Designs to Account for Trust in Social Network-based Sybil Defenses

Abedelaziz Mohaisen
University of Minnesota
Minneapolis, MN 55455, USA
mohaisen@cs.umn.edu

Nicholas Hopper
University of Minnesota
Minneapolis, MN 55455, USA
hopper@cs.umn.edu

Yongdae Kim
University of Minnesota
Minneapolis, MN 55455, USA
kyd@cs.umn.edu

ABSTRACT

Social network-based Sybil defenses exploit the trust exhibited in social graphs to detect Sybil nodes that disrupt an algorithmic property (i.e., the fast mixing) in these graphs. The performance of these defenses depends on the quality of the algorithmic property and assuming a strong trust model in the underlying graph. While it is natural to think of trust value associated with the social graphs, Sybil defenses have used the social graphs without this consideration. In this paper we study paramagnetic designs to tune the performance of Sybil defenses by accounting for trust in social graphs and modeling the trust as modified random walks. Our designs are motivated by the observed relationship between the algorithmic property required for the defenses to perform well and a hypothesized trust value in the underlying graphs.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems – *Distributed Applications*; C.2.0 [**Computer-Communication Networks**]: General – *Security and Protection*

General Terms

Security, Design, Algorithms, Experimentation

Keywords

Sybil Attack, Social Networks, Trust

1. INTRODUCTION

There has been a great interest in the research community for the potential of defending against Sybil attacks using social networks [8]. In these defenses, 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 to non-Sybil). The sociological assumption is a constraint on the trust in the underlying social graph: the graph needs to have strong trust as evidenced, for example, by face to face interaction demonstrating social nodes knowledge of each other [10, 11]. While the first assumption has been questioned recently in [8], where it is shown that even the honest subgraph may have some cuts that disrupt the algorithmic property on which Sybil defenses are based, the trust, though being a crucial requirement for these designs, was not considered carefully. Even worse, these defense [10, 11, 2, 4] — when verified against real-world networks — have considered samples of online social graphs, which are known to possess weaker value of trust.

Recently, we have shown that the mixing time, a concrete measure of the algorithmic property required in social networks used for building Sybil defenses, is greater than anticipated and used in literature [5]. We also relaxed the assumption by showing that a faster mixing graph not necessary for these designs to work [5]. Most importantly, we have shown "variable" mixing times even for the same sized social graphs implying that they, even algorithmically, cannot be taken equally for these designs. Also, the variable mixing time turned out not to be 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). To this end, we study designs to incorporate information on social graphs to reflect their trust value.

2. PRELIMINARIES

Network model: the social network is viewed as an undirected and unweighted graph $G = (V, E)$ where $|V| = n$, $V = \{v_1, \ldots, v_n\}$, $|E| = m$, $e_{ij} = (v_i \rightarrow v_j) \in E$ 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 also refer to $\mathbf{P} = [p_{ij}]^{n \times n}$ as the *transition matrix* where $p_{ij} = \frac{1}{\deg(v_i)}$ if $e_{ij} \in E$ and 0 otherwise, where $\deg(v_i)$ is the degree of v_i which is $\deg(v_i) = \sum_{k=1}^{n} \mathbf{A}_{ik}$. The set of neighbors of v_i is $N(v_i)$ and $|N(v_i)| = \deg(v_i)$.

Simple random walks: walking randomly on G is captured by a Markov Chain (MC), so a simple random walk of length w is a sequence of vertices in G beginning from v_i and ending at v_t using the transition matrix $\mathbf{P}$. The MC is said to be *ergodic* if it is irreducible and aperiodic, meaning that the MC 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 invariant to the transition matrix $\mathbf{P}$ (i.e., $\pi\mathbf{P} = \pi$). For this simple walk $\pi = [\pi_i]$ where $\pi_i = \frac{\deg v_i}{2m}$ [5]. The mixing time of the MC parameterized by a variation distance parameter ϵ is $T(\epsilon) = \max_i \min\{t : |\pi - \pi^{(i)}\mathbf{P}^t|_1 < \epsilon\}$, where $\pi^{(i)}$ is the initial distribution at vertex v_i, $\mathbf{P}^t$ is the transition matrix after t steps, and $|\cdot|_1$ is the total variation distance. The graph is fast mixing if $T(\epsilon) = \text{poly}(\log n, \log \frac{1}{\epsilon})$. [10, 11] strengthen the definition by requiring $\epsilon = \Theta(\frac{1}{n})$ and $T(\epsilon) = O(\log n)$. We recently found that $\epsilon \approx 1/4$ is sufficient, and smaller ϵ is unneces-

CCS'10, October 4–8, 2010, Chicago, Illinois, USA.
ACM 978-1-4503-0244-9/10/10.

sary, for the designs [10, 11, 4] to operate for 99% admission rate.

