# SocialTrust++: Building Community-Based Trust in Social Information Systems

James Caverlee, Zhiyuan Cheng, Brian Eoff, Chiao-Fang Hsu,
Krishna Kamath, Said Kashoob, Jeremy Kelley, Elham Khabiri, and Kyumin Lee
Department of Computer Science and Engineering
Texas A&M University
College Station, TX 77845
Contact email: caverlee@cse.tamu.edu

*Abstract*—**Social information systems – popularized by Facebook, Wikipedia, Twitter, and other social websites – are emerging as a powerful new paradigm for distributed social-powered information management. While there has been growing interest in these systems by businesses, government agencies, and universities, there remain important open challenges that must be addressed if the potential of these social systems is to be fully realized. For example, the presence of poor quality users and users intent on manipulating the system can disrupt the quality of socially-powered information and knowledge sharing applications. In this paper, we outline the SocialTrust++ project at Texas A&M University. The overall research goal of the SocialTrust++ project is to develop, analyze, deploy, and test algorithms for building, enabling, and leveraging community-based trust in Social Information Systems. Concretely, we are developing a trustworthy community-based information platform so that each user in a Social Information System can have transparent access to the community's trust perspective to enable more effective and efficient social information access.**

## I. Introduction

The past few years have seen the explosive rise of Web-based social networks, online social media sites, and large-scale information sharing communities – all part of a social computing push that has attracted increasing media, industry, and research interest. Beyond popular successes like Facebook, Wikipedia, YouTube, Delicious, and Twitter, the emergence of *Social Information Systems* is promising to fundamentally transform what information we encounter and digest, how businesses market and engage with their customers, how universities educate and train a new generation of researchers, how healthcare and medical advances are managed and disseminated, how the government investigates terror networks [8], and even how political regimes interact with their citizenry (e.g., the use of Twitter and Facebook in the recent Iranian election controversy [10]). Indeed, both the database and information retrieval communities have recently recognized the immense research challenges inherent in these emerging social systems [2], [6].

One of the key features of Social Information Systems is their reliance on users as primary contributors of content and as annotators and raters of other content. This reliance on users can lead to many positive effects, including large-scale growth in the size and content in the community (e.g., YouTube, Wikipedia), bottom-up discovery of "citizen-experts" with specialized knowledge, serendipitous discovery of new resources beyond the scope and intent of the original system designers, and so on. But the relative openness and self-supervision of many Social Information Systems places great demands on users to ascertain the relative quality and authority of other users in these systems, of the messages passed between users, of resources encountered, of facts asserted, and so on.

At one extreme, malicious adversaries have been observed to exploit the perceived social bonds inherent in Social Information Systems via impersonated (or fraudulent) digital identities [27], targeted malware dissemination [5], social network enhanced phishing [14], and corrupt user-generated metadata (or tags) [21]. Even participants with no malicious intent may hurt the quality of a system by submitting poor quality content and by polluting the system with low value messages and incorrect information. Perhaps most importantly, though, even for non-malicious users and resources of the system, the relative value of these entities is fundamentally tied to the *community-oriented perspective of each user*. For example, an intelligence analyst may assign little value to the latest sports scores but extremely high value to breaking intelligence about a potential military coup. Similarly, a medical researcher may place great value in a scholarly paper written by a chemist, but have little trust for a well-regarded economist. And of course, users may belong to multiple, overlapping communities, placing even greater demands on assessing the relative worth of users and resources in the system.

With these issues in mind, we identify three key features motivating our research on Social Information Systems:

- **Focus on Community:** First, fundamental to these social systems is community – be it friendships on Facebook, groups of similarly-interested users who comment on YouTube videos, collections of Wikipedia contributors who specialize in certain topics, and so on. Whether explicitly declared or implicitly revealed through user actions, *these communities contextualize resources in these systems* (e.g., a right-wing community's view of an Obama web video versus a left-wing community's view) and mark a shift in the balance of power from trusted third-parties (e.g., Google) back to users and their community-oriented perspective.
- **Trust is Key:** Second, users have moved from being pas-

sive consumers of information (via querying or browsing) to being active participants in the creation of data and knowledge artifacts and in the active sorting, ranking, and annotation of other users and artifacts. This shift to users as first-class objects places new demands on providing dependable capabilities for *knowing whom to trust* and *what information to trust*, given the open and unregulated nature of these systems.

