Uncovering the Mystery of Trust in An Online Social Network

Guangchi Liu*, Qing Yang*, Honggang Wang†, Shaoen Wu‡ and Mike P. Wittie*
*Department of Computer Science, Montana State University, Bozeman, MT, USA
†Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, North Dartmouth, MA, USA
‡Department of Computer Science, Ball State University, Muncie, IN, USA

***Abstract*—Trust is a hidden fabric of online social networks (OSNs) that enables online interactions, e.g., online transactions on Ebay. The fundamental properties of trust in OSNs, however, have not been adequately studied yet. In this work, we advance the understanding of trust in OSNs by analyzing the Advogato dataset [1]. We study the properties of direct trust, indirect trust, and trust community detection in Advogato. We found that 1) the trust between users are asymmetric, 2) high degree users are usually associated with high trust, 3) diversity in people's opinions on the same person will affect indirect trust inference, 4) users live in many separate "small small worlds" from the perspective of trust and it is difficult to identify these "small small worlds" with existing random walk-based community detection algorithms, e.g., ACL [2]. It in fact motivates the need for a new community detection algorithm to identify clusters of user connected by trustful relations. Although our findings are from a specific OSN, they can significantly impact how OSNs are designed and configured in the future, e.g., a better user crowdsourcing setting based on trust information.**

***Index Terms*—Online Social Networks, Computational Trust, Three Valued Subjective Logic**

I. Introduction

Online social networks (OSNs) are among the most frequently visited sites on the Internet. Trust is the enabling factor behind user interactions in OSNs. For example, in recommendation and crowdsourcing system, trust helps to identify useful opinions [3], [4]. In Twitter, spams undermine the trust between users by distributing false links [5], hence seriously impacts user experience. Despite its significance, little is known about the trust in OSNs, which motivates us to conduct a comprehensive study of trust in OSNs.

We investigate the fundamental properties of trust in an OSN by looking at direct trust, indirect trust, and trust community detection. For direct trust, we study trust asymmetry, trust assortativity, and the correlation between trust and user's degree in OSNs. For indirect trust, we investigate 1) whether co-citation, coupling, and propagation of trust relations exist in OSNs, 2) how diversity of trust relations affects indirect trust inference, and 3) how users' distance in Advogato network affects their trust. For trust community detection, we evaluate the effectiveness of existing random walk-based community detection algorithms, e.g., ACL [2], in detecting community connected by trustful relations in OSNs.

The model used to quantify and assess trust is the three-valued subjective logic (3VSL) [6]. 3VSL is able to model *direct* and *indirect* trust in OSNs, where direct trust is formed from a user's direct interactions with another user and indirect trust is inferred from others' recommendations or opinions. With the trust model in place, we rely on the dataset [1] derived from an OSN, Advogato.com [7], to advance our study of trust in OSNs. In Advogato, users share their views, ideas or comments about software developments. A user certifies the capability and knowledge of other users, in terms of software development, into various levels based on his/her interactions with them.

Trust was widely studied in many domains including psychology, sociology, management and computer science. A widely accepted definition of trust is: *"Trust is a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviors of another."* [8]. According to this definition, trust in Advogato can be considered a user's psychological state with the intention to accept vulnerability based upon positive expectations of the software development knowledge and capability of another. Here, expectation and vulnerability are integrated together, which is quantified by the certification levels given by a user. Previous work confirms high expectation yields strong intention to accept vulnerability [9], i.e., it is reasonable to only use the expectation component to model trust in Advogato.

Leveraging 3VSL and Advogato dataset, we have the following key findings:

- In Advogato, the mutual trust between two users are asymmetric. However, the difference in the mutual trust is not substantial.
- There is a strong correlation between a user's trust and his/her degree in Advogato, i.e., a user who receives more certifications implies he/she is more trustful.
- Co-citation, coupling and propagation of indirect trust relations are confirmed to be existed in Advogato. Particularly, propagation is the most applicable in Advogato.
- Diversity in people's trust opinions on the same person will impact indirect trust inference.
- From the perspective of trust, Advogato is more meaningfully viewed as many separate "small small worlds" instead of one "small world".
- ACL algorithm achieves a coarse-grained trust community detection in Advogato, which necessities the de-

velopment of new community detection algorithms by considering trust relations between users in OSNs.

Although our study is based on a specific dataset, it is a starting point of understanding the trust in OSNs. We believe more interesting and useful results will be found from other OSN datasets in the future.

II. Related Work

Trust in cloud computing Recently, trust has been introduced in the concept of social cloud. In [10], Mohaisen et al. employ trust as a metric to identify good workers for an outsourcer through her social network. In [11], Moyano et al. proposed a framework to employ trust and reputation for cloud provider selection. In [12], Pietro et al. proposed a multi-round approach called AntiCheetah to dynamically assign tasks to cloud nodes by accounting for their trustworthiness.

Trust in spam detection and Sybil defense One of the domains in which trust analysis is widely applied is the Sybil defense and spam detection [13], [14], [15], [16], [17], [18]. The goal of these works is to identify forged multiple identities and spam information in OSNs. In [13], [14], the basic idea is to employ random walk to rank the neighbors in a given OSN from a seed node, and extract the trust community composed of high ranking nodes. Then, the users outside the trust community will be considered not trustful, i.e., potential Sybil nodes. In [17], Tan et al. integrated traditional Sybil defense techniques with analysis of user-link graph. In [18], Mohaisen, A. et al. proposed a derivation of random walk algorithm, which employs biased random mechanism, to account for trust and other social ties. In [19], besides graph-based features, Yang et al. introduced some other features e.g., neighbor-based, automation-based, etc. to identify spammers. In addition, in [16], [15], spam detection approaches based on user similarity and content analysis are studied.

