussed earlier, this is based on the ratings given by other users in cluster V_{it}, that is, the cluster of user u_i. It denotes the credibility associated by the cluster or local community of u_i to the message m_k written by u_j. We assume that opinions of users in the same cluster will contribute only to adding context to messages; their contribution to completeness is already accounted for through public credibility explained next.
- p_{kt} = *public credibility*: Another sub-type of reputed credibility, this is based on ratings by all the users, and reflects the general public opinion about the credibility for the message m_k written by u_j. Public credibility contributes only to the completeness of messages across all users, including the users who's opinions have already been accounted in the cluster credibility construct.
- e_{ikt}= *experienced credibility*: Identical to the concept of experienced credibility discussed earlier in Section 2.2, this is based only on ratings given by user u_i in the past, and denotes the credibility that u_i associates with the message m_k written by u_j. We distinguish between the contributions experienced credibility would make to adding context to the message, or adding completeness.
- l_{ikt}= *role based credibility*: Similar to presumed credibility discussed earlier, this denotes the credibility that u_i associates with the message m_k written by users having the same role as that of u_j; for example, based on whether the messages' authors are students, or professors, or journalists, etc.

We reason that cluster credibility will only influence contextual credibility, while public credibility will only influence completeness credibility. This is because general public opinion is by definition averaged over different contexts, and hence it will only add noise to any context specific credibility. Similarly, cluster credibility will double count the opinion of a specific cluster when judging the degree of completeness or diversity in a message. Other types of credibilities, experienced and role based, will influence both contextual and completeness credibility since they are based on the personal beliefs of the user.

Each of the credibilities can be expressed as a real number $\in [0, 1]$. Our aim is to learn the distribution for $P_{it}(\mathbf{C}|\mathbf{E,L,S,P})$ for each user and topic based on ratings given by various users to older messages; here, $\{\mathbf{E,L,S,P}\}$ are evidence variables for the four types of credibilities for a message, and $\mathbf{C}$ is a variable denoting the credibility that u_i associates with the message. Thus, for each topic t, a set of messages M about t will be used during the training phase with samples of $(c_{ik}, e_{ik}, l_{ik}, s_{ik}, p_k)$ for different messages $m_k \in M$ to learn the topic specific credibility models for u_i. Assuming that a user's behavior with respect to preferences for different kinds of credibilities remains consistent over time, the learned model can now be used to predict c_{ix} for a new message m_x about topic t, that is, $P_{it}(c_{ix}|e_{ix}, l_{ix}, s_{ix}, p_x)$.

3.3 Meeting the design principles

Our modeling method is able to satisfy three out of the four design principles listed in Section 1. (*D-1*) The model takes into account personal and contextual opinions of people that may influence their credibility judgements. (*D-2*) The model is learned in a personalized manner for each user, and allows to accommodate varying degrees of openness of users to respect opinions of other users. (*D-3*) Different model instances are learned for different topics, making credibility judgements topic specific. (*D-4*) We will show in the next section that the fourth principle of allowing mistakes by credible users and useful messages by less-credible users can also be modeled in this framework.

4 Credibility computation

In this section, we describe how the different types of credibilities can be computed based on social network information, ratings given by users to messages, and authorship information. The notion of credibility of messages is extended to credibility of users as well. We first list the axioms that are the basis for our formulation to quantify the various types of credibilities, and then give the actual computation process.

4.1 Axioms to calculate credibility

We use the axiomatic assumptions embodied in the following relationships:

- *A-1*: A message is credible if it is rated highly by credible users.
- *A-2*: A user is credible if messages written by her are rated highly by other credible users.
- *A-3*: A user is also credible if ratings given by her are consistent with the ratings given by credible users.
- *A-4*: A user is also likely credible if she is linked to by many other credible users in the social network.

There is clearly a recursive relationship between these axioms. We solve the recursion using fixed-point Eigenvector computations, as described next.

4.2 Calculation of evidence variables

We henceforth assume that we are operating within some topic t, and drop the subscript for simplicity. As stated in the knowledge assumptions earlier, we start with the following information that will be a part of our training set.

- **A[k,n]**: A matrix for k messages and n users, where $a_{ij} \in \{0, 1\}$ indicates whether message m_i was written by u_j;
- **R[k,n]**: A ratings matrix for k messages and n users, where $r_{ij} \in \{0, 1\}$ indicates the rating given to message m_i by user u_j;[2]
- **N[n,n]**: A social network matrix where $n_{ij} \in \{0, 1\}$ indicates the presence or absence of a link from user u_i to user u_j. We also assume that the clustering algorithm can identify clusters of strong ties among users, connected to other clusters through weak ties.

Our goal is to find a method to compute the evidence variables for the Bayesian model using the axioms given above. The evidence variables can be expressed as the matrices **E[n,k]**, **L[n,k]**, **S[n,k]**, and **P[k]**, containing the credibility values for messages. Here, p_k is the public credibility for message m_k authored by user u_j. e_{ij} and l_{ij} are the experienced and role based credibilities respectively for message m_k according to the self-beliefs of user u_i. Similarly, s_{ij} is the cluster credibility for message m_k according to the beliefs of the users in u_i's cluster V_i. Once these evidence variables are computed for older messages, they are used to learn the Bayesian model for each user. Subsequently, for a new message, the learned model for a user is used to predict the credibility of the new message for the user.

We begin with computation of the evidence variable matrix for public credibility **P**; we will explain later how other credibilities can be computed in a similar fashion.

1. Let $\mathbf{P}'[\mathbf{n}]$ be a matrix containing the public credibilities of users, and consider the credibility of a message as the mean of the ratings for the message, weighted by the credibility of the raters (*A-1*):

$$p_k = \sum_i r_{ki}.p'_i / |r_{ki} > 0|$$

   Here, the denominator counts the number of occurrences of ratings greater than 0. This is the same as writing $\mathbf{P} = \mathbf{R}_r \cdot \mathbf{P}'$, where $\mathbf{R}_r$ is the row-stochastic form of $\mathbf{R}$, ie. the sum of elements of each row $= 1$.
2. The credibility of users is calculated as follows:

   2a. Consider the credibility of a user as the mean of the credibilities of the messages written by her (*A-2*):

$$p'_i = \sum_k p_k / |p_k|$$

   This is the same as writing $\mathbf{P}' = \mathbf{A}_c^T \cdot \mathbf{P}$, where $\mathbf{A}_c$ is the column-stochastic form of $\mathbf{A}$; and $\mathbf{A}_c^T$ is the transpose of $\mathbf{A}_c$.

[2] We assume in this paper that the ratings are binary. However, our method can be easily generalized to real-valued ratings as well. In the future, we also plan to extend the method to accept explicit negative ratings using distrust propagation [12].

2b. The above formulation indicates a fixed point computation:

$$\mathbf{P}' = A_c^T \cdot \mathbf{R}_r \cdot \mathbf{P}' \tag{1}$$

Thus, $\mathbf{P}'$ can be computed as the dominant Eigenvector of $\mathbf{A}_c^T \cdot \mathbf{R}_r$. This formulation models the first two axioms, but not yet the ratings-based credibility (*A-3*) and social network structure of the users (*A-4*). This is done as explained next.

