Keep your friends close: Incorporating trust into social network-based Sybil defenses

Abedelaziz Mohaisen, Nicholas Hopper, and Yongdae Kim
University of Minnesota, Minneapolis, MN 55455, USA
Emails: {mohaisen | hopper | kyd}@cs.umn.edu

***Abstract*—Social network-based Sybil defenses exploit the algorithmic properties of social graphs to infer the extent to which an arbitrary node in such a graph should be trusted. However, these systems do not consider the different amounts of trust represented by different graphs, and different levels of trust between nodes, though trust is being a crucial requirement in these systems. For instance, co-authors in an academic collaboration graph are trusted in a different manner than social friends. Furthermore, some social friends are more trusted than others. However, previous designs for social network-based Sybil defenses have not considered the inherent trust properties of the graphs they use. In this paper we introduce several designs to tune the performance of Sybil defenses by accounting for differential trust in social graphs and modeling these trust values by biasing random walks performed on these graphs. Surprisingly, we find that the cost function, the required length of random walks to accept all honest nodes with overwhelming probability, is much greater in graphs with high trust values, such as co-author graphs, than in graphs with low trust values such as online social networks. We show that this behavior is due to the community structure in high-trust graphs, requiring longer walk to traverse multiple communities. Furthermore, we show that our proposed designs to account for trust, while increase the cost function of graphs with low trust value, decrease the advantage of attacker.**

I. Introduction

The Sybil attack is a well-known and powerful attack in distributed systems, such as sensor networks and peer-to-peer systems. In the basic form of this attack, a peer representing the attacker generates as many identities as she can and acts as if she is multiple peers in the system, which are then utilized to influence the behavior of the system [1]. The number of identities that an attacker can generate depends on the attacker's resources such as bandwidth, memory, and computational power. With the sharp hardware growth—in terms of storage and processing capacities—and the popularity of broadband Internet, even an attacker who uses "commodity" hardware can cause a substantial harm to large systems.

Despite being known for long time, this attack lacked technical defenses and many papers have reported its existence without suggesting any defense while many proposed defenses are limited in many aspects [2]. The majority of defenses proposed in literature to defend against, limit, or mitigate the Sybil attack can be classified into centralized defenses and decentralized defenses. In the centralized defenses (e.g., [1], [3], [4], [5]), a centralized authority is responsible for verifying the identity of every user in the systems. Because they depend on a centralized authority, these defenses are ruled out in many distributed settings. On the other hand, the decentralized defenses (e.g., [6], [7], [8], [9]) utilize collaborative and distributed approaches to bind credentials to the identities of peers, and verify the peers authenticity.

A recent class of the decentralized defenses uses social networks, where peers in the network are not merely computational entities—the human users behind them are tied to each other to construct a social network. The social network is then used for bootstrapping the security and detecting Sybils under two assumptions: algorithmic and sociological. The algorithmic assumption is the existence of a "sparse cut between the Sybil and non-Sybil subgraphs" in the social network, which implies a limited number of attacker edges; edges between Sybil and non-Sybil nodes. The sociological assumption is a constraint on the trust in the underlying social graph: the social graph used in these defenses needs to exhibit strong trust as evidenced, for example, by face-to-face interaction demonstrating social actors' knowledge of each other [9], [10]. While the first assumption has been recently questioned in [11], where it is shown that even honest subgraphs may have cuts that disrupt the algorithmic property, the trust—though being a crucial requirement for these designs to perform well—was not considered carefully. Even worse, many of these defenses [9], [10], [12], [13]—when verified against real-world social networks—have considered samples of online social graphs, which are known to possess weaker value of social trust.

We have recently measured the mixing time, a concrete measure of the algorithmic property required in social networks, in [14], and demonstrated that it is greater than the values used in literature. Also, we pointed out that social graphs with same size have different mixing times implying that social networks, even algorithmically, cannot be taken equally for the purpose of these designs (see sec. V). However, the different mixing times are not arbitrary: social graphs that exhibit knowledge (e.g., co-authorship) or intensive interaction (e.g., social blogs) are slower mixing than social graphs that require less interaction or where edges are less meaningful (e.g., wiki-vote and online social networks such as Orkut and Facebook), which suggest that the algorithmic and trust properties in social graphs are at odds. To this end, we explore designs to model trust in social graphs in order to base the performance of the Sybil defenses more accurately on both assumptions: algorithmic and sociological.

We model the trust exhibited in the social graph as parameters of modified and biased random walks, as opposed

to the uniform random walks used in Sybil defenses—where social graphs are presumed to have similar trust value. The proposed designs use two observations: nodes in the social graph trust themselves more than they trust others, and they trust other nodes unequally. We use the first observation to incur gravitational probability in the random walk – at either the current or originator node of the walk – and use the second observation to incur weights on edges between the different nodes. In the first direction we introduce the lazy and originator-biased random walks. In the second direction we introduce the similarity and interaction-biased random walks to model trust. We investigate their power in modeling trust and influencing the Sybil defenses.

Contributions: The novel and original contributions of this paper are as follows. First, motivated by the observed relationship between the quality of the algorithmic property and hypothesized trust in social graphs, we propose several designs, each in the form of modified random walk, to model trust in social networks. Second, we learn the impact of the different designs on the performance of the Sybil defenses by comparing them to each other when operated on top of SyilLimit, a design for defending against the Sybil attack using social networks. For experiment part, we use several real-world social graphs that exhibit different levels of knowledge between nodes. We provide several insights through discussions that relate to observations on the measurements.

Organization: Some of the related work is reviewed in section II and preliminaries in section III. In section IV we introduce several designs to model trust in social networks, which are used for Sybil defenses. In section V we discuss the main results, which include experiments on real-world social networks. In section VI we present implications of the findings followed by the conclusion and future work in section VII.

II. Related Work

Sybil defenses based on social networks include SybilGuard [10], SybilLimit [9], SybilInfer [12], SumUp [15], and Whānau [13]. In principle, the performance of these defenses depends on the quality of the algorithmic property and assuming strong trust in the underlying social graph. These studies can benefit from our findings in quantifying their performance more accurately by making up for variable trust exhibited in the social graphs which they operate on top of. A study on analyzing these designs can be seen in [11].

Also, several other studies were introduced in the literature on using the trust in social graphs. For instance, Daly et al. [16] used social networks for routing in disconnected delay tolerant networks. In [17], Marti et al. constructed DHTs over social networks. In [18], Pai et al. used the trust in social graphs for bootstrapping trust in ad-hoc networks. The social capital exhibited in social networks is used in [19] to replace the tit-for-tat model in peer-to-peer systems. In all of these studies, trust is considered binary and they can benefit from our findings though our designs are not intended for them.

Understanding, predicting, and analyzing interactions in social networks are studied by Viswanathet al. in [20] and by Wilson et al. in [21]. We use the latter model for our interaction-biased random walk in this study.