Trust in recommendation and crowdsourcing systems In addition to Sybil defense in OSNs, trust analysis is found useful in recommendation systems [3], [4], [20]. In [4], Zou et al. proposed a belief propagation algorithm to identify untrustful recommendations generated by spam users. In [3], Basu et al. proposed a privacy preserving trusted social feedback scheme to help users obtain opinions from friends and experts whom they trust. In [20], Andersen et al. proposed a trust-based recommendation system that generates personalized recommendations by aggregating the opinions from other users. In addition, five axioms about trust in a recommendation system are studied.

Different from all these above-mentioned work, our research is the first attempt to give a comprehensive study on the fundamental properties of trust in an OSN rather than apply trust in various applications.

III. Experiment Setup

A. 3VSL

Among the existing trust models (e.g., [21], [22], [23], etc.), 3VSL is proved to be able to accurately (1) model *direct trust* and (2) infer *indirect trust* in OSNs with arbitrary topologies. 3VSL models A's trust on X, where X is called trustee and A is called trustor, as a trust opinion. It is expressed as an opinion vector (b, d, n, e), where b, d, n, e are belief, distrust, posteriori uncertainty, and priori uncertainty. A trust opinion can be further transferred into a single value (ranging from 1 to 0) to represent A's overall trust on B, by 3VSL. The *individual trust* (see Fig. 1) from A to X is a trust opinion obtained from the trust relations between the trustor to the trustee. It can be expressed as a single value by 3VSL, as is introduced before. Similarly, an overall trust opinion from X's neighbors to X itself, called *public trust* (see Fig. 1), can also be expressed as a single value by 3VSL. Individual and public trust should not be confused with direct and indirect trust. As shown in Fig. 1, while individual trust is computed by accounting for both direct or indirect trust from the trustor to the trustee, a trustee's public trust is computed by aggregating all the direct trust from its neighbors. Further details about 3VSL are available in [6].

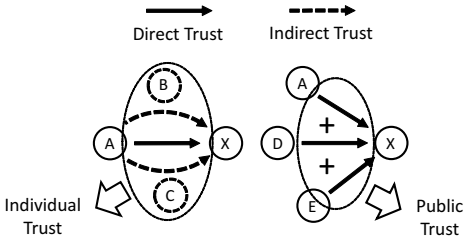


Fig. 1. Individual and public trust. A's individual trust on X can be formed from 1) direct trust and 2) indirect trust, which is inferred from his intermediate friends' (i.e., B and C) opinions. X's public trust can be formed by aggregating his in neighbors' (i.e., A, D and E) direct trust.

B. Dataset Preparation

Trust in Advogato is classified into 4 ordinal levels: *observer*, *apprentice*, *journeyer* and *master*. A level, which is called *trust level*, is given by a trustor individually as a certification to indicate the trustor's direct trust on the trustee. Advogato provides a directed graph where users are nodes and certifications are edges with trust levels. Since the trust levels are in ordinal scales without specifically defined numerical values, a transformation is needed to convert a trust level into a single trust value (ranging from 0 to 1), which has been introduced in section III-A. We employ normal score transformation [24] to achieve this transformation. Firstly, the trust levels are converted to z-scores by normal score transformation based on their distribution in Advogato. Then, these z-scores are mapped to opinion vectors linearly. More specifically, we set the belief components (i.e. b) of observer and master as 0.36 and 0.9 respectively, which indicate poor and excellent interaction histories. Then we assign these two belief components with the z-scores of observer and master, and interpolate the belief components of apprentice and Journeyer as 0.54 and 0.72, according to the distances between the adjacent z-scores. The opinion vector of each trust level can be derived from the belief component as:

$$(b \times (1-e), (1-b) \times (1-e), 0, e) \quad (1)$$

TABLE I
SCALE CONVERSION

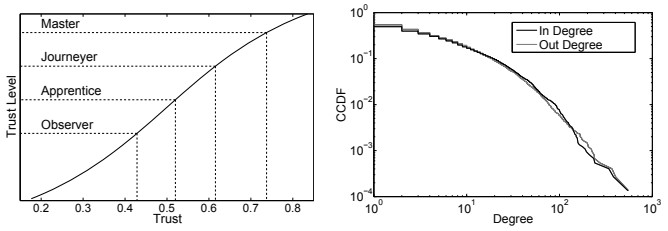
Trust Value	0.43	0.52	0.62	0.74
Belief	0.36	0.54	0.72	0.9
Z-Score	-2.20	-1.11	0.01	1.10
Trust Level	Observer	Apprentice	Journeyer	Master

Finally, the trust value of each trust level is transferred by 3VSL from these opinion vectors. Table I shows the procedure of conversion. Notice that the resulting trust values indicate a non-linear relation with the corresponding trust levels, which can be seen in Fig. 2(a).

The graph we used in this paper is a snapshot taken on 3/16/2014. It consists of 7422 nodes and 56507 edges. Despite its relatively small size, Advogato is an useful and representative dataset in starting a preliminary understanding of trust in OSNs because it is the only publicly accessible dataset that provides an interpersonal relation network with contextual trust information. Notice that, trust should not be confused with reputation despite of the strong relation between them [11]. While reputation is an overall view from the public [11], trust is an expectation and intention from individual experience and preference. Hence, existing rating and reputation datasets, e.g., [25], [26], can hardly be used for trust analysis.

C. Preliminary Statistics

To help grasping the profile of Advogato graph, some preliminary statistics of Advogato are given here. The distribution of trust levels in Advogato are 8.1%, 16.9%, 41.4% and 33.6% for *observer*, *apprentice*, *journeyer* and *master*, respectively. Obviously, the majority of the trust levels lay in journeyer and master, indicating that the overall trust level of users in Advogato is high. The in and out degree distributions of Advogato can be seen in Fig. 2(b), which indicate that both in and out degree range from 1 to 1000. The Spearman's rank correlation coefficient between in and out degree is 0.643, indicating a moderate correlation between in and out degree. In other words, there exists reciprocity in giving and receiving trust certifications. In addition, the global clustering coefficient of advogato is 0.43, indicating that users in Advogato tend to cluster to each other.