- **Community-Powered Opportunities:** Third, the mode of information seeking within these systems often departs from the traditional keyword search paradigm, with social searchers adopting new approaches like tag clouds and social recommendations. Coupled with the challenges of building trust in social systems and exploring community, there is a need for new modes of social information discovery powered through *community-oriented social search and social navigation functionalities*.

Understanding each of these features is an important research challenge on its own and is necessary for fully realizing the vision of Social Information Systems. In concert, the inter-relationships among the three features pose challenging and important questions. For example, how does community impact the trust placed in a user in a social system? How does the trust placed in a user impact community formation? How can we incorporate trust and community into more effective and efficient information discovery algorithms? How does the specific information discovery algorithm impact what resources are trusted and what users are trusted?

In this paper, we outline the SocialTrust++ project at Texas A&M University – a sustained research effort directed at this research framework – which is especially important as these social systems gain traction in businesses, government agencies, and universities. We survey our recent and ongoing research toward building community-based trust and provide some open questions guiding our continuing efforts.

## II. SocialTrust++ Project Overview

The overall research goal of the SocialTrust++ project is to develop, analyze, deploy, and test algorithms for building, enabling, and leveraging community-based trust in Social Information Systems. Concretely, we are developing *a trustworthy community-based information platform so that each user in a Social Information System can have transparent access to the community's trust perspective* to enable more effective and efficient social information access even in the presence of poor quality users and users intent on manipulating the system.

### A. Community-Oriented Trust

As the basis of the SocialTrust++ project, we view a Social Information System $\mathcal{S}$ as consisting of users $\mathcal{U}$, resources $\mathcal{R}$, and metadata $\mathcal{M}$, and the links $\mathcal{L}$ that exist between pairs of entities: $\mathcal{S} =< \mathcal{U}, \mathcal{R}, \mathcal{M}, \mathcal{L} >$. Resources can be images, videos, Web pages, etc. Metadata can be comments, tags (which are typically simple keywords or phrases), ratings (e.g., a thumbs-up or a 3-star rating), etc. The connections across users, resources, and metadata define a social graph, where the linkages indicate the nature of the relationship. For example, a user may be connected to another user via a friend relationship; a resource may be connected to another resource via a hyperlink; a user may be connected to a resource via a comment; and so on. While trust is an overloaded term, for the purpose of this paper, we consider trust as a form of authority measure for users and resources in social systems, much like how PageRank can be considered an authority measure for web pages.

In the overall framework, all users, metadata, and resources can be assigned a community-wide trust score indicating the aggregate trust perspective of the entire community. For example, a user $i \in \mathcal{U}$ may have a community trust rating from community $k$'s perspective at time $t$, denoted by $ST^{[k]}(i,t)$. Hence, for any two users (or resources or metadata), we may evaluate their relative community-based trustworthiness, e.g., that user $i$ is more trustworthy than user $j$ (i.e., $ST^{[k]}(i,t) > ST^{[k]}(j,t)$), from the perspective of each community. This aggregated trust information may be used by users for enhancing the quality of their experiences in the community. Since users will typically have direct relationships with only a small fraction of all users in the network, community-oriented trust values may be used to evaluate the quality of the vast majority of other users (and resources and metadata) for which the user has no direct experience. Over time, these community-oriented trust values can be balanced with each user's personalized experiences. Since users may belong to multiple communities, the community-oriented trust values provide each user with a richer and more robust perspective on entities in the system.

### B. Building Trust with SocialTrust

Concretely, a trust rating for any resource, user, or metadata should be designed with the open and dynamic nature of social systems in mind.[1] For example, a robust trust rating should incorporate features to incent long-term good behavior and to penalize users who build up a good trust rating and suddenly "defect."