Understanding the negative and positive links in social networks – which we can base our designs on – are studied by Leskovec et al. in [22]. The similarity and centralities in social graphs are studied and evaluated in [23], [24], and [25]. All of these studies can be further used to derive similarity metrics, where these metrics can be utilized in our designs.

Influential studies on analyzing the topological structures in online social networks are in [26] and in [27].

Trust in social networks has been also studied since most systems built on top of social networks exploit it – samples of related work on characterizing trust in social networks can be seen in [28] and [29], which are not clear how to use in the context of the problem in hand. Finally, the power of social graphs as good mixers is studied in [30] and [14]. For more detailed exposition on related work, please see [31].

III. Preliminaries

A. Network Model

We view the social network as an undirected unweighted graph $G = (V, E)$ where $|V| = n$, $V = \{v_1, v_2, \ldots, v_n\}$, $|E| = m$, $e_{ij} \in E = v_i \rightarrow v_j$ if $v_i \in V$ is adjacent to $v_j \in V$ for $1 \leq i \leq n$ and $1 \leq j \leq n$. We refer to $\mathbf{A} = [a_{ij}]^{n \times n}$ as the adjacency matrix where $a_{ij} = 1$ if e_{ij} is in E and $a_{ij} = 0$ otherwise. We refer to $\mathbf{P} = [p_{ij}]^{n \times n}$ as the transition matrix

$$p_{ij} = \begin{cases} \frac{1}{\deg(v_i)} & e_{ij} \in E \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $\deg(v_i)$ is the degree v_i, or the row-norm of $\mathbf{A}$:

$$\deg(v_i) = \sum_{k=1}^{n} \mathbf{A}_{ik}. \tag{2}$$

The set of neighbors of v_i is $N(v_i)$ and $\deg(v_i) = |N(v_i)|$.

B. Simple Random Walks and Mixing Time

The "event" of moving from a node to another is captured by a Markov Chain (MC) which represents a random walk over G. A random walk of length w over G is a sequence of vertices in G beginning from an initial node v_i and ending at v_t, the terminal node, using the transition matrix (1). The MC is said to be ergodic if it is irreducible and aperiodic, meaning that it has a unique stationary distribution π and the distribution after random walk of length w converges to π as $w \rightarrow \infty$. The stationary distribution of the MC is a probability distribution that is invariant to the transition matrix $\mathbf{P}$ (i.e., $\pi\mathbf{P} = \pi$). The mixing time of the MC, T is defined as the minimal length of the random walk in order to reach the stationary distribution. More precisely, Definition 1 states the mixing time of MC on G parameterized by a variation distance parameter ϵ.

Definition 1 (Mixing time): The mixing time (parameterized by ϵ) of a Markov chain is defined as

$$T(\epsilon) = \max_i \min\{t : |\pi - \pi^{(i)}\mathbf{P}^t|_1 < \epsilon\}, \tag{3}$$

where π is the stationary distribution, $\pi^{(i)}$ is the initial distribution concentrated at vertex v_i, $\mathbf{P}^t$ is the transition

matrix after t steps, and $|\cdot|_1$ is the total variation distance, which is defined as $\frac{1}{2}\sum_j |\pi_j - \pi_j^{(i)}|$. Notice that $|\cdot|_1$ is at most 1. The MC is rapidly mixing if $T(\epsilon) = \text{poly}(\log n, \log \frac{1}{\epsilon})$. Papers such as [12], [13], [9], [10] refer to this as "fast mixing" and strengthen the definition by considering only the case of $\epsilon = \Theta(\frac{1}{n})$, and requiring $T(\epsilon) = O(\log n)$.

Theorem 1 (Stationary distribution): For undirected unweighted graph G, the stationary distribution of the MC over G is the probability vector $\pi = [\pi_{v_i}]$ where $\pi_{v_i} = \frac{\deg v_i}{2m}$. This is, $\pi = [\frac{\deg(v_1)}{2m}\ \frac{\deg(v_2)}{2m}\ \frac{\deg(v_3)}{2m} \cdots \frac{\deg(v_n)}{2m}]$.

Theorem 2 (Second largest eigenvalue [32]): Let $\mathbf{P}$ be the transition matrix of G with ergodic random walk, and λ_i for $1 \leq i \leq n$ be the eigenvalues of $\mathbf{P}$. Then all of λ_i are real numbers. If we label them in decreasing order, $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_{n-1} \geq \lambda_n > -1$ holds. We define the second largest eigenvalue modulus (SLEM) as $\mu = \max(|\lambda_2|, |\lambda_{n-1}|)$. Then, $T(\epsilon)$ is bounded by $\frac{\mu}{2(1-\mu)}\log(\frac{1}{2\epsilon}) \leq T(\epsilon) \leq \frac{\log(n)+\log(\frac{1}{\epsilon})}{1-\mu}$.

We observe that the mixing time captures the connectivity of the graph. Well-connected graphs have small mixing time while weakly connected graphs have large mixing time [32]. Also, the second largest eigenvalue used for measuring the mixing time bounds the graph conductance, a measure for the community structure [11]. In short, the conductance $\Phi \geq 1-\mu$.

C. Social Network based Sybil Defenses

As mentioned in section II, there are several defenses to the Sybil attack using social networks. Here we limit ourselves to SybilLimit, which we use to measure our designs.

Unlike SybilGuard which uses one long random route for verification, SybilLimit [9] uses several shorter instances of random routes. A verifier as well as the suspect perform $O(\sqrt{m})$ random routes each of length $w = O(\log n)$ to obtain samples of the honest region – since $O(\sqrt{m}) = r_0\sqrt{m}$, SybilLimit fixes $r_0 = 4$ to ensure high intersection probability. The verifier determines to accept a suspect if he is registered at one of the tails in his sample. SybilLimit accepts a suspect if intersection with the verifier happens on a tail, which is the last edge of the random routes. In SybilLimit, if a tail ends up in the Sybil region, it will always end-up in it due to the random routes one-to-one pre-computed permutation structure. Also, if a tail ends up in the Sybil region, it may advertise many non-existent intersections with routes initiated by Sybil nodes. To avoid that, SybilLimit limits the number of intersections into $g \times w \times m$ intersections on honest tails – where g is the number of attack edges and w is the random walk length. This means that SybilLimit accept at most $w = O(\log n)$ Sybil identities per attack edge. SybilLimit greatly depends on w for its security and uses benchmarking techniques for estimating it. However, since these techniques are not provable, underestimating or overestimating the parameters is problematic. SybilLimit works as long as $g \leq o(\frac{n}{\log n})$.