2c. Perform a fixed-point computation to infer the credibilities **G[n]** acquired by users from the social network (*A-4*):

$$\mathbf{G} = (\beta \cdot \mathbf{N}_r^T + (\mathbf{1} - \beta) \cdot \mathbf{Z}_c \cdot 1^T) \cdot \mathbf{G} \tag{2}$$

Here, $\beta \in (0, 1)$ denotes a weighting factor to combine the social network matrix **N** with the matrix **Z** that carries information about ratings given to messages by users. We generate **Z** by computing z_i as the mean similarity in credibility ratings of user u_i with all other users. The ratings similarity between a pair of users is computed as the Jacquard's coefficient of common ratings between the users. Thus, z_i will be high for users who give credible ratings, that is, their ratings agree with the ratings of other users (*A-3*). In this way, combining the social-network matrix with ratings-based credibility helps to model the two remaining axioms as well. Note that $\mathbf{Z}_c$**[n]** is a column stochastic matrix and **1[n]** is a unit column matrix; augmenting **N** with $\mathbf{Z}_c \cdot \mathbf{1}^T$ provides an additional benefit of converting **N** into an irreducible matrix so that its Eigenvector can be computed.[3]

2d. The ratings and social network based scores are then combined together as:

$$\mathbf{P}' = (\alpha \cdot \mathbf{A}_c^T \cdot \mathbf{R}_r + (\mathbf{1} - \alpha) \cdot \mathbf{G}_c \cdot 1^T) \cdot \mathbf{P}' \tag{3}$$

Here again **1** is a unit column matrix, and $\alpha \in (0, 1)$ is a weighting factor. The matrix $\mathbf{P}'$ can now be computed as the dominant Eigenvector using the power method.

3. Once $\mathbf{P}'$ is obtained, **P** is calculated in a straightforward manner as $\mathbf{P} = \mathbf{R}_r \cdot \mathbf{P}'$.

Note that the above method is only one way of combining the different pieces of information we have. Our objective in presenting this method is to show that information about social networks, ratings, and authorship can be combined together and to then examine the performance of this method compared to competing approaches.

The above process is to compute the public credibilities **P[k]** of messages. The processes to compute cluster **S[n,k]**, experienced **E[n,k]**, and role based **L[n,k]** credibilities are identical, except that different cluster credibilities are calculated with respect to each cluster in the social network, and different experienced and role based credibilities are calculated with respect to each user. This is why cluster and

[3]This step is similar to the Pagerank or HITS computations for the importance of Internet web pages [20, 26]. The matrix **N** can be considered as the link matrix of web-pages, and the matrix **Z** as the pagerank personalization matrix. The output matrix **G** then essentially ranks the web-pages in order of their importance, after taking personalization into account.

Algorithm 1 Training set preparation

Input: $\mathbf{A[k,n]}$, $\mathbf{R[k,n]}$, $\mathbf{N[n,n]}$
Output: $\mathbf{P[k]}$, $\mathbf{E[n,k]}$, $\mathbf{S[n,k]}$, $\mathbf{P'[k]}$, $\mathbf{E'[n,n]}$, $\mathbf{S'[n,n]}$

1. Compute similarity matrix $\mathbf{Y[n,n]}$

```
forall i ∈ 1..n, j ∈ 1..n, i ≠ j do
    forall m ∈ 1..k do
        if R[m,i] = R[m,j] then
            Y[i,j] ← Y[i,j] + 1/k
```

2. Compute public credibilities $\mathbf{P[k]}$, $\mathbf{P'[n]}$

```
Z[n] ← 0
forall i ∈ 1..n do
    forall j ∈ 1..n do
        Z[i] ← Z[i] + Y[j,i]
```

Solve for $\mathbf{G[n]}$: $\mathbf{G}=(\beta.\mathbf{N}_r^T+(\mathbf{1}\text{-}\beta).\mathbf{Z}_c\cdot\mathbf{1}^T).\mathbf{G}$
Solve for $\mathbf{P'[n]}$: $\mathbf{P'}=(\alpha\cdot\mathbf{A}_c^T\cdot\mathbf{R}_r + (\mathbf{1}\text{-}\alpha)\cdot\mathbf{G}_c\cdot\mathbf{1}^T)\cdot\mathbf{P'}$
$\mathbf{P} \leftarrow \mathbf{R}_r\cdot\mathbf{P'}$

3. Compute cluster credibilities $\mathbf{S[n,k]}$,

$\mathbf{S'[n,n]}$

forall *Cluster* $V_c \in$ *clusters in social network* **do**

- $\mathbf{Z[n]} \leftarrow 0$
- $\underline{\mathbf{G}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{P}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{P'}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{R}}\mathbf{[k,n]} \leftarrow 0$
- **forall** $j \in$ *users in* V_c **do**
  - **forall** $i \in 1..n$ **do**
    - $\mathbf{Z[i]} \leftarrow \mathbf{Z[i]} + \mathbf{Y[j,i]}$
  - **forall** $m \in 1..k$ **do**
    - $\underline{\mathbf{R}}\mathbf{[m,j]} \leftarrow \mathbf{R[m,j]}$
- Solve for $\underline{\mathbf{G}}\mathbf{[n]}$: $\underline{\mathbf{G}}=(\beta\cdot\mathbf{N}_r^T+(\mathbf{1}\text{-}\beta)\cdot\mathbf{Z}_c\cdot\mathbf{1}^T)\cdot\underline{\mathbf{G}}$
- Solve for $\underline{\mathbf{P}}'\mathbf{[n]}$: $\underline{\mathbf{P}}'=(\alpha\cdot\mathbf{A}_c^T\cdot\underline{\mathbf{R}}_r+(\mathbf{1}\text{-}\alpha)\cdot\underline{\mathbf{G}}_c\cdot\mathbf{1}^T)\cdot\underline{\mathbf{P}}'$
- $\underline{\mathbf{P}}=\underline{\mathbf{R}}_r\cdot\underline{\mathbf{P}}'$
- **forall** $j \in$ *users in* V_c **do**
  - **forall** $m \in 1..k,\ u \in 1..n$ **do**
    - $\mathbf{S'[j,u]} \leftarrow \underline{\mathbf{P'}}\mathbf{[u]}$; $\mathbf{S[j,m]} \leftarrow \underline{\mathbf{P}}\mathbf{[m]}$