In our design of the baseline SocialTrust framework, all users, resources, and metadata are initially treated equally. We support trust maintenance through dynamic revision of trust ratings according to three critical components: the current quality component of trust $Tr_q(i,t)$, the history component, and the adaptation to change component (where, for simplicity in presentation we drop the $k$ community superscript).

$$ST(i,t) = \alpha \cdot Tr_q(i,t) + \beta \cdot \frac{1}{t}\int_0^t I(x) Tr_q(i,x)dx + \gamma \cdot Tr_q'(i,t)$$

[1] Reputation systems are an important feature of many e-marketplaces and online communities (like eBay, Amazon, and Digg), and reputation-based trust systems have received considerable attention in P2P systems (e.g., [1], [15], [23]). Most existing approaches, however, ignore the social constructs and social network topology inherent in online communities, and typically provide less personalized criterion for providing feedback and computing reputations.

where $i$ is a user, resource, or metadata, $Tr'_q(i,t)$ is the derivative of $Tr_q(i,x)$ at $x = t$, where $I(x)$ is an importance function, and where $\alpha$, $\beta$, and $\gamma$ are tunable parameters. This approach is similar to a Proportional-Integral-Derivative (PID) controller used in feedback control systems [24].

In the context of user-based trust, the trust quality component $Tr_q(i,t)$ indicates how well the community believes that user $i$ can be trusted at a point-in-time, but without any consideration of user $i$'s behavior in the past nor any consideration for sudden changes in a user's behavior. Hence, the second and third components of SocialTrust consider the evolution and trajectory of a user's trust rating. The history component – $\frac{1}{t}\int_0^t I(x)Tr_q(i,x)dx$ – considers the integral of the trust value over the lifetime of the user in the network, say, from time 0 to the current time $t$, weighted by an importance function $I$. This history component is important for (i) providing an incentive to all users in the network to behave well over time; and (ii) limiting the ability of malicious participants to whitewash their trust ratings by repeatedly leaving and re-entering the network. The importance function $I$ allows us to optimize the history component by balancing the weight given to more recent periods versus less recent periods. The adaptation to change component – $Tr'_q(i,t)$ – tracks shifts in a user's behavior. This change component can mitigate the impact of malicious participants who build up a good trust rating over time (through the other two components) and suddenly "defect."

Depending on the application domain – e.g., Web-based social network, enterprise information sharing network, etc. – it may be reasonable to adjust the model based on domain knowledge. There are three tunable knobs to balance the current quality component of trust ($\alpha$), the history component ($\beta$), and the change component ($\gamma$). By tuning $\alpha$, $\beta$, and $\gamma$, the SocialTrust model can be optimized along a number of dimensions, e.g., (i) to emphasize the most recent behavior of a user in the network (by choosing higher values of $\alpha$); (ii) to de-emphasize the current user's behavior in the context of his entire history of behavior (by choosing higher values of $\beta$); or (iii) to amplify sudden fluctuations in behavior (by choosing higher values of $\gamma$).

In our initial investigation of community-based trust building in Social Information Systems, we have studied several PageRank-style [25] random walk based trust models that rely on the relationship structure of the social network alone (JCDL'08 [7]). Such random walk models have been studied in both the peer-to-peer file-sharing domain (EigenTrust) [15] and in the context of trust management for the Semantic Web [26]. Over a real social network dataset consisting of 5 million users and 19 million relationship links, we have evaluated the quality of this initial SocialTrust approach over a simple community-based information sharing application. For increasing numbers of malicious users, we have evaluated the quality of information discovered in the community using a simple precision metric. In our preliminary work, the basic SocialTrust approach maintains high quality even as malicious participants increase, as compared to several alternative random walk models. Even when malicious users behave in a strategic fashion by forming cliques, the basic SocialTrust approach performs well.