IV. Designs to Account for Social Trust

In most of the literature that considered social networks for building Sybil defenses, the simple uniform random walk highlighted in section III has been used. In this section, we introduce several designs of modified random walks that consider a "trust" parameter between nodes. In all of the proposed modified random walks, the purpose is to assign "trust-driven" weights and thus deviate from uniform. We do this by either capturing the random walk in the originator or current node, as the case of originator-biased and lazy random walks, or by biasing the random walk probability at each node, as the case of interaction and similarity-biased random walks, or a combination of them. The intuition of the lazy and originator-biased random walk is that nodes trust "their own selves" and other nodes within their community more than others. On the other hand, interaction and similarity-biased trust assignments try to weigh the natural social aspect of trust levels. Given the motivation for these designs, we now formalize them by deriving $\mathbf{P}$ and π required for characterizing them. We omit the details for lack of space (see [31] for the complete proofs).

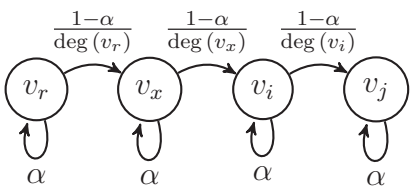


Fig. 1. An illustration of the lazy random walk. For simplicity, α is equal for each node though it can be determined by each node locally to reflect what that node perceive as the trust of the network.

A. Lazy Random Walks

To accommodate for the trust exhibited in the social graph, we assume a global single parameter α in the network which is used to characterize this trust level and used in the different schemes to enforce and apply the trust along with other parameters used (e.g., driven from the algorithmic property in the graph). The transition matrix

$$\mathbf{P}' = \alpha\mathbf{I} + (1-\alpha)\mathbf{P} \tag{4}$$

which yields a transition according to p_{ij} defined as follows:

$$p_{ij} = \begin{cases} \frac{1-\alpha}{\deg(v_i)} & v_j \in N(v_i) \\ \alpha & v_j = v_i \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

We note that for the transition probability defined in (4), by adding self loops it does not alter the final stationary distribution from that in Theorem 1. This is, since $\mathbf{P}' = \alpha\mathbf{I} + (1-\alpha)\mathbf{P}$, by multiplying both sides by π, we get $\pi\mathbf{P}' = \pi(\alpha\mathbf{I} + (1-\alpha)\mathbf{P}) = \alpha\pi\mathbf{I} + (1-\alpha)\pi\mathbf{P} = \alpha\pi + \pi - \alpha\pi = \pi$.

B. Originator-biased Random Walk

We incorporate the concept of biased random on the social graph walks to characterize the bias introduced by the trust among different social actors (nodes). At each time step, each node decides to direct the random walk back towards the node that initiates the random walk, i.e., node v_r, with a fixed probability α or follow the original simple random walk by *uniformly* selecting among its neighbors with the total remaining probability $1-\alpha$. The transition probability that captures the movement of the random walk, initiated by

a random node v_r, and moving from node v_i to node v_j is defined according to p_{ij} as follows

$$p_{ij} = \begin{cases} \alpha & j = r, v_r \notin N(v_i) \\ \alpha + \frac{1-\alpha}{\deg(v_i)} & j = r, v_r \in N(v_i) \\ \frac{1-\alpha}{\deg(v_i)} & j \neq r, v_j \in N(v_i) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

We note that, unlike the lazy random walks, the transition probability here considers moving the state back to the originator of the random walk, a state that may not be connected to the current state in the social graph. This requires a virtual connection between each node through the walk – every node in the graph – and each originator of a random walk. To mathematically model this transition loop, for each node $v_r (1 \leq r \leq n)$, we define $\mathbf{A}_r$ as an all-zero matrix with the exception of the r^{th} row which is 1's. Using $\mathbf{A}_r$, we further define the originator-biased transition matrix, for the walk originated from v_r, as

$$\mathbf{P'} = \alpha \mathbf{A}_r + (1-\alpha)\mathbf{P}. \quad (7)$$

We can show that $\mathbf{P'}$ is stochastic since each row in it sums to 1. Furthermore, since $\mathbf{P'}$ depends on the initial state v_r, we observe that the "stationary" distribution is not unique among all initial states, and so we refer to it as the "bounding distribution" for the walk initiated from v_r. The bounding distribution in that case is $\pi^{(v_r)} = [\pi_i]^{1\times n}$ where π_i is

$$\pi_i = \begin{cases} (1-\alpha)\frac{\deg(v_i)}{2m} & v_i \in V \setminus \{v_r\} \\ \alpha + \frac{\deg(v_i)}{2m} & v_i = v_r \end{cases} \quad (8)$$

We note also that the bounding distribution in (8) is a valid probability distribution since $\alpha + \frac{\deg(v_r)}{2m} + \sum_{v_i \in V\setminus\{v_r\}}(1-\alpha)\frac{\deg(v_i)}{2m} = \alpha + \sum_{i=1}^{n}(1-\alpha)\frac{\deg(v_i)}{2m} = \alpha + (1-\alpha)\sum_{i=1}^{n}\frac{\deg(v_i)}{2m} = \alpha + (1-\alpha) = 1$. It is also easy to show that given distribution bounds the random walk since $\pi\mathbf{P'} = \pi$.

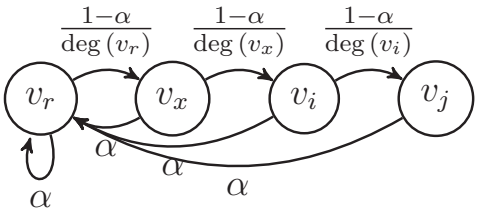


Fig. 2. An illustration of the originator biased random walk.

C. Interaction-biased Random Walk

The interaction between nodes can be used to measure the strength of the social links between nodes in the social network [21]. In this model, high weights are assigned to edges between nodes with high interaction and low weights are assigned to edges between nodes with low interaction. Formally, let $\mathbf{B}$ be the raw interaction measurements between nodes in G and $\mathbf{D}$ be a diagonal matrix representing the row norm of $\mathbf{B}$, computed as in (2). The transition matrix $\mathbf{P}$ of the random walk based on interaction is then computed as $\mathbf{P'} = \mathbf{D}^{-1}\mathbf{B}$. The stationary distribution of the random walk on G following to the probability in $\mathbf{P'}$ is $\pi = [\pi_i]^{1\times n}$ where $\pi_i = (\sum_{j=1}^{n}\sum_{k=1}^{n} b_{jk})^{-1}(\sum_{z=1}^{n} b_{zi})$. We observe that this distribution makes a valid probability distribution since $\sum_{i=1}^{n}\pi_i = 1$ and is a stationary distribution since $\pi\mathbf{P} = \pi$.

Wilson et al. [21] introduced a slightly different model to capture interaction between nodes in the social graph. The interaction graph $G' = (V, E')$ is defined for a social graph $G = (V, E)$ where $E' \subseteq E$ and $e_{ij} \in E'$ if $I(v_i, v_j) \geq \delta$, where I is an interaction measure to assign weights on edges between v_i and v_j for all i, j, and δ is a threshold parameter. The interaction measure used in [21] is the number of interactions over a period of time. This later model further simplifies the random walk where the $\mathbf{P'}$ is defined over G', as well as the stationary distribution.