4. Compute experienced credibilities $\mathbf{E[n,k]}$,

$\mathbf{E'[n,n]}$

forall *User* $i \in 1..n$ **do**

- $\mathbf{Z[n]} \leftarrow 0$
- $\underline{\mathbf{G}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{P}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{P'}}\mathbf{[n]} \leftarrow 0$, $\underline{\mathbf{R}}\mathbf{[k,n]} \leftarrow 0$
- **forall** $j \in 1..n$ **do**
  - $\mathbf{Z[j]} \leftarrow \mathbf{Y[j,i]}$
- **forall** $m \in 1..k$ **do**
  - $\underline{\mathbf{R}}\mathbf{[m,i]} \leftarrow \mathbf{R[m,i]}$
- Solve for $\underline{\mathbf{G}}\mathbf{[n]}$: $\underline{\mathbf{G}}=(\beta\cdot\mathbf{N}_r^T+(\mathbf{1}\text{-}\beta)\cdot\mathbf{Z}_c\cdot\mathbf{1}^T)\cdot\underline{\mathbf{G}}$
- Solve for $\underline{\mathbf{P}}'\mathbf{[n]}$: $\underline{\mathbf{P}}'=(\alpha\cdot\mathbf{A}_c^T\cdot\underline{\mathbf{R}}_r+(\mathbf{1}\text{-}\alpha)\cdot\underline{\mathbf{G}}_c\cdot\mathbf{1}^T)\cdot\underline{\mathbf{P}}'$
- $\underline{\mathbf{P}} \leftarrow \underline{\mathbf{R}}_r.\underline{\mathbf{P}}'$
- **forall** $m \in 1..k,\ u \in 1..n$ **do**
  - $\mathbf{E'[i,u]} \leftarrow \underline{\mathbf{P'}}\mathbf{[u]}$; $\mathbf{E[i,m]} \leftarrow \underline{\mathbf{P}}\mathbf{[m]}$

experienced credibility matrices are 2-dimensional, while the public credibility is only 1-dimensional. For example, considering a message m_3 and a recipient user u_1, **P[3]** is the public credibility of message m_3; **E[1,3]** is the experienced credibility of message m_3 according to the self-belief of recipient u_1; **L[1,3]** is the role based credibility of message m_3 also according to the self-belief of recipient u_1; and **S[1,3]** is the cluster credibility of message m_3 according to the beliefs of users in cluster V_1 of recipient u_1. The processing steps for computing these quantities are outlined in Algorithm 1; A description of the processing steps for computing these quantities is as follows:

- The cluster credibilities **S[n,k]** are computed in the same manner as the public credibilities, but after modifying the ratings matrix **R** to contain only the ratings of members of the same cluster. Thus, the above process is repeated for each cluster, modifying **R** in every case. For each users u_i belonging to cluster V_i, s_{ik} is then equal to the cluster credibility value for message m_k with respect to u_i. The matrix **Z** in the computation on the social network matrix is also modified. When computing the cluster credibilities for cluster V_i, element z_j of **Z** is calculated as the mean similarity of user u_j with users in cluster V_i. Thus, z_j will be high for users who are regarded credible by members of cluster V_i because their ratings agree with the ratings of the cluster members.
- The experienced credibilities **E[n,k]** are computed in the same manner as well, but this time for each user by modifying the ratings matrix **R** to contain only the ratings given by the user. The matrix **Z** is also modified each time by considering z_j as the similarity between users u_i and u_j, when calculating the experienced credibilities for u_i.
- Role based credibility is computed as the mean experienced credibilities of users having the same role. However, we do not use role based credibility in our evaluation because sufficient user profile information was not available in the digg dataset used by us. Henceforth, we ignore **L[n,k]** in our computations.

Algorithm 2 Inference phase (ratings based)

Input: User i, Cluster V_i of user i, Message m;
    Ratings $\boldsymbol{R}[\mathbf{n,m}]$ given by other users to m;
    Learned model for user i
Output: P(user i will find m to be credible | $\boldsymbol{R}[\mathbf{k}]$)

$p_m \leftarrow \text{mean}(\boldsymbol{R}[\mathbf{j,m}]\cdot\mathbf{P}'[\mathbf{j}])_{j\in 1..n}$
$s_{im} \leftarrow \text{mean}(\boldsymbol{R}[\mathbf{j,m}]\cdot\mathbf{S}'[\mathbf{i,j}])_{j\in 1..n}$
$e_{im} \leftarrow \text{mean}(\boldsymbol{R}[\mathbf{j,m}]\cdot\mathbf{E}'[\mathbf{i,j}])_{j\in 1..n}$
$\text{P}(c_{im}|p_{im}, s_{im}, e_{im}) \leftarrow$ MCMC on learned model for i

4.3 Model learning

Once various types of credibilities for messages are calculated with respect to different users, this data is used to learn the Bayesian model for each user and topic of interest using the Expectation-Maximization (EM) algorithm [29]. Model parameters are learned to predict for user u_i interested in topic t, the probability $\text{P}_{it}(c_{ix}|e_{ix}, s_{ix}, p_x)$ that u_i will find a new message m_x to be credible.

4.4 Inference

Now, for a new message m_x, the evidence variables are calculated with respect to a recipient user u_i in one of two ways as described next, and the learned model is used to produce a probabilistic prediction of whether u_i would find m_x to be credible.

- *Authorship*: The four types of credibilities of the message are considered to be the same as the corresponding four types of credibilities of its author with respect to u_i.
- *Ratings*: The cluster and public credibilities are calculated as the weighted mean of ratings for the message given by other users and the credibilities of these users with respect to u_i. The experienced and role based credibilities are the same as the corresponding credibilities of the message author with respect to u_i.

As we will show in the evaluation, the ratings method performs better than the authorship method. This also meets the fourth design principle (*D-4*) listed in Section 1. Since credibility is evaluated through ratings given to the message by various users, it allows new users to popularize useful messages written by them because their own credibility does not play a role in the computations. It also allows credible users to make mistakes because the credibility of the author is not taken into account.

Given the evidence variables for the new message, and the learned Bayesian model, the probability of u_i finding the message to be credible is computed using standard belief propagation methods such as Markov–Chain–Monte–Carlo (MCMC) [29]. The outline is given in Algorithm 2.

5 Evaluation

We present our evaluation in two parts. First, we validate our hypothesis that social networks give an indication of context and completeness. This is done through surveys of real users of orkut.com, a popular social networking website. Second, we test our Bayesian model for credibility assessment on a dataset obtained from digg.com, a popular knowledge sharing website.

5.1 Hypothesis testing for context and completeness

The hypothesis we wish to test can be stated as follows:

> *Given a classification of ties as strong or weak in a topic specific social network, people with strong ties linking each other share a similar social context than with weak ties, and people with weak ties linking each other provide more completeness than strong ties.*

We crawled a popular social networking website, Orkut, and validated the hypothesis through surveys of real users.

5.1.1 Preparation of the dataset

We wrote a web-crawler that screen-scraped a snow-ball sample from Orkut.com to obtain a social network graph of almost 800,000 users. Snowball sampling has a tendency to oversample hubs, and therefore we identified a core-set of approximately 42,000 users whose social network graph was known to a high degree of completion. This was done by recording only those users whose indegree was close to their outdegree in the initial larger dataset, making use of the evidence that the Orkut social graph has been seen to have a high degree of reciprocity [25]. This core-set of 42,000 users was used for further analysis, and the graph was considered as undirected for our experiments.