### C. Open Issues and Challenges

The results presented above, while encouraging, are based on several strong simplifying assumptions and leave open many important questions that we are investigating as part of our next-generation SocialTrust effort (which we dub SocialTrust++):

- What community does a user belong to and how do we find these communities? Community is fundamental to determining the scope of trust building, so that all users and resources may be viewed through multiple community-based lenses, reflecting the values and preferences of each community.
- How do we collaboratively monitor users and resources for determining the current quality component of trust $Tr_q(i,t)$ in the first place? Careful monitoring is necessary so that the SocialTrust++ model accurately and dependably reflects community-based values and preferences.
- What is the relationship between the concrete trust model and how users access and interact with information in the system? Do we find that community-based search and browsing functionality impacts effective trust building and vice versa?
- How robust is the trust framework to strategic malicious behavior by adversarial users? And what other vulnerabilities are there for sustainable community-based trust building?

As part of the SocialTrust++ project, we are engaged in a systematic study of these and related questions (which echo the three key features identified in the introduction of this paper). In the next three sections, we identify and outline our current and ongoing research in: (i) Modeling and mining implicit communities; (ii) Community-based monitoring; and (iii) Community-driven social information access.

## III. Modeling and Mining Implicit Communities

To support our vision of SocialTrust++, we are investigating new approaches for discovering underlying communities of interest that exist in Social Information Systems, in addition to the explicit linkages among users and resources. Our hypothesis is that the observed social graph is "generated" from a core set of underlying latent semantic structures – for example there may exist underlying communities of users, groups of related resources, and so on. In practice, latent communities are *hidden* from us; all we may observe are the artifacts of these underlying groups – e.g., the tags applied to a particular resource or the links between a pair of users. By uncovering these implicit communities, we can establish for each user in the system multiple, overlapping trust groups that reflect each user's community affiliation. These trust groups can serve as the basis for community trust building and guide the development of more effective community-based information exploration in social systems.

Inspired by recent work on text-based topic models, we are developing *probabilistic social structure models* for uncovering these implicit communities. A topic model typically views the words in a text document as belonging to hidden (or "latent") conceptual topics. Prominent examples of latent topic models include Latent Semantic Analysis (LSA) [9], Probabilistic Latent Semantic Analysis (pLSA) [12], and Latent Dirichlet Allocation (LDA) [4]. While these conceptual topic models have classically been applied to text documents, we are investigating new approaches for uncovering latent social structure by (i) optimizing on features unique to social systems, e.g., linkages among users, information resources, and metadata (like tags); and (ii) creating new dynamic community discovery algorithms for efficiently tracking the evolution, volatility, and trajectory of social communities in web-scale datasets.

In our initial study (IEEE SocialCom'09 [18], AAAI Weblogs and Social Media'09 [16]), we have focused on the challenge of modeling and mining community from social tagging systems. These tagging (or annotation) systems are a popular and growing type of Web-based social information systems that aggregate thousands of user's perspectives on web content via simple keywords or phrases that are used to annotate (or "tag") web pages, images, videos, and other web media. For example, the social tagging site Delicious, alone, has led to the annotation of over 150 million web pages by millions of users.

In a collaborative tagging environment, an image of a Tyrannosaurus rex may be annotated by a scientist e.g., with tags like `cretaceous` and `theropod`), by an elementary school student (e.g., with tags like `meat-eater` and `t-rex`) and by a French-speaking tagger (e.g., with tags like `carnivore` and `lézard-tyran`). We view the underlying groups that form around these interests and expertise as distinct *communities*. Concretely, we assume the existence of $L$ distinct communities that are implicit, where each community is a mixture of users that view the world. In our initial design, we model the tags applied to a resource by (potentially) hundreds of authors as a single *social annotation document*. Formally, let $\mathbf{S}_i$ and $\mathbf{c}$ be vectors of length $N_i$ representing $\langle user, tag \rangle$ pairs, and community assignments, respectively, in a social annotation document. The generative process is outlined below:

1) for each community $c = 1, ..., L$
   - Select $U$ dimensional $\tau_c \sim$ Dirichlet($\alpha$)
   - Select $V$ dimensional $\phi_c \sim$ Dirichlet($\gamma$)
2) for each object $\mathbf{S}_i$, $i = 1, ..., D$
   - Select $L$ dimensional $\theta \sim$ Dirichlet($\beta$)
   - For each position $S_{i,j}$, $j = 1, ..., N_i$
     - Select a community $\mathbf{c}_{i,j} \sim$ multinomial($\theta_i$)
     - Select a user $\mathbf{S}^u_{i,j} \sim$ multinomial($\tau_{\mathbf{c}_{i,j}}$)
     - Select a tag $\mathbf{S}^t_{i,j} \sim$ multinomial($\phi_{\mathbf{c}_{i,j}}$)