(a) (b)

Fig. 2. a) Mapping from trust levels to trust values. b) In and out degree distribution.

IV. DIRECT TRUST

In Advogato, direct trust is formed from accumulative interactions and experience in software development cooperation between a trustor and a trustee. It is expressed as a certification given by the trustor to the trustee. A user's received direct trust is represented as an in coming edge with trust level. On the other hand, a direct trust sent by the user is represented as his out coming edge. Hence, the quantities of received and sent direct trust can be seen as a node's in and out degree. As direct trust is the foundation of indirect trust, in this section we systematically analyze the properties of direct trust in Advogato, including public trust distribution, trust variance, trust assortativity, trust asymmetry, and the relationship between users' public trust and in degree.

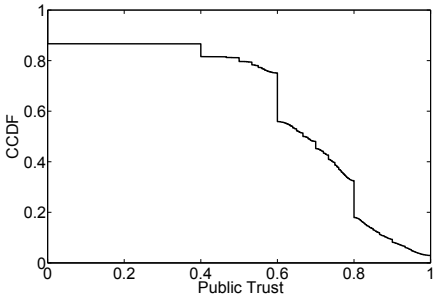


Fig. 3. CCDF of public trust.

A. Public Trust Distribution

We firstly investigate the public trust of users and plot the CCDF of public trust in Advogato, as shown in Fig. 3. Notice that this is different from the trust level distribution discussed before. While trust level is the edge attribute of the Advogato graph, public trust is the overall trust opinion a user received from the neighbors who give him certifications. As shown in Fig. 3, the public trust of users mainly ranges from 0.36 to almost 1. The sharp decrease at 0 indicates the users who have not received certification from others. The other sharp decreases at 0.36, 0.54, 0.72 and 0.9 indicates that considerable quantity of users have received only one certification, which is because of the discrete configuration in Table I

B. Trust Variance

A common phenomenon about trust is that trust opinions from different people on a certain person may be inconsistent. Hence, a question attracted our interests is the inconsistence of user's received direct trust.

We define *trust variance* as the average difference between a user's received direct trust and his/her public trust. We plot the CDF of trust variance in Fig. 4, and also add a null graph as comparison. The topology of the null graph is same as that of Advogato, but its trust levels are randomly assigned from the trust level distribution of Advogato (see section III-C). Compared to the null graph, the trust variance of advogato is smaller. We generated the null graph repeatedly and always get similar results. This indicates that people's direct trust on the same person is inconsistent but not arbitrary as randomly formed in Advogato. A possible reason is that people who give certifications to the same person yield diversity in interactions and subjective preference.

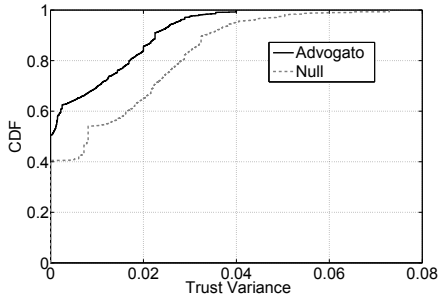


Fig. 4. CDF of trust variance.

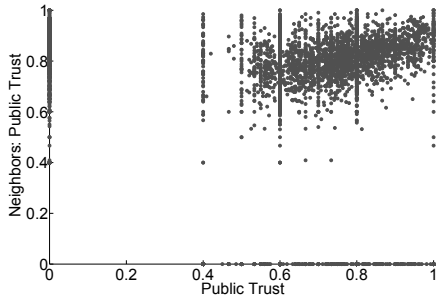


Fig. 5. Trust assortativity. The Spearman's rank correlation coefficient is 0.14. Nodes on the x and y axis are users have not received or given any certification.

C. Assortativity

In traditional OSNs, assortativity is used to quantify how the nodes with same attributes tend to be connected with each other [27]. An interesting question is: does assortativity of trust exist in Advogato? In other words, do users with high public trust tend to connect to each other?

We extend the traditional definition of assortativity to trust assortativity by considering the correlation of public trust between a user and his/her neighbors. We use a scatter plot to show the trust assortativity between Advogato users, as shown in Fig. 5. The x axis is users' public trust and the y axis is the average pubic trust of users' neighbors. We observe that there is a weak correlation between the public trust of a user and his/her neighbors (The Spearman's rank correlation coefficient is 0.14.). We conclude that trust assortativity does not exist in Advogato, i.e., trustful and distrustful Advogato users are mixed together. It indicates that the trust certifications are given without preference in Advogato. A user, no matter how his/her public trust is, can give certifications to anyone else. In other words, Advogato is an open community for people to express their opinions on others.

D. Asymmetry

The mutual trust between two persons are often believed to be asymmetric [21], [28], i.e., if A trusts B, B may not trust A. Some previous works, however, assume the mutual trust is symmetric [13], [14]. Hence, we want to figure out whether they are symmetric or not in Advogato.

Given any two users A and B, we denote A's direct trust on B and B's direct trust on A as T_B^A and T_A^B. Then, we define trust asymmetry between these two trust relations as $|(T_B^A - T_A^B)|/(T_B^A)$. Fig. 6, which is the CDF plot of trust asymmetry in Advogato, shows the results. We also add a

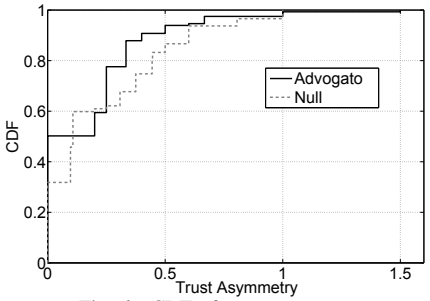


Fig. 6. CDF of trust asymmetry.

null graph aforementioned in IV-B as comparison. Comparing to the null graph, we conclude that *the mutual trust are asymmetric, but the difference is not substantial as randomly formed*. We repeat the experiment of multiple times and always get the similar trend like Fig. 6. A possible reason for this phenomenon is that people who give trust certifications to each other are more likely have frequent interactions, hence will more likely have similar (but inconsistent) capabilities in software development. Combined with the observation in section IV-B, we can see that the distribution of trust relations in Advogato is not arbitrary, which deserves more attentions for the future work in OSNs trust study.