D. Similarity-biased random walk

The similarity between social nodes in social networks is used for measuring the strength of social links and predicting future interactions [23], [25]. For two nodes v_i and v_j with sets of neighbors $N(v_i)$ and $N(v_j)$, respectively, the similarity is $\frac{N(v_i)\cap N(v_j)}{N(v_i)\cup N(v_j)}$. For $\mathbf{a}_i$ and $\mathbf{a}_j$, two rows in $\mathbf{A}$ corresponding to the entries of v_i and v_j, we use the cosine similarity measure given as $S(v_i, v_j) = \frac{\mathbf{v}_i \cdot \mathbf{v}_j}{|\mathbf{v}_i|_2 |\mathbf{v}_j|_2}$, where $|\cdot|_2$ is the L2-Norm. To avoid disconnected graphs resulting from edge cases, we augment the similarity by adding 1 to the denominator to account for the edge between the nodes. Also, we compute the similarity for adjacent nodes only, so that $\mathbf{S} = [s_{ij}]$ where $s_{ij} = S(v_i, v_j)$ if $v_j \in N(v_i)$ or 0 otherwise. The transition matrix $\mathbf{P}$ of a random walk defined using the similarity is given as $\mathbf{P} = \mathbf{D}^{-1}\mathbf{S}$ where $\mathbf{D}$ is a diagonal matrix with diagonal elements being the row norm of $\mathbf{S}$. Accordingly, the stationary distribution of random walks on G according to $\mathbf{P}$ is $\pi = [\pi_i]^{1\times n}$ where $\pi_i = (\sum_{z=1}^{n} s_{zi})(\sum_{j=1}^{n}\sum_{k=1}^{n} s_{jk})^{-1}.\sum_{i=1}^{n}\pi_i = 1)$.

E. Mixed random walks

It is intuitive and natural to consider a hybrid design that constitutes more than one of the aforementioned random walks. In particular, the interaction and similarity-biased models "rank" different nodes differently and "locally" assign weights to them. Though this limits the mixing time of social graphs as we will see later, it does not provide nodes any authority on the random walk once they are a "past state". On the other hand, benefits of these models are shortcomings in other models. It's hence technically promising and intuitively sound to consider combinations of these designs. Another potential of a mixed design is to use both the lazy and originator-biased random walk in a single walk. As we will see later, in some rapidly mixing social graphs where the underlying social trust is hypothesized to be weak, the lazy random walk poorly captures the behavior of the Sybil defense.

V. Results and Discussion

In this section we outline the results of this study. We first measure the mixing time of the social graphs used in this study (in Table I) and highlight its variable nature among

TABLE I

DATASETS, THEIR SIZE AND THEIR SECOND LARGEST EIGENVALUES OF THE TRANSITION MATRIX. PHYSICS 1, 2, 3 ARE RELATIVITY, HIGH ENERGY AND HIGH ENERGY THEORY CO-AUTHORSHIP RESPECTIVELY [33].

Social network	Nodes	Edges	SLEM
Physics 1 [33]	4,158	13,428	0.998133
Slashdot [34]	82,168	582,533	0.987531
Physics 2 [33]	11,204	117,649	0.998221
Physics 3 [33]	8,638	24,827	0.996879
Wiki-vote [22]	7,066	100,736	0.899418
Enron [33]	33,696	180,811	0.996473
Epinion [35]	75,879	13,428	0.998133
DBLP [36]	614,981	1,155,148	0.997494
Facebook A [21]	1,000,000	20,353,734	0.982477
Facebook B [21]	1,000,000	15,807,563	0.992020
Livejournal A [26]	1,000,000	26,151,771	0.999387
Livejournal B [26]	1,000,000	27,562,349	0.999695
Youtube [26]	1,134,890	2,987,624	0.997972

TABLE II

SOCIAL GRAPHS WITH THEIR SIZE, DIAMETER, AND RADIUS. PHYSICS 1, 2, 3 ARE RELATIVITY, HIGH ENERGY AND HIGH ENERGY THEORY CO-AUTHORSHIP RESPECTIVELY [33].

Social network	Nodes	Edges	Diameter	Radius
Physics 1 [33]	4,158	13,428	17	9
Sdot [34]	10,000	14,6469	6	3
Physics 2 [33]	11,204	117,649	13	7
Physics 3 [33]	8,638	24,827	18	10
Wiki-vote [22]	7,066	100,736	7	4
Enron [33]	10,000	108,373	4	2
Epinion [35]	10,000	210,173	4	2
DBLP [36]	10,000	20,684	8	4
Facebook [21]	10,000	81,460	4	2
Livejournal [26]	10,000	135,633	6	3
Youtube [26]	10,000	58,362	4	2
Rice-cs-grad [37]	501	3255	9	5
Rice-cs-ugrad [37]	1221	43153	6	3

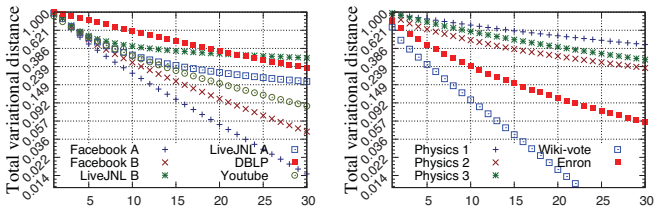
networks with similar size. We follow this by examining the impact of using proposed models on the mixing time and the performance of SybilLimit, a well-known Sybil defense. We limit ourself to this defense scheme though our conclusions apply to all other schemes that using the mixing time as the underlying property for their performance.

A. Social graphs—the datasets

The social graphs used in our experiments are in Table I. These graphs are carefully selected to feature different models of knowledge between nodes in the social networks. These networks are categorized as follows. (1) social networks that exhibit knowledge between nodes and are good for the trust assumptions of the Sybil defenses; e.g., physics co-authorships and DBLP. These are slow mixing (see Fig. 3). (2) Graphs of networks that may not require face-to-face knowledge but require interaction; e.g., Youtube and Livejournal, which are slow mixing, but faster than the first category. (3) Datasets that may not require prior knowledge between nodes or where the social links between nodes are less meaningful to the context of the Sybil defenses; e.g., Facebook and wiki-vote, which are shown to be very fast mixing. While these graphs are used for demonstrating the first part of the results, measuring the performance of SybilLimit and the impact of our designs on the mixing time is done over samples of these graphs. For feasibility reasons, we sample only 10K nodes, using the breadth-first search algorithm, from each graph larger than 10K in Table I. The resulting sub-graphs are in Table II. The diameter is the maximal eccentricity (set of maximal shortest paths from each source in the graph) and the radius is the minimal eccentricity. We compute them to show some insight on the structure of the graphs. For Facebook and Livejoural datasets, the sub-graphs are from dataset A of each.