The generative process creates a social annotation document by sampling for each position $\mathbf{S}_{i,j}$ a community $\mathbf{c}_{i,j}$ from a multinomial distribution with parameter $\theta_i$. A user is then sampled for that position from a multinomial distribution with parameter $\tau_{\mathbf{c}_{i,j}}$ and a tag is sampled from a multinomial distribution with parameter $\phi_{\mathbf{c}_{i,j}}$. This initial generative model naturally encodes the underlying communities of interest and tag-based categories through a generative process, from which we may infer the unobserved community structures by learning model parameters using Gibbs sampling. In our initial investigation (IEEE SocialCom'09, AAAI Weblogs and Social Media'09), we have evaluated the quality of this model over real tagging data collected from CiteULike (for scholarly articles) and Delicious (for Web pages) and have found that it performs well in predicting previously unseen data and in the quality of the user-based communities and tag-based categories.

In our ongoing work, we are investigating several research questions that naturally arise from this initial work, including:

- Adapting the model to new systems, including Twitter-like social messaging systems and Foursquare-like location-based social systems. How generalizable are the probabilistic community discovery algorithms across domains? And what features impact the quality of community discovery?
- Comparing explicit versus implicit structures. For example, do we identify clear implicit communities of users that also reflect explicitly declared group memberships? Or do implicit structures more naturally capture the vibrant activity inherent in social systems?
- Analyzing community evolution and dynamics. As a first step, we are developing a time-sensitive community model that independently generates the social graph at discrete intervals; as we continue, we anticipate augmenting this memoryless model to consider probabilistically linked communities from interval to interval to track community evolution, volatility, and trajectory.

## IV. Community-Based Monitoring

Our second thrust for supporting community-oriented trust building is focused on collaboratively monitoring and assessing the quality of users and resources in a Social Information System. Recall that the SocialTrust++ framework relies on a snapshot trust rating (or authority) for user $i$ at time $t$: $Tr_q(i, t)$. Our goal is to accurately and dependably provide a community-oriented perspective on each user's current authority through a collaborative monitoring framework. By studying how a community can self-regulate, we may gain insights into what a community values and how to sustain the positive growth of the community in the presence of a flood of new users and user-contributed resources. This effort is especially challenging in the context of maintaining up-to-date quality assessments since there are no quality guarantees of published content and since the SocialTrust++ system must update information for each user in near real-time (meaning heavyweight offline analysis of the content and link structure – as in PageRank and many NLP techniques – is infeasible). Compounding the challenge is the subjective and variable nature of community preference. That is, the perceived quality of any piece of information in the system may vary from user to user and from community to community. Dealing with this variation in perceived quality is a difficult and important challenge.

The general approach we take is community-based preference modeling that can predict for each community whether a new user or resource is trustworthy (or authoritative). Toward this goal, we focus on (i) first, identifying and filtering social spammers that degrade the quality of a social system; and (ii) then, building predictive community preference models for estimating how much each community values other users and resources (even in the absence of explicit trust ratings).

For the first task, we have developed a novel honeypot-based approach for uncovering social spammers in online social systems. Social spam is a large and growing problem, however, little is known about social spammers, their level of sophistication, or their strategies and tactics. Filling this need is challenging, especially in social networks consisting of 100s of millions of user profiles (like Facebook, MySpace, Twitter, YouTube, etc.). Traditional techniques for discovering evidence of spam users often rely on costly human-in-the-loop inspection of training data for building spam classifiers; since spammers constantly adapt their strategies and tactics, the learned spam signatures can go stale quickly. Concretely, the honeypot approach is designed to (i) automatically harvest spam profiles from social networking communities, avoiding the drawbacks of burdensome human inspection; (ii) develop robust statistical user models for distinguishing between social spammers and legitimate users; and (iii) actively filter out unknown (including zero-day) spammers based on these user models. Drawing inspiration from security researchers who have used honeypots to observe and analyze malicious activity (e.g., for characterizing malicious hacker activity, generating intrusion detection signatures, and observing email address harvesters), we deploy and maintain *social honeypots* for trapping evidence of spam profile behavior, so that users who are detected by the honeypot have a high likelihood of being a spammer (i.e., low false positive rate).