The graph followed a power-law in degree distribution with a truncation at 200, as also noticed in multiple other related studies [1, 25]. Orkut users can subscribe to various communities of interest and participate in discussions; we also crawled the community memberships for the set of users in our dataset, and a large number of discussions (ie. *messages*) in these communities. Orkut allows communities to link to other related communities; we then crawled the community graph and clustered it to derive coarse topics using a flow-stochastic graph clustering algorithm [37]. Some examples of topic clusters of communities that were identified were {*Books*, *Literature*, *Simply books*}, and {*Mumbai*, *Mumbai that I dream about*, *Mumbai bloggers*}. This indicates that related communities were indeed present in the same clusters to determine broad topics of interest. From this we were able to obtain the interests of users in different topics. Knowledge of user interests allowed us to extract *topic specific social networks* consisting of only those users and edges among users who were interested in a particular topic. We then selected four topic specific networks for our analysis: *Economics*, *Orissa* (a state in India), *Books*, and *Mumbai* (a city in India). Henceforth, any statistics about these clusters will be described in the same order.

5.1.2 Tie classification

Our hypothesis is based on the assumption that the nature of ties being strong or weak is given as a prior. We therefore first classify the ties between people in our dataset as being strong or weak, and then use this classification to validate the hypothesis. We assume that strong and weak ties can be differentiated from each other based on some clustering algorithm. Although significant research exists on the identification of such clusters [36], since we were agnostic to the actual choice of the clustering algorithm, we use the same flow-stochastic graph clustering algorithm [37] used earlier, to cluster the social network graph of all the 42,000 users. This algorithm has a configurable parameter to control the granularity of clustering, and hence produces different clusterings for different parameter values. We choose the parameter value that produces the closest agreement with user surveys, as described next.

We randomly chose {300, 250, 200, 500} users from the four topics respectively, and sent them a personalized survey in which we asked them questions about 5 of their randomly chosen friends with whom they had an explicitly declared reciprocal relationship. We asked these users to rate their five friends on a 5-point scale, giving a score of 1 to an acquaintance and a score of five to a close friend, and to also indicate the frequency of communication with them. This data allowed us to use communication frequency, emotional intensity, and reciprocity of a tie as proxies for the strength of a tie [11]. A total of 314 responses were obtained across the four topics, with ratings for 1,473 links. We then compared these ratings with the classification into strong and weak ties produced by the clustering algorithm.

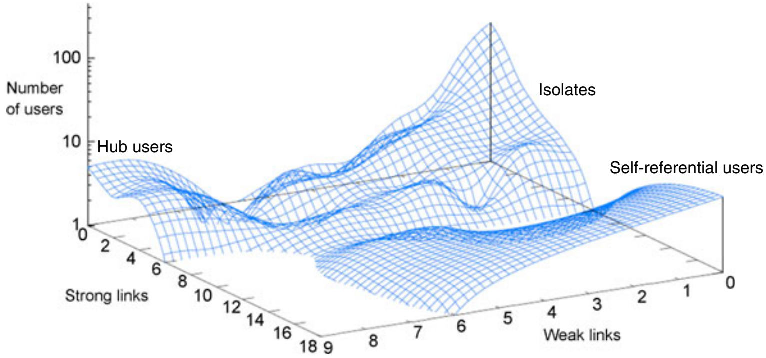


Figure 3 Mass-distribution of strong and weak ties in the topic cluster for Orissa

The best choice of parameter gave a correlation of 0.76 between the classifications produced by the clustering algorithm and the classifications obtained from the user-surveys. The clustering produced by this choice of parameter was used for subsequent analysis. To give a visual explanation, Figure 3 shows the mass-distribution of the number of strong and weak links of users over the population of users interested in the topic of Orissa. It can be seen that a few *self-referential* users, ie. those users who link back to other users in their own cluster, have many strong ties but few weak ties, while other *hub* users have many weak ties. The distribution characteristics in fact depend on the topic; other topic clusters have very different characteristics [32].

5.1.3 Role of social ties

To test this hypothesis, we randomly chose 125 users per topic, and sent them a list of 5 of their friends who were also interested in the same topic. The users were not told which of their friends had been classified as strong or weak by our clustering algorithm. They were only asked to rank their friends on a 5-point scale to assess how much contextual and complete information each friend contributes to the user. We did this by framing a different question for each topic such that it captured the notion of context and completeness that we have defined. For example, we asked the users interested in Orissa to assume that they have to rely on their friends for the latest news about happenings in the state. Then we asked them to rank their friends based on how well the friends knew about their specific interests in Orissa (∼*context*), and how often the friends provided diverse viewpoints about happenings in Orissa (∼*completeness*).

We received replies from {57, 46, 64, 63} users across the 4 topics, with information about {195, 204, 187, 188} links respectively. Each tie was then assigned three labels:

– {*strong, weak*}, given by the clustering algorithm.
– {*provides, does not provide*} context, given by the user surveys.
– {*provides, does not provide*} completeness, given by the user surveys.

Welch t-test We produced two sub-samples of ties: (*strong tie*, {*provides, does not provide*} *context*), and (*weak tie*, {*provides, does not provide*} *context*). We then used the Welch t-test to compare the means of the first and second sub-samples by forming

Table 1 Comparison of four scenarios: {*strong, weak* ties} promote {*context, completeness*}

	Context	Completeness
Strong ties	$\mu = .87, n = 133$ $z = 1.24^{***}$	$\mu = .50, n = 133$ $z = -0.38$
Weak ties	$\mu = .35, n = 71$ $z = -3.95$	$\mu = .70, n = 71$ $z = -0.88^{***}$

the null-hypothesis $\mu_1 = \mu_2$ and the alternative hypothesis $\mu_1 > \mu_2$ [39]. The mean of the first sub-sample was statistically much greater than the mean of the second sub-sample, and confirmed with a p-value < 0.001 (reject the null hypothesis) that strong ties are indeed more likely to provide context than weak ties. In the same way, we produced two sub-samples of (*weak tie*, {*provides, does not provide*} *completeness*), and (*strong tie*, {*provides, does not provide*} *completeness*). Results again confirmed that weak ties are more likely to provide completeness than strong ties. This did not falsify our hypothesis about the relationship between social networks and the context and completeness of messages. We next proceed to analyze the data more closely, to study what proportion of strong and weak ties can be expected to provide context and completeness respectively.

Explicit scenario tests We categorized our samples into four scenarios, {*strong, weak* ties} provide {*context, completeness*}. For each scenario, we performed the z-test by forming the null hypothesis (true mean $\mu = 0.8$) to indicate that at least 80 % of the subjects believe in the scenario with an error-rate of 10 % ($\alpha = 0.1$), and compared it with the alternative hypothesis $\mu < .8$ [39]. The choice of 0.8 as the true-mean is quite subjective, and only reflects an intuition that a majority of the subjects (aka. 80 %) believe some scenario to be true. According to statistical tables, a z-value greater than -1.28 is considered as sufficient evidence to not reject the null-hypothesis. The results are shown in Table 1, and indicate that there is sufficient reason to not reject the claim that more than 80 % of the subjects believe that strong ties provide context and weak ties provide completeness (marked as $***$). The test also succeeds for the scenario that strong ties provide completeness, although the mean is only 0.5, showing that strong ties also provide completeness but to a lesser extent than weak ties. Results from hypothesis tests on other topics are available in [32].