B. Measuring the mixing time

While measuring the mixing time using SLEM as explained in section III requires computing μ, the computed mixing time might be an overestimation for quality which is necessary in the Sybil defenses. In principle, the overestimation occurs because the computed mixing time using SLEM is the maximal, where a few outlier nodes may capture the mixing time of the entire graph, while the majority of nodes may have relatively smaller mixing time than these outliers [14]. For that, we limit ourselves to measuring the mixing time using Definition 1, and considering a few initial distributions. We classify graphs, shown in Table I, based on their size into large ($> 600,000$ nodes) and small ($< 100,000$ nodes) graphs. For each social graph, we compute the mixing time according to Definition 1 for a sample of $1,000$ initial distributions (nodes). We then compute the total variation distance for a given walk length w as the *average* distance among the $1,000$ nodes. The results are shown in Fig. 3. In short, two things to observe from these measurements [14]. First, the mixing time is larger than used in literature (e.g., 10 to 15 in [8], [9] for 10^6-node graphs). For example, for $\epsilon \approx 1/4$, which is required for $\approx 99\%$ admission rate in SybilLimit, $w = 30$ is required in Physics 1. Second, we observe that the mixing time is variable among social graphs with similar size where graphs with meaningful edges are slower mixing than others with less meaningful links.



(a) Large datasets (b) Small/meduim datasets

Fig. 3. The average mixing time of a sample of 1000 initial distributions in each graph in Table I using the sampling method for computing the mixing time by its definition over $\mathbf{P}$.

C. Implication of the designs on the mixing time

Along with the simple random walk-based design, we implement three of the proposed designs: lazy, originator, and similarity biased random walks. We use the simple random walk-based implementation over the interaction graph of Wilson et al.'s [21] to learn the performance of the interaction-based model. We examine the impact of each design on the mixing time on some graphs from Table II. The results are shown in Fig. 4 and Fig. 5. We observe that, while they bound

the mixing time of the different social graphs, the originator-biased random walk is too sensitive even to a small α. For example, as in Fig. 5(a) for Facebook social graph in Table II, $\epsilon \approx 1/4$ is realizable at $w = 6$ with the simple random walk, $w > 10$ for both lazy and originator-biased random walk. However, this happens with $\alpha = 0.5$ in the lazy against $\alpha \approx 0.1$ in the originator-biased walk. This observation is made clearer on Fig. 5 which compares the mixing time of four different social graphs with different characteristics when using the simple and modified random walks.

We also observe in Fig. 4 and Fig. 5 that the linear increments in the parameters do not necessarily have linear effect on the measured mixing time. Furthermore, this behavior is made clearer in the experiments performed on SybilLimit and shown in Fig. 6 and Fig. 7. This however is not surprising, at least with the originator-biased random walk since the probability of intersection when sampling from the stationary distribution is $\leq e^{-8(1-\alpha)^4}$ [31] from which one can see the exponential effect of α on the admission rate. While this explains the general tendency in the admission rates of SybilLimit, it does not answer some inconsistency shown in Fig. 7(b) for the transition between $\alpha = 0.12, 0.16$, and 0.20. One additional explanation for that is the community structure in this graph, which is shown in [11] to be clear in Physics 1 and problematic for Sybil defenses (results for the same graph are in Fig. 6(b) and Fig. 7(b)). On the other hand, some graphs are less sensitive to the same value of these parameters, e.g., Facebook with the results shown in figures 4(a), 5(a), 6(d), and 7(d). One possible explanation for this behavior is that this graph has less community structure. Reasoning about this behavior and its quantification is to be our future work.

D. Sybil defense performance over simple random walks

To understand the necessary mixing time quality required for the operation of SybilLimit, we measure the performance of SybilLimit using simple random walks, where the evaluation metric is the percent of honest nodes accepted by other honest nodes. For each walk with length $w(0 \leq w \leq 30)$, we compute the number of accepted nodes as a percent out of $n(n-1)$—total verifier/suspect pairs. Since SybilLimit accepts nodes on edges only, it works for $w \geq 2$. The results are shown in Fig. 8 and the variable mixing time shown earlier is further highlighted by observing the percent of accepted nodes when varying w. We observe that, unlike claims in SybilLimit where one would expect 95% admission rate at $w = 4$, some graphs require $w = 30$; where graphs which admit high percent of nodes for small w are those with poor trust.

E. Sybil defense performance over modified random walks

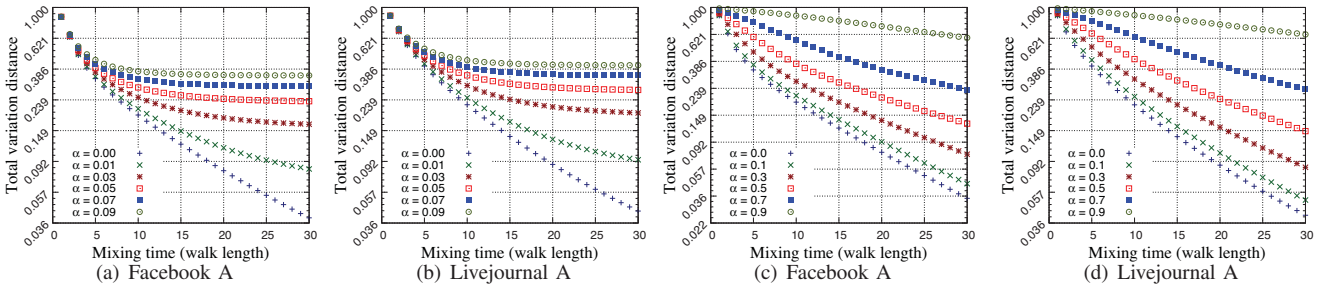
Now we study the impact of the modified random walks on the performance of SybilLimit. We select four datasets with different characteristics from Table II: DBLP, Facebook, Facebook (Rice grad), and Physics 1 (relativity theory). We implement modified SybilLimit versions that consider changes introduced by the modified random walks and test the admission rate of honest nodes under different values of α and w.

1) Performance over lazy random walk: we measure the performance of SybilLimit operating with the lazy random walks – results are shown in Fig. 6. We vary w from 0 to 30 with steps of 2. We further vary α associated with the lazy random walk from 0 to 0.80 with steps of 0.16—$\alpha = 0$ means simple random walk. While the performance of SybilLimit is generally degraded when increasing α, we observe that the amount of degradation varies and depends on the initial quality of the graph. For example, by comparing DBLP (Fig. 6(c)) to Facebook (Fig. 6(d)) we observe that for $w = 10$, DBLP and Facebook admit about 97% and 100% of the honest nodes respectively for $\alpha = 0$. For the same w and $\alpha = 0.64$, the accepted nodes in Facebook are still close to 100% while the accepted nodes in DBLP are only 50% suggesting variable sensitivity of different graphs to same α. Once we raise α to 0.80, the number of accepted nodes in Facebook decreases to 80% while giving only 25% in DBLP. One explanation of this behavior is what we have discussed in section V-C. Also, since the ultimiate goal of this model is to characterize trust, which already differs in these graphs, we know that α should not necessarily be equal in both cases. For instance, if one is concerned about achieving same admission rate for the same w in both cases, one may choose $\alpha = 0.48$ in DBLP and $\alpha = 0.80$ in Facebook where $w = 10$ in both cases which yields 80% admission rate in both cases.