E. Public Trust and In Degree

In e-commerce websites, e.g., Amazon.com and Ebay.com, it is common to see a high rating of a seller accompanied with a large number of reviews. We are interested to know whether there is a correlation between a user's public trust and his/her in degree.

A scatter plot of the relation between the public trust and in degree of nodes is provided in Fig 7. The x and y axis are users' in degree and public trust, respectively. Looking at this figure, we see that when the in degree is low (fewer certifications), the public trust ranges from 0.4 to 1 (a significant fluctuation). On the other hand, as the in degree becomes high, the public trust increases accordingly. The Spearman's rank correlation coefficient between the public trust and in degree is 0.42. Notice that the increase of public trust is because of the overall high trust values on the in edges rather than high in degree. Therefore, we conclude that *there is a moderate correlation between the public trust and in degree*. A possible reason is that a trustful user in Advogato is more likely to get more certifications from others because of his/her rich interactions with others.

V. Indirect Trust

While direct trust is based on the direct interactions between a trustor and trustee, indirect trust is an inference based on the trust relation topology between the trustor and trustee. In fact, indirect trust is complementary to direct trust. When direct trust is unknown or weak, indirect trust can be used to estimate the trust relation between a trustor and a trustee. For example, if Alice trusts Bob who in turn trusts Claire whom Alice does not know herself, Alice may trust Claire based

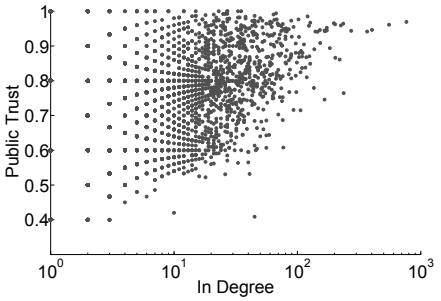


Fig. 7. Correlation between in degree and public trust. The Spearman's rank coefficient is 0.42.

on Bob's opinion on Claire. In this section, we investigate indirect trust properties in Advogato including the existence of various indirect trust relations, impact of the diversity in public's opinions on indirect trust inference and "small world" phenomenon in the perceptive of indirect trust.

A. Co-citation, Coupling and Propagation

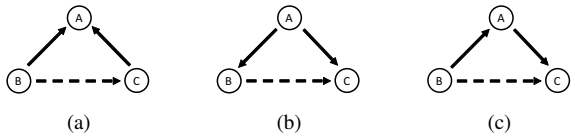
Co-citation, coupling and propagation are the most commonly used relations for indirect trust inference in OSNs [22], [29], [30]. As shown in Fig. 8, co-citation means that if A trusts both B and C, B will likely trust C; coupling means that if both B and C trust A, B will likely trust C; propagation means that if A trusts B and B trusts C, A will likely trust C. To figure out whether these relations are applicable in Advogato, we employ analysis of variance (ANOVA) [31] test to validate them against the Advogato dataset. Given a response variable and multiple independent variables, ANOVA test is used to analyze the effects of the independent variables on the response variable.

From the Advogato dataset, we first randomly pick three nodes (without replacement) that are connected to each other and group them based on the topologies shown in Fig. 8. For each group, we select 1000 samples which are enough to make statistically significant conclusions. Then we set the direct trust of the solid-line edge and dashed-line edge as independent and response variables, respectively. We employ ANOVA tests to evaluate the effect of direct trust of the solid-line edges on the direct trust of dashed-line for the three relations.

TABLE II
STATISTICAL FINDINGS OF CO-CITATION, COUPLING AND PROPAGATION OF TRUST RELATIONS IN ADVOGATO

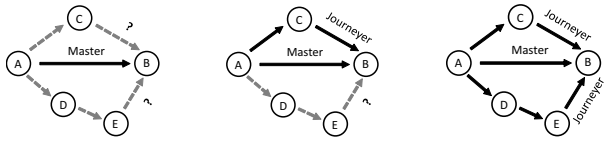
Co-Citation	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(<F)**
E_{BA}	3	33.3	11.09	24.85	<1.67e-15
E_{CA}	3	173.6	57.85	129.67	<2e-16
Residuals	992	442.6	0.45	–	–
Coupling	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(<F)**
E_{AB}	3	58.3	19.433	41.69	<2e-16
E_{AC}	3	88.2	29.409	63.08	<2e-16
Residuals	992	462.5	0.466	–	–
Propagation	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(<F)**
E_{BA}	3	105.0	35.00	117.2	<2e-16
E_{AC}	3	150.7	50.22	168.1	<2e-16
Residuals	992	296.3	0.30	–	–

Table II shows the findings of the ANOVA tests, where E denotes the solid-line edges. We find the F-value of E_{BA} and



(a) (b) (c)

Fig. 8. Co-citation, coupling and propagation.



(a) (b) (c)

Fig. 9. Varying search scope in a social network. Node A and B are trustor and trustee, the rest are intermediate nodes. a) Search neighbors within 1 hop. b) Search neighbors within 2 hops. c) Search neighbors within 3 hops.

E_{AC} in propagation relation are 117.2 and 168.1 respectively. They are both higher than their counterparts in coupling and co-citation relations. This result indicates that the direct trust of both E_{BA} and E_{AC} have significant effects on the direct trust of the dashed-line edge (E_{BC}). Hence, among the three relations, *the propagation model is the most applicable in Advogato*. Notice that the residuals of each relation are very high, indicating a significant effect from other potential variables. The reason is that to investigate the pure topology, we truncate the other edges connecting the three nodes, hence ignored the effect from these edges.