5.2 Evaluation of the credibility model

We evaluate our method over a dataset of ratings by real users obtained from a popular knowledge sharing website, digg.com [23]. The website allows users to submit links to news articles or blogs, which are called *stories* in the terminology used by the website. Other users can vote for these stories; this is known as *digging* the stories. Stories that are *dugg* by a large number of users are promoted to the front-page of the website. In addition, users are allowed to link to other users in the social network. Thus, the dataset provides us with all the information we need:

- Social network of users: We use this information to construct the social network link matrix between users **N[n,n]**. The social network is clustered using MCL, a flow-stochastic graph clustering algorithm [37], to produce classifications of

ties as strong or weak [32]. The cluster of users strongly connected to user u_i is referred to as V_i.
- Stories submitted by various users: This is used to construct the authorship matrix **A[k,n]**. Since all the stories in the dataset were related to technology, we consider them as belonging to a single topic.
- Stories dugg by various users: We use this information to construct the ratings matrix **R[k,n]**. We consider a vote of 1 as an evidence for credibility of the story.

Although the dataset is quite large with over 200 stories, we are able to use only 85 stories which have a sufficiently large number of ratings by a common set of users. This is because we require the same users to rate many stories so that we have enough data to construct training and test datasets for these users. Eventually, we assemble a dataset of 85 stories with ratings by 27 users. A few assumptions we make about the validity of the dataset for our experiments are as follows:

- The submission of a story to Digg may not necessarily be made by the author of the story. However, we regard the submitting user as the message author because it distinguishes this user from other users who only provide further ratings to the messages.
- The ratings provided on the Digg website may not reflect credibility, but rather usefulness ratings given to messages by users. We however consider them to be equivalent to credibility and do not include users who rate more than 65 stories as all credible or all non-credible. We argue that in this pruned dataset, all the users are likely to be interested in the topic and hence all the stories; therefore, the only reason for their not voting for a story would be its credibility.

We use an open-source package, OpenBayes, to program the Bayesian network. We simplify the model by discretizing the evidence variables **E,S,P** into 3 states, and a binary classification for the hidden variables **N, M**, and the credibility variable **C**. The discretization of the evidence variables into 3 states is performed by observing the Cumulative Distribution Frequency (CDF) and Complementary Cumulative Distribution Frequency (CCDF) of each variable with respect to the credibility rating of users. The lower cutoff is chosen such that the product of the CDF for rating=0 and CCDF for rating=1 is maximum, indicating the point at which the evidence variable has a high probability of being 0 and a low probability of being 1. Similarly, the upper cutoff is chosen such that the CCDF for rating=0 and CDF for rating=1 is maximum, indicating the point at which the evidence variable has a low probability of being 0 and a high probability of being 1. This gives a high discrimination ability to the classifier because the cutoffs are selected to maximize the pair-wise correlation of each evidence variable with the credibility rating given by the user.

5.2.1 Choice of parameters

The first set of experiments shown here find good values of α (3) and β (2), and compare ratings with authorship based evidence variable computation (Section 4.4). We evaluate the performance of the model for each user by dividing the 85 stories into a training set of 67 stories and a test set of 17 stories (80 and 20 % of the dataset respectively). We then repeat the process 20 times with different random selections of stories to get confidence bounds for the cross validation. For each evaluation, we use two kinds of performance metrics [6]:

- *Matthew's correlation coefficient (MCC)*:

$$\mathrm{MCC} = \frac{(t_p \times t_n - f_p \times f_n)}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}}$$

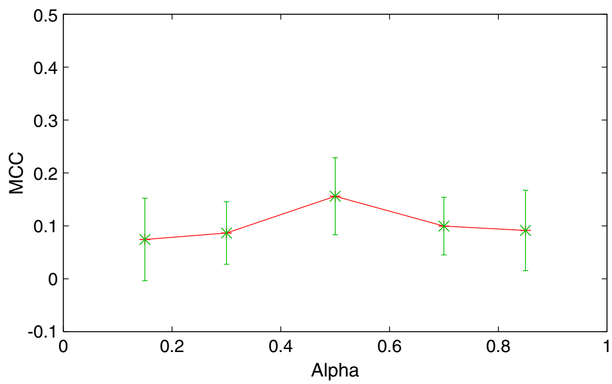
  Here, f_p is the number of false positives, t_p is the number of true positives, f_n is the number of false negatives, and t_n is the number of true negatives. The MCC is a balanced measure for skewed binary classifications, and is convenient because it gives a single metric for the quality of binary classifications.
- *TPR Vs FPR*: This plots on an XY-scale the true positive rate (TPR) with the false positive rate (FPR) of a binary classification. Maximum accuracy implies TPR=1.0 and FPR=0.0, while TPR=FPR is the random baseline. Therefore, points above the random baseline are considered to be good.

Figure 4 shows the mean MCC across all users for different values of α (3) to combine the ratings and social network matrices. The best performance happens at $\alpha = 0.5$, conveying our message that all of authorship, ratings, and social networks provide valuable credibility information. All the experiments are done using ratings-based inference with $\beta = 0.85$ (2). Larger or smaller values of β both give poorer results.

5.2.2 Inference methods

Figure 5 shows the TPR-FPR plot for ratings and authorship based evidence variable computation when $\alpha = 0.5$ and $\beta = 0.85$. As can be seen visually, the ratings-based method performs better than the authorship-based method. The former gives $\mathrm{MCC} = 0.156$ ($\sigma = 0.073$), while the latter gives $\mathrm{MCC} = 0.116$ ($\sigma = 0.068$). However, the authorship performance is still successful for a majority, which is encouraging. This indicates that authorship information may be used to solve the problem of cold-start for new messages that have not acquired a sufficient number of ratings.

Figure 4 Performance with different parameters



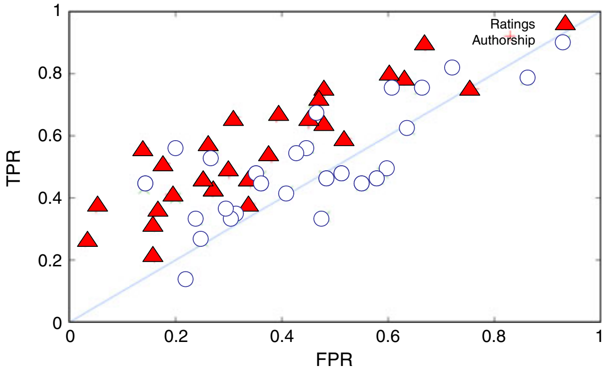


Figure 5 Performance of Bayesian credibility model

Similarly, ratings may be used to solve cold-start for new users who have not acquired sufficient credibility.

We notice that the classifier performs very well for some users, but close to random for some other users. We therefore investigate various characteristics that may prove useful to determine for which users our method may work well and when it may not.