Even with effective social spam filtering, we still face the challenge of assessing the relative trust (or authority) of users and resources in the Social Information System. Concretely, we consider a set of $d$ entities (e.g., a user or a resource). Each entity $E_i$ is represented by a set of $m$ features: $F_i = \{f_1, f_2, ..., f_m\}$. Each feature refers to some observable characteristic of the entity (e.g., the content of their messages, their social network characteristics, demographics, and so on) or derived feature (e.g., graph centrality over a user's social link structure, part-of-speech tagging, etc.). We assume there exists some training data that has the form: $\{(F_1, r_1), (F_2, r_2)...(F_d, r_d)\} \subset F \times \mathcal{R}$ where the pair $(F_i, r_i)$ corresponds to the feature set and a community preference rating $r_i$ for entity $E_i$. As a baseline, we can train a regression model over this training data. Since the predictive preference models must be evaluated for each entity in near real-time to support effective trust building, we are evaluating the quality of preference predictions over computationally efficient features (e.g., statistical properties of users and their associated content, etc.) versus more expensive features.

In our initial study, we have independently investigated social spam filtering and community preference modeling. First, we have deployed and monitored around 50 social honeypots in MySpace and Twitter (CEAS'08 [28], SIGIR'10 [22]). We find that the deployed social honeypots identify 1000s of social spammers with low false positive rates and that the harvested spam data contains signals that are strongly correlated with observable profile features (e.g., content, friend information, posting patterns, etc.). Based on these profile features, we developed classifiers (e.g., decision tree-based classifiers, SVM) with spam classification accuracy ranging from 90% to 99%. This empirical evaluation shows how the general principles of (i) social honeypot deployment, (ii) robust spam profile generation, and (iii) adaptive and ongoing spam detection can effectively harvest spam profiles and support the automatic generation of spam signatures for detecting new and unknown spam. To test the predictive preference modeling approach (SocialCom'09 [13], AAAI Weblogs and Social Media'08 [19]), we have developed models for learning the *community preference* of short text user-contributed comments, like those found on blogs and Twitter-like services. These short text comments are a rich source of contextual information about web content but in a potentially "messier" form, considering the wide variability in quality, style, and substance of comments generated by a legion of contributors. In this work, we modeled each comment by a number of observable features, including the visibility of the comment, the influence and reputation of the user contributing the comment, and the content of the comment itself. We found that community preference modeling can lead to high-quality identification of interesting and important comments, customized to each community's perspective.

In our ongoing work, we are exploring new avenues for community-oriented monitoring by fully integrating the social honeypot spam filtering approach with community preference modeling. Through effective spam filtering, we can support more efficient and more precise community-based preference models. Some of our ongoing research directions include:

- Augmented spam detection, by exploring how social honeypots can be augmented by other recent approaches to deal with spam in social systems, including Heymann et al. [11] and Benevenuto et al. [3]. These prior approaches have focused on particular communities (e.g., social tagging systems, online video sharing sites); in what ways can their domain-specific techniques be incorporated into the social honeypot approach?
- Expansion and diversification of social honeypots, to both scale up the number of social honeypots (say, to the 1000s) and to consider more variation in the demographics and behaviors of the social honeypot profiles (say, by constructing clique-based social honeypots to measure whether honeypots that are more "connected" induce more spammer activity than "loner" honeypots.).