2) Performance over originator-biased random walk: The same settings in section V-E1 are used in this experiment but here we vary α from 0 to 0.2 with 0.02 steps since the originator-biased walk is more sensitive to smaller α than the lazy-random walk. Similar to the lazy walk, the originator-biased walk, as shown in Fig. 7, influences the performance of SybilLimit on different graphs differently, and depending on the underlying graph. However, two differences are specific to the originator-biased walk over the lazy random walk.

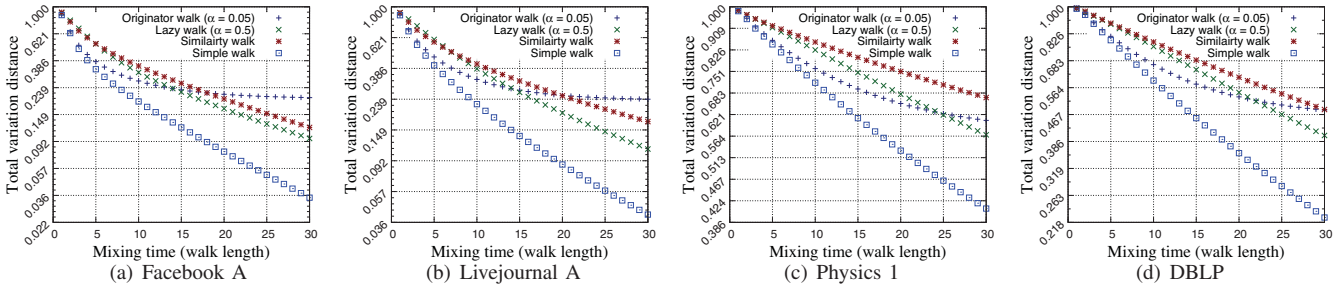
First, the insensitivity shown earlier is even clearer in the originator-biased model. Second, while the end result of SybilLimit operating with lazy random walk is identical to the simple random walk if one allows long enough walk to compensate for the laziness, the behavior of the originator-biased walk is different. The indirect implication of the originator-assigned probability to herself is "discontinuity" in the graph (with respect to each node), where each node gives up some of the network by not trusting nodes in it. To cover the whole graph with that same α, w needs to be exponentially large. To challenge the insensitivity of the fast mixing social graphs, we extend α beyond the values used in Fig. 7 with Facebook from Table II and use $\alpha(0 \leq \alpha \leq 0.5)$ with 0.1 steps and compute the admission rate. The result shows (not included here) that the originator-biased walk limits the number of accepted nodes, even in fast mixing graphs, but for larger α.

3) Performance over similarity and interaction-biased walk: The similarity and interaction-biased random walks as used in this paper are unparameterized. We compute the similarity for Facebook in Table II, as explained in IV-D. The similarity is then used to assign weights to edges between nodes, and bias the transition matrix. We run SybilLimit with



(a) Facebook A (b) Livejournal A (c) Facebook A (d) Livejournal A

Fig. 4. The impact of the originator and lazy walks on the mixing time—(a) and (b) are for originator-biased while (c) and (d) are for lazy random walks.



(a) Facebook A (b) Livejournal A (c) Physics 1 (d) DBLP

Fig. 5. The mixing time of four different social graphs when using simple vs. lazy, originator, and similarity-biased random walks, for each graph. While they are similar in size, a mixing time (parameterized by the same ϵ) is variable.

similarity-biased random walks on Facebook in Table II, where the result is shown in Fig. 9. In short, the similarity – while expected to capture some truth about the underlying graph – has less influence on the behavior of SybilLimit. It is however worth noting that the impact of the similarity-biased random walk is clearer on other social graphs, such as DBLP and Physics, which have clearer community structures.

For the interaction-biased design, we borrow the interaction graph of Wilson et al. [21] on Facebook (same dataset in Table II). The interaction model introduces a richer model than the mere connections between nodes: it shows how strong are the links between nodes in the graph. With the same settings as earlier, we run SybilLimit – as a simple random walks – over the interaction graph. The results are shown in Fig. 9.

F. All designs: comparative study

Finally, we consider all designs at the same time. Because we only have interaction measurements for the Facebook dataset, we limit ourselves to that dataset. The result is shown in Fig. 9. While the performance of the similarity-biased random walk produces *almost* same results as the simple random-walk, the interaction-biased walk affects the number of the accepted nodes. Furthermore, the lazy random walk captures the behavior of model when deviated from the simple random-walk. As shown for this dataset, the interaction model behavior is characterized by the behavior of the lazy random walk for two given parameters ($\alpha = 0.48$ and $\alpha = 0.64$) suggesting that the interaction model can be further modeled as a lazy random walk where the problem is to find the proper parameters to match its behavior. Note that the value of α works for this dataset in particular. However, other datasets may be characterized by other values. We also find that the number of escaping tails per node is also decreased using our design, as shown in Fig. 10. In this last experiment, we compute the average escaping tails per 100 honest node samples, and by running the experiment 5 times, independently

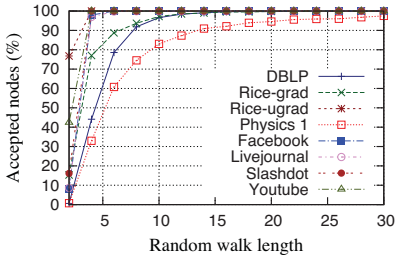


Fig. 8. Accepted honest nodes in SybilLimit versus random walk length – with simple random walk. Different graphs have different quality of the algorithmic property though being with same size.