- We compute the variance of cluster and experienced credibility scores for different users. We then compare the variances by good performing users ($\frac{TPR}{FPR} > 1.5$) with the variances by the remaining users. We find that for both cluster and experienced credibilities, the variances by good performing users are more than twice the variances by poorly performing users.
  This shows the more the discrimination produced in the cluster and experienced credibility scores by a user, the better the performance of the user, because greater discrimination ability implies higher entropy in the information theoretic sense.
- We find that on an average, 85 % of users in the same cluster are likely to be all good performing or all poorly performing. This is an interesting result because we also find that users in the same cluster are four times more similar to each other in their credibility ratings than to users in other clusters. Although the similarity of ratings explains why the majority of users also perform similarly, an open question is whether the performance of a user goes up or down because of the cluster in which she is a member, or simply because the ratings given by her are too inconsistent to be captured by the Bayesian model.

As part of future work, we will try to identify more features to classify ratings, authorship, and social network matrices in terms of their characteristics to yield good or bad performance for users.

5.2.3 Comparison with other methods

We next compare our method with other well known methods for trust and reputation computation meant for different applications. All these methods perform very close to random, even with personalization. We believe this to be due to a fundamental drawback of these methods: they try to form an objective assessment of credibility for users and messages, which is not appropriate for participatory media content.

- An Eigenvector computation on $\mathbf{A}_c^T.\mathbf{R}_r$ by leaving out the social network part (1), is identical to the Eigentrust algorithm [19]. The best choice of parameters could only give a performance of MCC $= -0.015$ ($\sigma = 0.062$). Eigentrust has primarily been shown to work in P2P file sharing scenarios to detect malicious users that inject viruses or corrupted data into the network. However, P2P networks require an objective assessment of the trustworthiness of a user, and does not allow for subjective differences, as desired for participatory media.
- An Eigenvector computation on the social network matrix (2), personalized for each user, is identical to the Pagerank algorithm used to rank Internet web pages [26]. However, this too performs poorly with an MCC $= 0.007$ ($\sigma = 0.017$). This suggests that users are influenced not only by their own experiences, but also by the judgement of other users in their cluster, and by public opinion. Methods ignoring these factors may not perform well.
- The beta-reputation system [40] is used in e-commerce environments to detect good or bad buying and selling agents. It estimates the credibility of agents in an objective manner using a probabilistic model based on the beta probability density function. Only the public opinion is considered; ratings are filtered out if they are not in the majority amongst other ratings. It too does not perform well in participatory media environments, giving an MCC $= 0.064$ ($\sigma = 0.062$).

Our conclusion is that our approach which subjectively models credibility using Bayesian networks, allowing users to be influenced in different ways by different sources, performs better than objective modeling approaches that consider a uniform function for credibility across all users.

5.3 Use in recommender systems

As mentioned earlier, our method for credibility computation can be used in two ways to improve recommender systems: (i) Since our method serves to predict the probability of a user finding a message to be credible, it can be used as a pre- or post-filtering stage with existing recommendation algorithms. (ii) As shown in this section, our proposed model can be adapted to integrate closely with recommendation algorithms; we show how to do this with collaborative filtering (CF) [2].

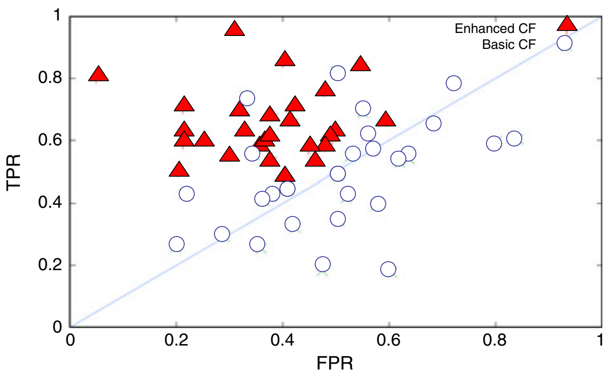
A basic CF algorithm works in two steps. First, similarity coefficients are computed between all pairs of users, based on the similarity of message ratings given by each pair. Second, to make a decision whether or not to recommend a new message to a user, the mean of the message ratings given by other similar users is computed, weighted on the coefficients of similarity to these users. If the mean is greater than a threshold, the message is recommended; else it is rejected.

The drawback of the CF method is that it only learns the average user behavior. However, as we have argued, user behavior can be different in different circumstances. We therefore develop an adaptation of our method. Rather than computing a single similarity coefficient between each pair of users, we compute four similarity coefficients based upon whether messages are believed to be highly contextual by both users, or highly complete by both users, or contextual by the first user and complete by the second user, or vice versa. Essentially, we break down the average user behavior into four components based upon the context and completeness of messages to users, as follows:

- For each user, we run the EM algorithm on the training set to learn the model.
- We use the learned model to infer the probabilities of the hidden variables of context **CN** and completeness **CM** for each story in the training set: $P_i(\mathbf{CN}|\mathbf{E,S,P,C})$ and $P_i(\mathbf{CM}|\mathbf{E,S,P,C})$ shown in Figure 2. That is, for each story m_j, we infer $P(cn_{ji} = 0{,}1|e_{ji}, s_{ji}, p_{ji}, c_{ji})$ and $P(cm_{ji} = 0{,}1|e_{ji}, s_{ji}, p_{ji}, c_{ji})$.
- We then discretize the probabilities for **CN** and **CM** in same way as we did earlier, by finding cutoffs that maximized the product of the CDF for $c_{ji} = 0$ and CCDF for $c_{ji} = 1$. This gives us samples of $(c_{ji} \in \{0, 1\}, cn_{ji} \in \{0, 1\}, cm_{ji} \in \{0, 1\})$, that is, which stories appear contextual or complete to a user, and the rating given by the user to these stories.
- For every pair of users, their samples are then compared to produce four similarity coefficients on how similar the users are in their contextual opinion, completeness opinion, and cross opinions between messages that appear contextual to one user and complete to the other, or vice versa.
- Finally, when evaluating the decision to recommend a test message to a user, the mean of the message ratings is computed over all the four coefficients of similarity, rather than over a single coefficient as in the basic CF algorithm.

Figure 6 shows the performance of the basic CF scheme and our enhanced version. The basic scheme performs worse than random for many users, but when enhanced with breaking up the average user behavior into contextual and completeness components, the performance improves considerably. The mean MCC for the basic scheme is 0.017 ($\sigma = 0.086$), and for the enhanced scheme is 0.278 ($\sigma = 0.077$), a sixteen-fold improvement. We consider this to be a huge improvement over the existing methodologies for trust, reputation, and recommendation algorithms, especially to build applications related to participatory media. Our results reinforce the value of using sociological insights in recommender system design.

Figure 6 Enhancement of collaborative filtering



6 Related work

Our credibility model allows the credibility of messages to be evaluated, makes use of information about social network of users and ratings of messages, and learns for each user a Bayesian model to combine different types of credibilities. In this section, we provide a brief summary of some existing research and point out how they are different from our approach.