B. Individual Trust: Effect of Scope

When computing the individual trust of a trustor to a trustee, it is important for a trustor to form indirect trust from his/her social network. For example, if A wants to know whether he could trust B, he may refer to his own opinion (direct trust) and the opinions from his/her friends who know B through his/her OSN (indirect trust). Intuitively, the larger scope he search in his/her OSN, the more people who have direct trust on B he will reach and the more complete individual trust on B he can form (see Fig. 9). To validate this intuition, we carry out the following experiments.

First, we randomly select 1000 trustor-trustee pairs which are directly connected. Then, we compute the individual trust of each trustor-trustee pair by running 3VSL with various search scope, e.g., 1 hop, 2 hops and 3 hops, as can be seen in Fig. 9. Then, we compute the public trust of the trustee for each pair, and compare the absolute difference between the individual trust (with various search scope) and public trust.

Fig. 10 plots the CDF of absolute difference between individual trust (with various search scope) and public trust. We see that the larger the search scope, the closer the results approach to the public trust (i.e., smaller absolute difference). At the same time, however, the results can never reach the public trust. This observation indicates that enlarging search scope pushes the computed individual trust closely to the public one. Notice that an 1 hop search scope actually only account for direct trust. Hence, we can conclude that indirect trust adjusts the direct trust to the public trust. In other words,

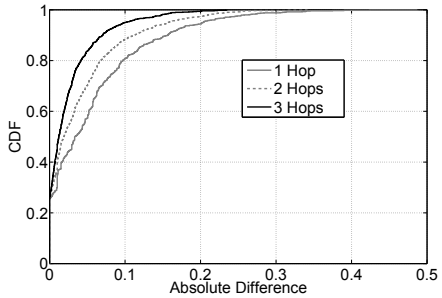


Fig. 10. CDF of difference between public and individual trust, with various search scope.

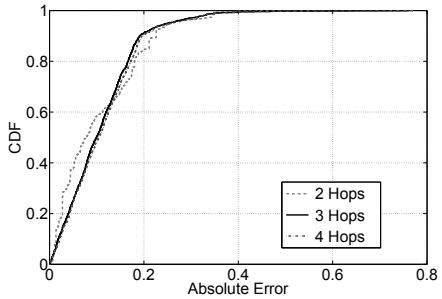


Fig. 11. CDF of error with 2, 3, and 4 hops search scope.

A can adjust his/her direct trust on B by considering other's opinions on B.

C. Individual Trust: Subjective vs. Objective

Due to the trust variance (see section IV-B), it is possible that a trustor's direct trust on a trustee is inconsistent with the trustee's public trust (see Fig. 9(c)). Considering the case when we want to infer a trustor's direct trust (which is unknown) on a trustee through the indirect trust between them, we are interested in whether enlarging search scope, i.e., approaching the public trust, will accurately infer the direct trust.

To answer this question, it is necessary to compare a trustor's computed individual trust and direct trust on a trustee. Hence, we carry out the following experiments. First, we randomly select 1000 trustor-trustee pairs which are directly connected. Then, we remove the direct trust edge and compute the individual trust for each trustor-trustee pair by running 3VSL with various search scope. We define the difference between the computed individual trust and removed direct trust as *error*, which is expressed as $|(T_I - T_D)|/(T_D)$, where T_I is the computed individual trust and T_D is the direct trust.

Fig. 11 plots the CDF of errors with various search scope. From Fig. 11, we can clearly see that more than 60 % the results obtained with a 2 hops search scope have errors less than 0.11. However, the other 40% results contain larger errors than those with 3 or 4 hops search scopes. This implies the results computed with a 2 hops search scope diverges significantly, i.e., they can be either very accurate or imprecise.

To further understand the results, we carried out the following experiments. We identify two groups of data (trustor-trustee pairs):

G1: trustor-trustee pairs whose individual trust is accurate (error < 0.1) with a 2 hops search scope but inaccurate (error > 0.1) with a 4 hops search scope.

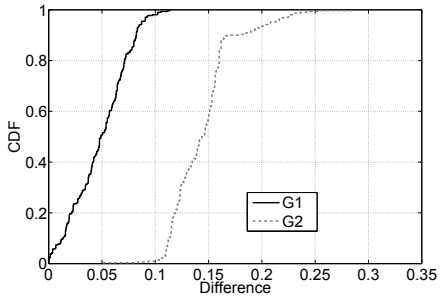


Fig. 12. Difference between individual and public trust.

G2: trustor-trustee pairs whose individual trust is accurate (error < 0.1) with a 4 hops search scope but inaccurate (error > 0.1) with a 2 hops search scope.

Then, we compute the public trust of the trustee for both **G1** and **G2**. Finally, the differences between the public and individual trust, are computed, which are shown as a CDF plot in Fig. 12.

For the trustor-trustee pairs in **G1**, the differences between public and individual trust are much smaller than those in **G2**. That means *if the direct trust is close to the public trust, accounting for larger search scope helps approaching the public trust, as well as the direct trust, hence improves the accuracy. Otherwise, shorter search scope provides more accurate results.*

This observation matches the common sense. Considering in an OSN where the trust variance is small, e.g., the direct trust is estimated based on objective standards, asking for opinions from the public (i.e., larger search scope) yields more accurate results. On the other hand, when the trust variance is large, e.g., the trust is estimated based on subjective preference, asking for opinions from close friends (i.e., smaller search scope) is better.

D. The Small Small World Phenomena

Because Advogato is an OSN, small world phenomena exists in it. In Fig. 13, which plots the distribution of shortest path distance in Advogato, we see the majority of Advogato users are 3 hops away (shortest distance) from each other, and almost all of them are within 6 hops. In this section, we revisit this property by considering whether the small world property holds in Advogato, from the perspective of trust. In other words, even though most users are few hops away from Kevin Bacon, would they trust him? To answer these questions, we analyze the individual trust between users in Advogato as follows.