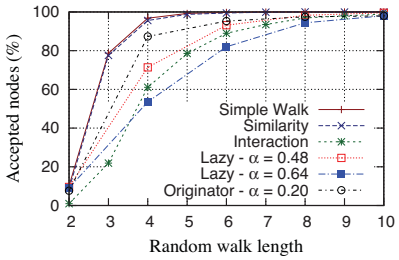
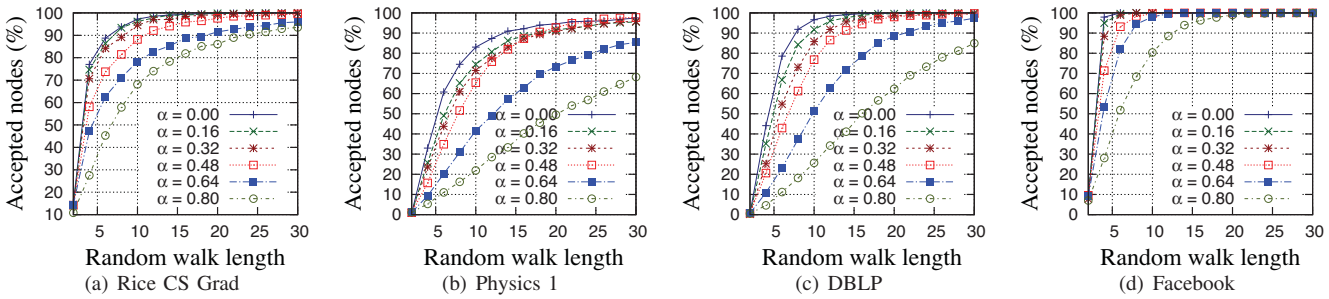


Fig. 9. Accepted honest nodes in SybilLimit versus random walk length, when using the different designs to model the of trust in the social graph. The social graph of Facebook in Table II.

with a the given attackers edges for which nodes are selected uniformly at random from the honest region. In the experiment of Fig. 10, and for the interaction model, we assume that the attacker may infiltrate the social graph but cannot produce meaningful interactions, and thus the number of escaping tails to the attacker is always zero. It would be interesting in the future to generalize this model to an attacker with limited budget of interactions, and see how this changes the number of escaping tails with varying budgets. Finally to understand the impact of the different random walks on the accepted Sybil nodes per attack edge, we experiment with the same dataset (Facebook) and for varying g. The results are shown in Figure 11. Similar to above, our designs outperform the uniform design (more experiments are in [31]).

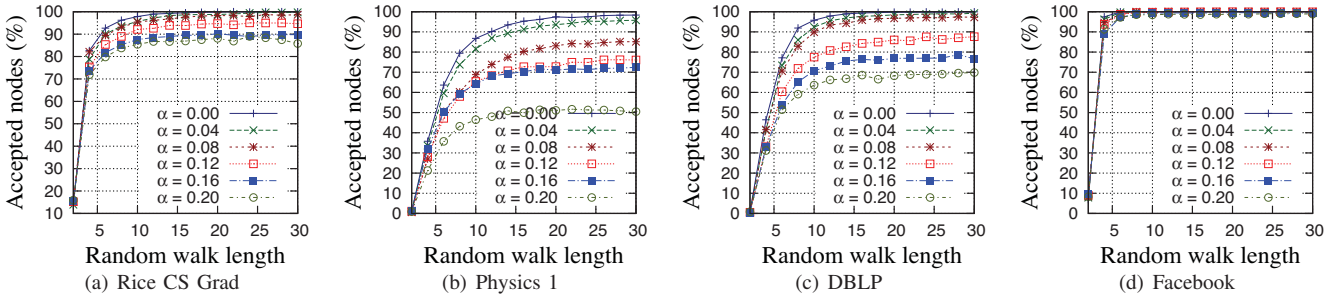
VI. Implications of Findings

To sum up, we find in this study that one can control the behavior of the social network-based Sybil defenses by incorporating parameters for trust. For this purpose, we introduced



(a) Rice CS Grad (b) Physics 1 (c) DBLP (d) Facebook

Fig. 6. The performance of SybilLimit measured for accepted honest nodes when using different lengths of lazy random walk for different social graphs.



(a) Rice CS Grad (b) Physics 1 (c) DBLP (d) Facebook

Fig. 7. The performance of SybilLimit depends on the underlying social graph, where different graphs require different walk lengths to ensure the same number of accepted nodes. The originator-biased random walk can further influence the number of nodes accepted in each graph.

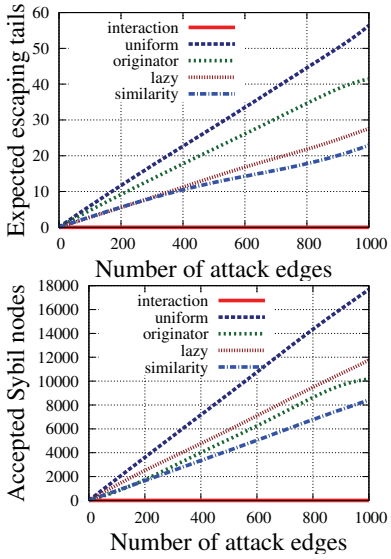


Fig. 10. Expected escaping walks per node (among 100 nodes, $r = 850$) in Facebook dataset (in Table II) where $w = 6$ and $\alpha = 0.4$ for both of the originator and lazy random walks.

Fig. 11. Accepted Sybil nodes over tainted tails when varying $g$in Facebook dataset (in Table II) where $w = 6$ and $\alpha = 0.4$ for both of the originator and lazy random walks.

and experimented the behavior of four designs. In graphs that are empirically-proven to be fast mixing and well-performing for the utility of the Sybil defense – though having poor value of trust – we have shown that one can select the necessary parameters to account for trust and make the performance of the defense on that graph equivalent to stronger and richer version of the same graph – e.g., the case of the interaction-based model versus the mere connections on the Facebook dataset. With these designs being intuitive in characterizing trust, the results being in agreement one another, and with this paper being the first of its own type in this direction, we believe that this study is a first step in the direction of bringing well-received theoretical results into practice. The implications of our findings can be summarized as follows.

First, the mixing time and utility of the Sybil defense depend on the underlying graph. Through measurements, we supported our hypothesis that the quality of the social graph depends on the characteristic of the social links between the nodes. On one hand, social links that are easier to make result in well-enmeshed graphs but are bad in principle for the Sybil defense since they already tolerate bad edges. However, these are shown to provide good honest nodes acceptance rate even with shorter random walks. On the other hand, social links that are harder to make result in graphs with more community structure, which are bad for the detection (as shown in [11]) and require longer walks to operate for the honest nodes.

Second, it is now possible for the Sybil defense operator, when given multiple options of social graphs, to further derive the utility of the Sybil defense using several criteria. Our study empowers the operators by an additional dimension that influences the behavior of the Sybil defense: trust.

Third, our findings answer a recently called for question in [11] of studying the behavior of Sbyil defenses when operated on the interaction-based model rather than the mere social connections, which are sometimes less meaningful. In short, our study shows that the interaction model can influence the behavior of the Sybil defense, by requiring longer random walk for the defense to work for honest nodes. However, this finding also suggests that a more community-structure is in the interaction model than in the mere social graph. This implies that, while the original social graph does not possess clear community structure, the use of the interaction model would add sensitivity for the detection part of the defense and result in weaker detection. However, the underlying graphs in both cases are different and the interpretation of the results should also consider the trust value in the interaction model, which is a better fit to the trust required in the Sybil defense.

Finally, online social graphs are known to possess weaker value of trust [29]. However, their potential for being used for Sybil defenses is very high since alternatives are limited,

too expensive, and may not fit into the Sybil defense settings. For example, co-authorship social graphs which are known for their trust value may not necessarily include most users of a particular online system that tries to deploy the Sybil defense. On the other hand, given the popularity of online social networks, Sybil defenses may benefit from them, across systems and networks. To this end, the main finding of the paper is to open the door wide open for investigating trust, its modeling, and quantification for these systems.