Various researchers in the P2P community have focused on Eigenvector based methods to compute the reputation of peers in sharing reliable content [19]. The ratio of successful to unsuccessful content exchanges is computed for each pair of peers who have interacted in the past, and these values are propagated in a distributed manner assuming a transitive trust relationship between peers. However, this is used to only compute the peer reputations (i.e. evaluating users) and not the reliability of content that is shared by the peers. A similar approach of Eigenvector propagation was also used in [27] to compute reputation scores in a blog network, but the reputation of individual blog-entries was not computed. Some other trust propagation methods [7, 14, 15] have also been proposed. For example, on the basis of a social network built from users' direct evaluation on each other, Hang et al. [14] design a new algebraic approach called *CertProp* to propagate trust. Their recent work of the *Shin* approach [15] also considers the difference on trust evaluation towards same peers between users. Fang et al. [7] propose a trust model based on the diffusion theory. The key element of this model is to make use of the social proximity between users in evaluating the trustworthiness of peers. All these methods also only compute peer trustworthiness and assume that trustworthy or reputable peers will always share credible message. In our approach, we make use of message ratings and compute the credibility of each message.

For P2P networks, a method was proposed in [38] where the object reputation is directly calculated to determine whether or not to accept a file being shared on a peer-to-peer network. Transaction history is used to assign edge weights between pairs of peers based on the similarity of ratings given by them to common objects rated in the past. Instead of using Eigenvector propagation to compute an absolute reputation score, a small set of shortest paths is found for each pair of peers, and the relative trust between the peers is computed as the mean of the product of edge weights along the paths. In our approach, we offer a richer multi-dimensional representation, integrating concepts of cluster, experienced and public credibility.

Researchers in the AI community have examined trust models for multi-agent based electronic marketplaces. For example, Zhang and Cohen [43] and Whitby et al. [40] offer systems that determine the trustworthiness of an agent (i.e. a user). In addition, the use of extensive and multi-faceted trust models is promoted in [24, 30], to include features of contextual, role-based and experienced trust. We also have a multi-dimensional model, but we place great emphasis on representing and making use of the social network of a user, in order to learn a user-specific credibility rating for messages.

Other related work in recommender systems that makes use of social networks includes [35, 41, 42]. Song et al. [35] makes recommendations based on stochastic simulations that replicate the observed patterns of information flow on social networks. Yang et al. [41] operates in a P2P setting, and uses decentralized CF algorithms executed within local social network neighborhoods of users. Yu and Singh [42] learns content-based gradients on links between users; this can be used to route messages along desired gradients to users who will be interested in these messages. However, unlike our method which is based on the real-world social network of users, all these methods consider an artificial social network that is formed by linking together users observed to be similar to each other. Furthermore, these methods do not explicitly model message features such as context and completeness. We have already begun to demonstrate the merit of using real social networks of users in delivering recommendations that match well with what real users prefer. As environments where friendship relationships are declared continue to be prevalent (e.g. Facebook [31]), our method for recommending messages will be of particularly great value. Guo et al. [13] propose an approach called *Merge* to make use of real social networks of users to improve the performance of personalized recommendation by particularly addressing the problems of cold start and data sparsity. However, this approach also does not consider the context and completeness of messages. And, the trust propagation mechanism in the *Merge* approach is rather simple.

7 Future work

Confidence bounds Methods for combining trust and confidence have been proposed by researchers such as [16, 22]. For future work, it may be valuable to explore how to incorporate the concept of confidence into our model, for example as a way of placing bounds on the statistical hypotheses that are formed at each processing step.

Model extensions We view our proposed method more as an extensible framework that can be extended to incorporate new insights or information. For example, we could explore the concept of *expert credibility* in the future, for which we would repeat the Eigenvector computations by considering ratings only by a specific set of users categorized as expert users by expert identification algorithms [21]. Another piece of information that is typically available in participatory media content, although it is not available in the digg dataset that we used, is the *message link matrix* based on hyperlinks between messages. An axiom that credible messages link to other credible messages can be modeled through pagerank or HITS, and included as an additional weighting factor in the Eigenvector computations. Alternatively, the polarity between links can be derived by sentiment analysis of the anchor text [17], and distrust propagation methods can be used to produce credibility scores based on the message link matrix [12].

Recommender systems In this paper, we showed how our model can be applied to collaborative filtering. We plan to apply the model to other recommendation algorithms as well, such as a model based algorithm we developed in prior work [33].

Dataset size Given the limited size of our dataset, we have not been able to form significant insights about the size of the training data required for our model to perform well. We will work with larger datasets in the future to understand this aspect in a better way.

Optimized computation The proposed credibility model may be computationally intensive when datasets get larger. However, Eigenvector optimization schemes are available that can decompose a large matrix into smaller matrices, and then combine the components together in an approximate fashion [18]. We will experiment with such schemes in future work.

Robustness to attacks It would be desirable to have our model be robust in the face of attacks by malicious users [44]. This may include scenarios where attackers could add noise to the ratings matrix by giving random ratings to various messages, or attackers could pollute the social network matrix by inviting unsuspecting users to link to them as friends, or even more sophisticated scenarios where attackers could collude with each other. In future work, we would like to examine the robustness of our model against such types of attacks and understand features that can classify ratings, authorship, and social network matrices in terms of their robustness. We also believe that attack analysis could give important insights about the implicit interactions between various pieces of information that are modeled together; such insights are likely to help improve performance.

Additional experimentation The problem of determining which messages are most credible to a user is one which arises in a variety of possible environments. For future research, it would be valuable to replicate our experiments for social networks other than Digg.com. We are currently examining such settings as discussion boards in Massively Open Online Courses (MOOCs), networks to exchange information for patient-led healthcare such as Patients Like Me [4] and the popular advice-providing online meeting place, Reddit.[4]

8 Discussion and conclusions

In this paper, we made use of insights from sociology, political and information science, and HCI, to propose a personalized credibility model for participatory media content. We formulated the model as a Bayesian network that can be learned in a personalized manner for each user, making use of information about the social network of users and ratings given by the users. We showed that our method outperforms both Eigenvector computations and a popular trust modeling system. In addition, we demonstrated that an adaptation of our method to recommendation algorithms such as collaborative filtering (CF) improves the performance of CF. This encourages the use of sociological insights in recommender system research. Our research presents an effective system drawn from the Social Web, to recommend participatory media messages; as such, it promotes social networking and demonstrates the feasibility of leveraging the social graph in forming recommendations.

Our relationship to the Semantic Adaptive Social Web is as follows. The semantic web deepens the labeling of web pages in order to enable users to discover the most valuable pages, due to proper analysis of the content of those pages. In a similar spirit, our research advocates the effective delivery of appropriate content to users, filtering out less relevant choices, by leveraging a deeper understanding of the content. In our case, that deeper understanding is provided by the social network of peers and

[4] http://www.reddit.com/

their assessment of the most credible messages which are then recommended for viewing by the user. Indeed, our discussions of the value of our particular approach in comparison with algorithms such as PageRank clearly explain our focus on enabling users to retrieve appropriate web content. As such, the model presented here not only serves to clarify how best to operate in contexts of social networks but also more specifically how to make use of social networks and message ratings over and above content analysis of web pages, when recommending the information to be processed by each user.