First, we randomly select trustor-trustee pairs and group them based on their shortest paths ranging from 1 to 6 hops. Within each group, we only keep 1000 pairs. Then, we compute the individual trust from the trustor to the trustee by 3VSL. Finally, we use public trust as the base line.

We plot the results in Fig. 14 which shows CDF of the individual trust of trustees which are $1 - 6$ hops away from the trustor. We see that the longer the distance, the lower the individual trust. If we define a trustee is trustful to a trustor

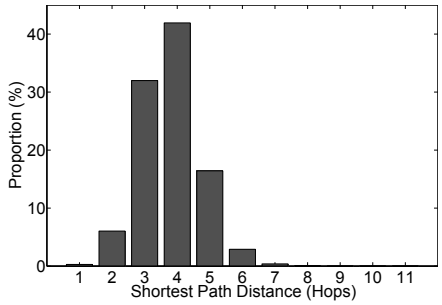


Fig. 13. Distribution of shortest path distance.

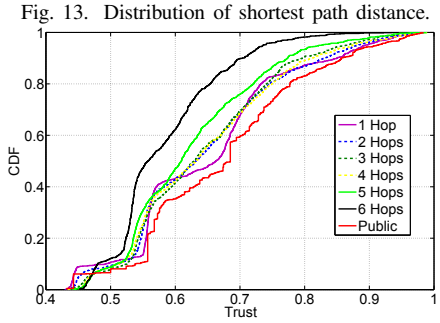


Fig. 14. Individual trust with various shortest path distances.

when its individual trust is greater than 0.7, the portion of "friends" a trustor can trust is around 33% if they are < 4 hops away. The number decreases to 24% and 10% if the trustees are 5 and 6 hops away, respectively. If we pick a trustor and define a "small small world" consisting of all the nodes within 4 hops from the trustor, then we can conclude that *users from a "small small world", rather than the "small world" of Advogato, are likely to be trustful to the trustor*. In other words, we may live in a "small world" where any people can be connected to another one by a chain of acquaintances, but some of these connections are too weak to be useful. In fact, longer the connection, weaker the trustfulness.

VI. Trust Community Detection in OSNs

Because of the fast mixing and clustering property of many OSNs, a lot of random walk-based Sybil defense algorithms were proposed to identify trust community and Sybil users in OSNs. We are hence interested in whether these algorithms can be used to identify the aforementioned " small small world" where users are closely connected and mutually trustful to a given seed node in Advogato.

Among the literature (e.g., [14], [13]), we select the ACL [13] algorithm for its efficiency and accuracy [13]. Because the original ACL algorithm in [13] is applicable to undirected graph only, we use the directed version of ACL [2] to detect the trust community in Advogato. ACL can be seen as a derivation of the Personalized PageRank algorithm. Given a graph, a seed node, a jumpback parameter α and an error parameter ϵ as input, ACL starts from the seed node, moves to a randomly selected neighbor with probability $1-\alpha$ and returns to the seed node with probability α, at each step. If many random walks are performed, the nodes in the "community" to which the seed nodes belongs will be visited most frequently. The weight (i.e., stationary distribution) that

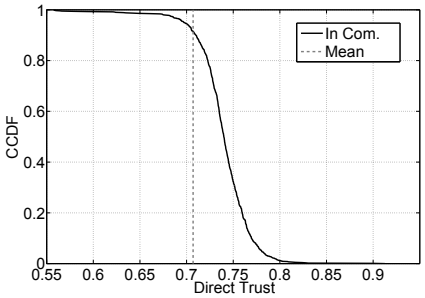


Fig. 15. CCDF of direct trust distribution: In Community vs. Mean.

a node u receives will be proportional to the number of times it is visited when many random walks are performed. Let $p(u)$ and $deg(u)$ denote the stationary distribution of node u and the number of edges pointing to u, respectively. Then, the metric for assessing the trust of node u is $p(u)/deg(u)$, as is defined in [13]. ACL will output a sequence of nodes that in the decreasing order of trust to the seed node. Then the trust community around the seed node can be formed by keeping selecting the top ranking nodes until the minimum conductance is reached [2].

A. Trust Communities Detected by ACL

The very first question we have is: whether the communities detected by the ACL algorithm are trustful or not?

To answer it, we first compute the mean direct trust within a community, and then compare it to the mean direct trust in the entire graph. The experiments are conducted as follows. We first compute the mean direct trust in Advogato. Then we run the ACL algorithm based on randomly selected seed node (from Advogato), which yields a community with the chosen seed node as the center. The α and ϵ are set as 10^{-3} and 10^{-6} according to the configuration in [13]. Then, we compute the mean direct trust within this community, which is treated as the in community trust. Finally, we repeat this process 1000 times and plot the CDF of the in community trust in Fig. 15. We see that the in community trust of 90% communities are higher than the mean direct trust of the entire graph, indicating the effectiveness of ACL in detecting trust communities from Advogato.

B. Trust of Out Community Nodes

So far we know it is possible to extract a community where the in community trust is higher than the mean direct trust of the entire Advogato graph. However, will the seed's individual trust be higher on the nodes within a community (detected by the ACL) than those that are out of this community?

To answer the above-mentioned question, we first randomly select a seed node from Advogato and generate a community around it (through executing ACL algorithm). Then, we randomly pick two nodes: one is inside the community and the other one is outside the community. Finally, we run 3VSL to compute the individual trust from the seed node (trustor) to these two nodes (trustees). Because the individual trust between nodes will decay as their shortest path distance became far (see section V-D), these two nodes must be the

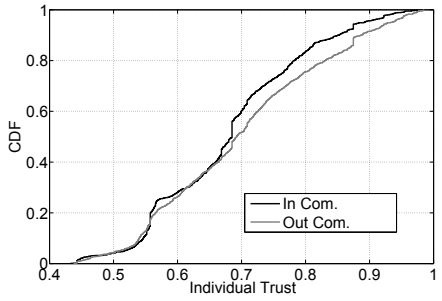


Fig. 16. CDF of individual trust: In Community vs. Out Community.