- Learning implicit preference, so that *explicit* aggregate community ratings that may not always be available (e.g., a discussion thread may not necessarily have star ratings or aggregate thumbs-up/thumbs-down) may be linked to implicit preference signals like page views and user clicks.
- Understanding community differences, to understand what factors most influence overall community preference and to support new techniques for biasing the preference model toward certain sub-communities (e.g., the faculty and student sub-communities may have different preferences for resources within the larger university community).

## V. Community-Driven Social Information Access

Finally, we are investigating new modes of community-driven social search and navigation for enhanced information access in social systems. Instead of guiding users to resources that a user already knows about (e.g., via his own self-managed bookmarks) or that are globally well-known (e.g., via a traditional search engine), we seek to develop new community-based exploration approaches that emphasize the community's implicit view (e.g., to identify resources that are relevant to the implicit *emergency responder* community). Whereas traditional approaches to organizing and accessing the Web's massive amount of information have focused on *content-based* and *hyperlink-based* approaches (e.g., PageRank [25], HITS [20]), these social systems offer rich opportunities for *community-based* exploration and analysis of the Web by building on the unprecedented access to the interests and perspectives of millions of users. As part of this project, we are investigating community-based exploration approaches that emphasize the community's implicit view (e.g., to identify resources that are relevant to the implicit *emergency responder* community). As part of this effort, we are studying the cross-cutting impacts of the discovered implicit communities (Section III) and information quality (Section IV) on the efficiency and effectiveness of community-driven information access. How can our findings in these areas lead to better coverage and more assurance over information and knowledge derived from these systems?

In the previous two sections, we outlined our approach for identifying community and for assessing the relative trust (or authority) of entities in a Social Information System. We now turn our attention to leveraging the discovered community structure to implicitly connect users and resources for more effective community-driven information exploration and discovery.

In our preliminary study, we have leveraged the tag-based community information to enhance exploration of socially tagged documents. The tag-based community discovery algorithm described in Section III results in several discovered distributions: (i) For each community, we have a probability distribution over all users $\tau_c = \{\tau_{c,i}\}_{i=1}^{|U|}$; (ii) For each community, we have a probability distribution over all tags $\phi_c = \{\phi_{c,i}\}_{i=1}^{|T|}$; and (iii) For each resource, we have a probability distribution over communities $\theta_i = \{\theta_{i,j}\}_{j=1}^{L}$. While the possibilities are quite large for applying the discovered community-based information from the model, we have developed and evaluated a query-community ranking approach that maps a user's topical interest (expressed as a query) to resources preferred by communities with a similar topical interest (since each user and resource has an underlying community distribution discovered by the model). In comparison with three state-of-the-art retrieval models: (i) BM25; (ii) Cluster-based retrieval using K-means clustering; and (iii) LDA-based retrieval, we find that the community-oriented ranking model results in a significant improvement over these alternatives (from 7% to 22%) in the quality of retrieved pages (ACM HyperText'10 [17]).

With this initial success in mind, we will explore several related research questions, including:

- User-community ranking, so that in addition to ranking resources by associating queries and communities, we can also explore techniques for personalizing the ranking of resources for each user. Knowing a user's community strength as derived from the implicit community modeling approach, we can favor resources that are most preferred from the user's community, even if the user has never encountered the resource.
- Integrated content and community-based resource exploration, so that users can browse from a candidate resource to other related resources through both the topic-based space (based on the resource text) and the community-based space (based on the implicit connection among resources via community interest).

## VI. Conclusion

We have described the overall research effort driving the SocialTrust++ project at Texas A&M University. The SocialTrust++ project is a comprehensive research effort focused on trustworthy community-oriented social information management, taking a unique three-pronged approach focused on (i) Modeling and mining implicit communities, (ii) Community-based monitoring, and (iii) Community-driven social information access. For more information on the efforts described here, as well as for more recent work, please visit the lab's website at: http://infolab.tamu.edu/.

## References

[1] K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *Proceedings of the ACM 10th Conference on Information and Knowledge Management (CIKM)*, 2001.