References

1. Adamic, L., Adar, E.: How to search a social network. Soc. Networks **27**(3), 187–203 (2005)
2. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
3. Baybeck, B., Huckfeldt, R.: Urban contexts, spatially dispersed networks, and the diffusion of political information. Political Geography **21**(2), 195–220 (2002)
4. Brownstein, C.A., Brownstein, J.S. III, D.S.W., Wicks, P., Heywood, J.A.: The power of social networking in medicine. Nat. Biotechnol. **27**, 888–890 (2009)
5. Bryant, J., Zillmann, D.: Media Effects: Advances in Theory and Research. Lawrence Erlbaum Associates (2002)
6. Davis, J., Goadrich, M.: The relationship between precision-recall and roc curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240 (2006)
7. Fang, H., Zhang, J., Thalmann, N.: A trust model stemmed from the diffusion theory for opinion evaluation. In: Proceedings of the 12th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2013)
8. Fogg, B.J., Tseng, H.: The elements of computer credibility. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 80–87 (1999)
9. Fortunato, S.: Community detection in graphs. Phys. Rep. **486**(3), 75–174 (2010)
10. Gillmor, D.: We the Media: Grassroots Journalism by the People, for the People. O'Reilly Media (2006)
11. Granovetter, M.S.: The strength of weak ties. Am. J. Sociol. **78**(6), 1360–1380 (1973)
12. Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: Proceedings of the 13th International Conference on World Wide Web (WWW), pp. 403–412 (2004)
13. Guo, G., Zhang, J., Thalmann, D.: A simple but effective method to incorporate trusted neighbors in recommender systems. In: Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization (UMAP) (2012)
14. Hang, C.W., Wang, Y., Singh, M.P.: Operators for propagating trust and their evaluation in social networks. In: Proceedings of the 8th International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), pp. 1025–1032 (2009)
15. Hang, C.W., Zhang, Z., Singh, M.: Generalized trust propagation with limited evidence. Computer (2012). doi:10.1109/MC.2012.116
16. Huynh, T.D., Jennings, N.R., Shadbolt, N.R.: FIRE: an integrated trust and reputation model for open multi-agent systems. In: Proceedings of the 16th European Conference on Artificial Intelligence (ECAI), pp. 18–22 (2004)
17. Kale, A., Karandikar, A., Kolari, P., Java, A., Finin, T., Joshi, A.: Modeling trust and influence in the blogosphere using link polarity. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2007)
18. Kamvar, S.D., Haveliwala, T.H., Manning, C.D., Golub, G.H.: Exploiting the block structure of the web for computing pagerank. Tech. rep., Stanford University (2003)
19. Kamvar, S.D., Schlosser, M.T., Garcia-Molina, H.: The eigentrust algorithm for reputation management in P2P networks. In: Proceedings of the 12th International Conference on World Wide Web (WWW) (2003)
20. Kleinberg, J.: Authoritative sources in a hyperlinked environment. In: Proceedings of ACM-SIAM Symposium on Discrete Algorithms (1998)

21. Kolari, P., Finin, T., Lyons, K., Yesha, Y., Yesha, Y., Perelgut, S., Hawkins, J.: On the structure, properties, and utility of internal corporate blogs. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2007)
22. Kuter, U., Golbeck, J.: SUNNY: a new algorithm for trust inference in social networks using probabilistic confidence models. In: Proceedings of the 22nd AAAI Conference on Artificial Intelligence (2007)
23. Lerman, K.: Social information processing in news aggregation. IEEE Internet Computing **11**(6), 16–28 (2007)
24. Minhas, U.F., Zhang, J., Tran, T., Cohen, R.: A multi-faceted approach to modeling agent trust for effective communication in the application of mobile ad hoc vehicular networks. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. (SMCC) **41**(3), 407–420 (2011)
25. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC) (2007)
26. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. Tech. rep., Stanford University (1999)
27. Pujol, J.M., Sanguesa, R., Delgado, J.: Extracting reputation in multi agent systems by means of social network topology. In: Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2002)
28. Rieh, S.Y.: Judgement of information quality and cognitive authority on the web. J. Am. Soc. Inf. Sci. Technol. **53**(2), 145–161 (2002)
29. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach. Prentice Hall (2009)
30. Sabater, J., Sierra, C.: REGRET: a reputation model for gregarious societies. In: Proceedings of the 5th International Joint Conference on Autonomous Agents (AAMAS) Workshop on Deception, Fraud and Trust in Agent Societies (TRUST), pp. 61–69 (2001)
31. Sakuma, P.: The Future of Facebook. Time. Retrieved on 5 Mar 2008 (2007)
32. Seth, A.: Design of a recommender system for participatory media. Ph.D. thesis, University of Waterloo (2008)
33. Seth, A., Zhang, J.: A social network based approach to personalized recommendation of participatory media content. In: Proceedings of International AAAI Conference on Weblogs and Social Media (2008)
34. Sifry, D.: The State of the Live Web. http://www.sifry.com/alerts/archives/000493.html (2007). Accessed 5 April 2007
35. Song, X., Tseng, B.L., Lin, C.Y., Sun, M.T.: Personalized recommendation driven by information flow. In: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (2006)
36. Tantipathananandh, C., Berger-Wolf, T., Kempe, D.: A framework for community identification in dynamic social networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD) (2007)
37. van Dongen, S.: Mcl: a cluster algorithm for graphs. Ph.D. thesis, University of Utrecht (2000)
38. Walsh, K., Sirer, E.G.: Experience with an object reputation system for Peer-to-Peer filesharing. In: Proceedings of the 3rd Conference on Networked Systems Design and Implementation (NSDI) (2006)
39. Wasserman, L.: All of Statistics: A Concise Course in Statistical Inference. Springer, New York (2004)
40. Whitby, A., Jøsang, A., Indulska, J.: Filtering out unfair ratings in bayesian reputation systems. In: Proceedings of the 3rd International Joint Conference on Autonomous Agenst Systems (AAMAS) Workshop on Trust in Agent Societies (TRUST) (2004)
41. Yang, J., Wang, J., Clements, M., Pouwelse, J., de Vries, A.P., Reinders, M.: An epidemic-based P2P recommender system. In: Proceedings of the ACM SIGIR Workshop on Large Scale Distributed Systems for Information Retrieval (LSDS-IR) (2007)
42. Yu, B., Singh, M.P.: Searching social networks. In: Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2003)
43. Zhang, J., Cohen, R.: A comprehensive approach for sharing semantic web trust ratings. Comput. Intell. **23**(3), 302–319 (2007)
44. Zhang, L., Jiang, S., Zhang, J., Ng, W.K.: Robustness of trust models and combinations for handling unfair ratings. In: Proceedings of the 6th IFIP WG 11.11 International Conference on Trust Management (IFIPTM) (2012)