VII. Conclusion and Future Work

In conclusion, we propose several designs to capture the trust value of social graphs in social networks used for Sybil defenses. Our designs filter weak trust links and successfully bound the mixing time which controls the number of accepted nodes using the Sybil defenses to account for variable trust. Our designs provide defense designers with parameters to model trust and evaluate Sybil defenses based on the "real value" of social networks.

Several directions are worth investigation in the near future. First, we would like to investigate generalized node-wise parameterized designs that consider different parameters for different users, or categories of them. Second, we would like to theoretically formulate the behavior of the different designs considering other features of the underlying graph, e.g., its eigenvalues, mixing time, etc. Finally, we would like to investigate the applicability of these designs in other contexts where the trust of social networks is used. For more on open problems and further work, please see [31].

Acknowledgement: we would like to thank the anonymous reviewers of INFOCOM'11, Aaram Yun, and Eugene Y. Vasserman for their feedback that improved the presentation of this work, Max Schuchard and Huy Tran for helping with the implementations, and Alan Mislove, Ben Y. Zhao, Christo Wilson, and Bimal Viswanath for providing the social graphs. This research was supported by the NSF under grant no. CNS-0917154 and a research grant from Korea Advanced Institute of Science and Technology (KAIST).

References

[1] J. Douceur, "The sybil attack," in *IPTPS*, 2002, pp. 251–260.

[2] B. Levine, C. Shields, and N. Margolin, "A survey of solutions to the sybil attack," University of Massachusetts Amherst, Tech. Rep., 2006.

[3] M. Castro, P. Druschel, A. J. Ganesh, A. I. T. Rowstron, and D. S. Wallach, "Secure routing for structured peer-to-peer overlay networks," in *OSDI*, 2002.

[4] A. Adya, W. J. Bolosky, M. Castro, G. Cermak, R. Chaiken, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. Wattenhofer, "Farsite: Federated, available, and reliable storage for an incompletely trusted environment," in *OSDI*, 2002.

[5] J. Ledlie and M. I. Seltzer, "Distributed, secure load balancing with skew, heterogeneity and churn," in *INFOCOM*, 2005, pp. 1419–1430.

[6] F. Lesueur, L. Mé, and V. V. T. Tong, "An efficient distributed pki for structured p2p networks," in *Proceeding of P2P*. IEEE, 2009, pp. 1–10.

[7] N. Borisov, "Computational puzzles as sybil defenses," in *Peer-to-Peer Computing*, 2006, pp. 171–176.

[8] H. Yu, M. Kaminsky, P. B. Gibbons, and A. D. Flaxman, "Sybilguard: defending against sybil attacks via social networks," *IEEE/ACM Trans. Netw.*, vol. 16, no. 3, pp. 576–589, 2008.

[9] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "Sybillimit: A near-optimal social network defense against sybil attacks," in *IEEE Symposium on Security and Privacy*, 2008, pp. 3–17.

[10] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman, "Sybilguard: defending against sybil attacks via social networks," in *SIGCOMM*, 2006, pp. 267–278.

[11] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove, "An analysis of social network-based sybil defenses," in *SIGCOMM*, 2010.

[12] G. Danezis and P. Mittal, "Sybilinfer: Detecting sybil nodes using social networks," in *NDSS*, 2009.

[13] C. Lesniewski-Lass and M. F. Kaashoek, "Whānau: A sybil-proof distributed hash table," in *USENIX NSDI*, 2010, pp. 3–17.

[14] A. Mohaisen, A. Yun, and Y. Kim, "Measuring the mixing time of social graphs," in *IMC*. ACM, 2010, pp. 383–389.

[15] N. Tran, B. Min, J. Li, and L. Subramanian, "Sybil-resilient online content voting," in *NSDI*. USENIX Association, 2009, pp. 15–28.

[16] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *MobiHoc '07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM, 2007, pp. 32–40.

[17] S. Marti, P. Ganesan, and H. Garcia-Molina, "Dht routing using social links," in *IPTPS*, 2004, pp. 100–111.

[18] S. Pai, T. Roosta, S. B. Wicker, and S. Sastry, "Using social network theory towards development of wireless ad hoc network trust," in *AINA Workshops*, 2007, pp. 443–450.

[19] R. Landa, D. Griffin, R. Clegg, E. Mykoniati, and M. Rio, "A sybilproof indirect reciprocity mechanism for peer-to-peer networks," in *Proceedings of IEEE Infocom*, vol. 9, 2009.

[20] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Social Networks*, August 2009.

[21] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in *EuroSys*. ACM, 2009, pp. 205–218.

[22] J. Leskovec, D. P. Huttenlocher, and J. M. Kleinberg, "Predicting positive and negative links in online social networks," in *WWW*, 2010, pp. 641–650.

[23] D. J. Crandall, D. Cosley, D. P. Huttenlocher, J. M. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *KDD*, 2008, pp. 160–168.

[24] E. Le Merrer and G. Trédan, "Centralities: capturing the fuzzy notion of importance in social graphs," in *SNS*. ACM, 2009, pp. 33–38.

[25] D. Liben-Nowell and J. M. Kleinberg, "The link prediction problem for social networks," in *CIKM*. ACM, 2003, pp. 556–559.

[26] A. Mislove, M. Marcon, P. K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Internet Measurement Comference*, 2007, pp. 29–42.

[27] Y.-Y. Ahn, S. Han, H. Kwak, S. B. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *WWW*, 2007, pp. 835–844.

[28] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Trans. Web*, vol. 3, no. 4, pp. 1–33, 2009.

[29] C. Dwyer, S. Hiltz, and K. Passerini, "Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace," in *AMCIS*, 2007.

[30] S. Nagaraja, "Anonymity in the wild: Mixes on unstructured networks," in *Privacy Enhancing Technologies*, 2007, pp. 254–271.

[31] A. Mohaisen, N. Hopper, and Y. Kim, "Incorporating trust into social network-based sybil defenses," UMN, Tech. Rep., 2010.

[32] A. Sinclair, "Improved bounds for mixing rates of marcov chains and multicommodity flow," *Comb., Probability & Computing*, vol. 1, pp. 351–370, 1992.

[33] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *KDD*. ACM, 2005, pp. 177–187.

[34] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *CoRR*, vol. abs/0810.1355, 2008.

[35] M. Richardson, R. Agrawal, and P. Domingos, "Trust management for the semantic web," in *ISWC*, 2003, pp. 351–368.

[36] M. Ley, "The DBLP computer science bibliography: Evolution, research issues, perspectives," in *String Processing and Information Retrieval*. Springer, 2009, pp. 481–486.

[37] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in Online Social Networks," in *WSDM*, New York, NY, February 2010.