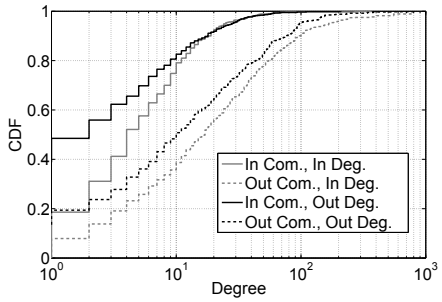


Fig. 17. CDF of degree: In community vs. Out community.

same number of hops (shortest path) away from the seed node. Notice that given this constraint, the nodes near the community boundary will more likely to be chosen as candidate trustees.

After repeating this process 1000 times, we obtain the CDFs of individual trust of nodes inside and outside a community in Fig. 16. We see that when the individual trust is less than 0.69, there are as many in community nodes as out community nodes. Surprisingly, we found when the individual trust is greater than 0.69, there are more out community nodes than in community nodes, i.e., the individual trust of out community nodes are even higher (although not that much) than those inside the community.

This result seems contradictory to the common sense and makes us want to know underlying reasons. As shown in Fig. 17, which plots the CDFs of in degree and out degree for out and in community nodes, both of the in degree and out degree of out community nodes are much higher than those of in community nodes. In other words, out community nodes of high individual trust are often associated with high in degree (see section IV-E), as well as high out degree (see section III-C). However, the ranking of a high degree node is not always higher than a low degree node. In addition, to reach the minimum conductance of a community, high out degree nodes are more likely to be eliminated from a community [2]. Therefore, a high out degree node tends to fall out of a community formed by the ACL algorithm.

In summary, the above-mentioned limitations lead ACL to incorrectly identify those boundary nodes (i.e., false positive). This observation is consistent with the findings in [13]: "Personalized PageRank offers honest nodes a path towards a realistic target for Sybil defense that is more limited than universal coverage but nonetheless useful: a way to white-list trustful nodes that proves efficient and robust in both theory and practice." Therefore, our finding indicates that *existing random walk and conductance-based Sybil defense methods, like ACL, may encounter problems in fine-grained trustful nodes identification, because random walk probability and conductance are not accurate metrics of trust.*

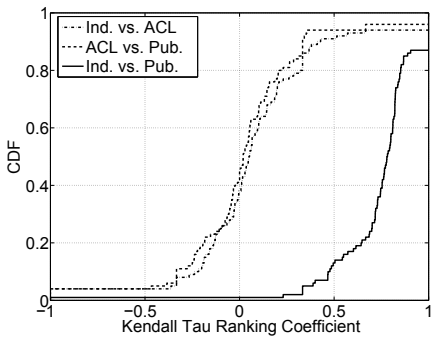


Fig. 18. Kendall tau rank correlation coefficient of trust: ACL, Individual (Ind.) and Public (Pub.)

C. Ranking of In Community Nodes

In the end, we are interested in whether ACL will generate an accurate ranking (based on users' trustworthiness) within a community.

We first randomly select a seed node, then generate a community and the ranking of its in community nodes by ACL. Then we rank the nodes by their individual trust to the seed node. We also consider the ranking of public trust of every node. Then, three rankings (ACL, Individual and Public) are generated. We compare the difference between the ACL and Public ranking. As comparison, we also compare the difference between the Individual and Public ranking. A pairwise Kendall tau coefficient is computed to evaluate the pairwise difference among the three rankings. Finally, we repeat the process 1000 times to make sure our observations are statistically significant.

The CDFs of the three Kendall tau coefficients are plotted in Fig. 18. We see that the coefficients of ACL vs. Individual and ACL vs. Public yield normal distribution with a mean of 0. Specifically, only 10% of the rankings can be seen as moderately correlated (> 0.5). The CDF of Individual vs. Public, however, has an obvious negative skew. We see that around 90% of the rankings are moderately correlated (> 0.5). In summary, ACL does not always accurately rank the users within a community from the perspective of trust. The underlying reason is also related to the phenomena that high degree nodes are treated as least trustful rather than trustful.

In summary, we found that the ACL, which is considered as an efficient random walk-based Sybil defense algorithm in the literature, is an appropriate way of extracting a coarse-grained trust community. However, it is incapable of giving convincing results in further fine-grained trust analysis. This implies that *the random walk and conductance based closeness is not a proper metric for accurate trust assessment in Advogato*. This observation motivates the research community to design new community detection algorithm to accurately identify trust community from the perspective of trust in an OSN like Advogato.

VII. Conclusions

In this work, we uncover the mystery of trust in an OSN called Advogato by looking at the properties of direct and indirect trust, as well as trust community detection.

After analyzing the Advogato dataset by 3VSL, we find that trust in Advogato is 1) asymmetric but without substantial difference; 2) inconsistent but not arbitrarily formed among different people. Interestingly, a strong correlation is identified between trust and node degree, which better explains the common phenomena existing in e-commerce domain that a high rating of a seller is always accompanied with a large number of reviews. Furthermore, we find propagation is the most applicable indirect trust relation in Adovgato. Inconsistent trust opinions on the same person will impact the the accuracy of indirect trust inference. The "small world" phenomena in Advogato can be better expressed as the "small small world" because only 10% of our friends are trustful if they are 6 hops away. In the end, We find applying the ACL algorithm to detect these "small small worlds" is efficient on coarse-grained level. However, ACL has troubles in further fine-grained identification of trustful users.

Although the conclusions of this paper are obtained based on Advogato, it opens a number of research directions for future trust study in other OSNs.

VIII. Acknowledgments

We appreciate our REU students Natalie Pollard from the University of Richmond and Brooke Kelsey from Swarthmore College who enthusiastically participated in this work.