[2] R. Agrawal, A. Ailamaki, P. A. Bernstein, E. A. Brewer, M. J. Carey, S. Chaudhuri, A. Doan, D. Florescu, M. J. Franklin, H. Garcia-Molina, J. Gehrke, L. Gruenwald, L. M. Haas, A. Y. Halevy, J. M. Hellerstein, Y. E. Ioannidis, H. F. Korth, D. Kossmann, S. Madden, R. Magoulas, B. C. Ooi, T. O'Reilly, R. Ramakrishnan, S. Sarawagi, M. Stonebraker, A. S. Szalay, and G. Weikum. The claremont report on database research. *SIGMOD Rec.*, 37(3):9–19, 2008.

[3] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Gonçalves. Detecting spammers and content promoters in online video social networks. In *SIGIR*, 2009.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. In *Journal of Machine Learning Research*, volume 3, pages 993–1022, April 2003.

[5] C. Boyd. Teenagers used to push zango on MySpace. http://www.vitalsecurity.org/2006/07/teenagers-used-to-push-zango-on.html, 2006.

[6] J. Callan, J. Allan, C. L. A. Clarke, S. Dumais, D. A. Evans, M. Sanderson, and C. Zhai. Meeting of the minds: an information retrieval research agenda. *SIGIR Forum*, 41(2):25–34, 2007.

[7] J. Caverlee, L. Liu, and S. Webb. Socialtrust: Tamper-resilient trust establishment in online communities. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 2008.

[8] H. Chen. Homeland security data mining using social network analysis. In *EuroISI '08: Proceedings of the 1st European Conference on Intelligence and Security Informatics*, pages 4–4, 2008.

[9] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6):391–407, 1990.

[10] L. Grossman. Iran Protests: Twitter, the Medium of the Movement. *Time Magazine*, Jun 17, 2009.

[11] P. Heymann, G. Koutrika, and H. Garcia-Molina. Fighting spam on social web sites: A survey of approaches and future challenges. *IEEE Internet Computing*, 11(6):36–45, 2007.

[12] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR '99*, pages 50–57, 1999.

[13] C.-F. Hsu, E. Khabiri, and J. Caverlee. Ranking comments on the social web. In *IEEE International Conference on Social Computing (SocialCom)*, 2009.

[14] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, to appear.

[15] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The EigenTrust algorithm for reputation management in p2p networks. In *Proceedings of the Twelth International World Wide Web Conference (WWW)*, 2003.

[16] S. Kashoob, J. Caverlee, and Y. Ding. A categorical model for discovering latent structure in social annotations. In *ICWSM '09: Proceedings of the Third International Conference on Weblogs and Social Media*, Menlo Park, California, USA, 2009. AAAI Press.

[17] S. Kashoob, J. Caverlee, and K. Kamath. Community-based ranking of the social web. In *Hypertext*, 2010.

[18] S. Kashoob, J. Caverlee, and E. Khabiri. Probabilistic generative models of the social annotation process. In *IEEE International Conference on Social Computing (SocialCom)*, 2009.

[19] E. Khabiri, C.-F. Hsu, and J. Caverlee. Analyzing and predicting community preference of socially generated metadata: A case study on comments in the digg community. In *AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2009.

[20] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[21] G. Koutrika, F. A. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *AIRWeb '07: Proceedings of the 3rd international workshop on Adversarial information retrieval on the web*, 2007.

[22] K. Lee, J. Caverlee, and S. Webb. Uncovering social spammers: Social honeypots + machine learning. In *SIGIR*, 2010.

[23] S. Marti and H. Garcia-Molina. Taxonomy of trust: Categorizing p2p reputation systems. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 50(4):472–484, March 2006.

[24] H. Ozbay. *Introduction to feedback control theory*. CRC Press Inc, 1999.

[25] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the Web. Technical report, Stanford University, 1998.

[26] M. Richardson, R. Agrawal, and P. Domingos. Trust management for the semantic Web. In *Web, Proceedings of the Second International Semantic Web Conference*, 2003.

[27] M. Sanchez. Pranksters posting fake profiles on myspace. *http://www.dfw.com/*, 2006.

[28] S. Webb, J. Caverlee, and C. Pu. Social honeypots: Making friends with a spammer near you. In *5th Conference on Email and Anti-Spam (CEAS)*, 2008.