References

[1] Danilo Tomasoni Paolo Massa, Martino Salvetti. Advogato Dataset. http://www.trustlet.org/datasets/advogato/, 2014.

[2] Reid Andersen, Fan Chung, and Kevin Lang. Local partitioning for directed graphs using pagerank. In Anthony Bonato and FanR.K. Chung, editors, *Algorithms and Models for the Web-Graph*, volume 4863 of *Lecture Notes in Computer Science*, pages 166–178. Springer Berlin Heidelberg, 2007.

[3] Anirban Basu, Jaideep Vaidya, Juan Camilo Corena, Shinsaku Kiyomoto, Stephen Marsh, Guibing Guo, Jie Zhang, and Yutaka Miyake. Opinions of people: Factoring in privacy and trust. *SIGAPP Appl. Comput. Rev.*, 14(3):7–21, September 2014.

[4] Jun Zou and Faramarz Fekri. A belief propagation approach for detecting shilling attacks in collaborative filtering. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13, pages 1837–1840, New York, NY, USA, 2013. ACM.

[5] Jinxue Zhang, Rui Zhang, Yanchao Zhang, and Guanhua Yan. On the impact of social botnets for spam distribution and digital-influence manipulation. In *Communications and Network Security (CNS), 2013 IEEE Conference on*, pages 46–54, Oct 2013.

[6] Guangchi Liu, Qing Yang, Honggang Wang, Xiaodong Lin, and M.P. Wittie. Assessment of multi-hop interpersonal trust in social networks by three-valued subjective logic. In *INFOCOM, 2014 Proceedings IEEE*, pages 1698–1706, April 2014.

[7] Raph Levin. Advogato. http://www.advogato.org/, 2014.

[8] Denise M Rousseau, Sim B Sitkin, Ronald S Burt, and Colin Camerer. Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404, 1998.

[9] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. Developing and validating trust measures for e-commerce: An integrative typology. *Information systems research*, 13(3):334–359, 2002.

[10] A. Mohaisen, Huy Tran, A. Chandra, and Yongdae Kim. Trustworthy distributed computing on social networks. *Services Computing, IEEE Transactions on*, 7(3):333–345, July 2014.

[11] Francisco Moyano, Carmen Fernandez-Gago, and Javier Lopez. A framework for enabling trust requirements in social cloud applications. *Requirements Engineering*, 18:321–341, Nov 2013 2013.

[12] Roberto Di Pietro, Flavio Lombardi, Fabio Martinelli, and Daniele Sgandurra. Anticheetah: Trustworthy computing in an outsourced (cheating) environment. *Future Generation Computer Systems*, 48(0):28 – 38, 2015. Special Section: Business and Industry Specific Cloud.

[13] L. Alvisi, A Clement, A Epasto, S. Lattanzi, and A Panconesi. Sok: The evolution of sybil defense via social networks. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 382–396, May 2013.

[14] Wei Wei, Fengyuan Xu, C.C. Tan, and Qun Li. Sybildefender: Defend against sybil attacks in large social networks. In *INFOCOM, 2012 Proceedings IEEE*, pages 1951–1959, March 2012.

[15] Hongyu Gao, Yi Yang, Kai Bu, Yan Chen, Doug Downey, Kathy Lee, and Alok Choudhary. Spam ain't as diverse as it seems: Throttling osn spam with templates underneath. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC '14, pages 76–85, New York, NY, USA, 2014. ACM.

[16] Xia Hu, Jiliang Tang, Yanchao Zhang, and Huan Liu. Social spammer detection in microblogging. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13, pages 2633–2639. AAAI Press, 2013.

[17] Enhua Tan, Lei Guo, Songqing Chen, Xiaodong Zhang, and Yihong Zhao. Unik: Unsupervised social network spam detection. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13, pages 479–488, New York, NY, USA, 2013. ACM.

[18] A. Mohaisen, N. Hopper, and Yongdae Kim. Keep your friends close: Incorporating trust into social network-based sybil defenses. In *INFOCOM, 2011 Proceedings IEEE*, pages 1943–1951, April 2011.

[19] Chao Yang, R. Harkreader, and Guofei Gu. Empirical evaluation and new design for fighting evolving twitter spammers. *Information Forensics and Security, IEEE Transactions on*, 8(8):1280–1293, Aug 2013.

[20] Reid Andersen, Christian Borgs, Jennifer Chayes, Uriel Feige, Abraham Flaxman, Adam Kalai, Vahab Mirrokni, and Moshe Tennenholtz. Trust-based recommendation systems: An axiomatic approach. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 199–208, New York, NY, USA, 2008. ACM.

[21] AUDUN JSANG. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 09(03):279–311, 2001.

[22] Yuan Yao, Hanghang Tong, Xifeng Yan, Feng Xu, and Jian Lu. Matri: A multi-aspect and transitive trust inference model. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, pages 1467–1476, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[23] Carmen Fernandez-Gago, Isaac Agudo, and Javier Lopez. Building trust from context similarity measures. *Computer Standards & Interfaces*, 36(4):792 – 800, 2014. Security in Information Systems: Advances and new Challenges.

[24] Daniel A Powers and Yu Xie. *Statistical methods for categorical data analysis*. Emerald Group Publishing, 2008.

[25] Amazon ratings network dataset – KONECT, August 2014.

[26] Epinions product ratings network dataset – KONECT, August 2014.

[27] M. E. J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89:208701, Oct 2002.

[28] Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07, pages 1551–1556, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[29] R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 403–412, New York, NY, USA, 2004. ACM.

[30] Carmen Fernandez-Gago, Isaac Agudo, and Javier Lopez. Building trust from context similarity measures. *Computer Standards & Interfaces, Special Issue on Security in Information Systems*, 36:792–800, 2014.

[31] Fred Ramsey and Daniel Schafer. *The statistical sleuth: a course in methods of data analysis*. Cengage Learning, 2012.