Measuring the mixing time [5]: while the mixing time of G can be brute-force computed according to its definition, Sinclair's result [6] can be used for bounding it. 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 $\mathbf{P}$. 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 μ as $\mu = \max(|\lambda_2|, |\lambda_{n-1}|)$. Then, the mixing time $T(\epsilon)$ is bounded by $\frac{\mu}{2(1-\mu)} \log(\frac{1}{2\epsilon}) \leq T(\epsilon) \leq \frac{\log(n)+\log(\frac{1}{\epsilon})}{1-\mu}$.

Social network based sybil defenses: Sybil defenses based on social networks exploit the trust exhibited in the social graphs. There has been a constant effort in this direction as reported in SybilGuard [11], SybilLimit [10], SybilInfer [2], SumUp [7], as well as applications to DHT in Whānau [4]. In principle, the quality of these defenses depends on the quality of the algorithmic property of the underlying graph. For a nice exposition some of these designs, see the recent work of Viswanath et al. in [8].

3. DESIGNS TO ACCOUNT FOR TRUST

Most defenses in literature uses the uniform random walk in section 2. In this section, we introduce several designs of modified random walks that consider a "trust" parameter between nodes that influences the random walk. In all of the proposed modified random walks, the purpose is to assign "trust-driven" weights and thus deviating from uniform. We do this by either capturing the random walk in the originator or current node, as the case of originator biased random walk and lazy random walk respectively, or by biasing the random walk probability at each node, as the case of interaction and similarity-based weights assignment over edges, or a combination of them. The intuition behind the different assignment mechanisms are similar in essence but motivated by different observations. For the lazy and originator-biased random walk the main intuition is that nodes tend to trust "their own selves" and other nodes within their community (up to some distance) more than others. On the other hand, interaction and similarity-based 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 walks over the graph they are applied on. We omit the details for lack of space.

Lazy random walks: lazy random walks accommodate for the trust exhibited in the social graph by assuming the parameter α used for characterizing this trust level. With the lazy random walk, each node along the path decides to capture the walk with probability α or to follow the simple random walk with $1-\alpha$ at each time step. The transition matrix is then defined as $\mathbf{P'} = \alpha\mathbf{I} + (1-\alpha)\mathbf{P}$. The stationary distribution of this walk is same like the simple walk in section 2. In particular, since $\mathbf{P'} = \alpha\mathbf{I} + (1-\alpha)\mathbf{P}$, by multiplying both sides by π, we get $\pi\mathbf{P'} = \pi(\alpha\mathbf{I} + (1-\alpha)\mathbf{P}) = \alpha\pi\mathbf{I} + (1-\alpha)\pi\mathbf{P} = \alpha\pi + \pi - \alpha\pi = \pi$.

Originator-biased random walks: The originator-biased random walk considers the bias introduced by the random walk initiator not to be fooled by Sybil nodes in a social graph that lacks quality of trust. 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 as $p_{ij} = \frac{1-\alpha}{\deg(v_i)}$ if $v_j \in N(v_i)$, $p_{ij} = \alpha$ if $v_j = v_r$, or 0 otherwise. For the α and $\mathbf{A}_r$ with all-zero but the r^{th} row, which is 1's, $\mathbf{P'}$ for the random walk originated from v_r is given as $\mathbf{P'} = \alpha\mathbf{A}_r + (1-\alpha)\mathbf{P}$. Since the "stationary distribution" is not unique among all initial distributions, it's called the "bounding distribution" and given for v_r as $\pi = [\pi_i]^{1\times n}$ where $\pi_i = (1-\alpha)\frac{\deg(v_i)}{2m}$ if $v_i \in V \setminus \{v_r\}$ and $\pi_i = \alpha + \frac{\deg(v_i)}{2m}$ if $v_i = v_r$. It's then easy to show that the bounding distribution is a valid probability distribution since $\alpha + \frac{\deg(v_r)}{2m} + \sum_{v_i \in V/\{v_r\}}(1-\alpha)\frac{\deg(v_i)}{2m} = 1$.

Similarity-biased random walks: The similarity between nodes in social networks is used for measuring the *strength of social ties and predicting future interactions* [1]. For two nodes v_i and v_j with sets of neighbors $N(v_i)$ and $N(v_j)$, respectively, the similarity is defined as the set of nodes common to both of v_i and v_j normalized by all their neighbors and expressed as: $S(v_i, v_j) = \frac{N(v_i) \cap N(v_j)}{N(v_i) \cup N(v_j)}$. For $\mathbf{v}_i, \mathbf{v}_j \in \mathbf{A}$ corresponding to the adjacency entries of v_i and v_j, the cosine similarity measure is used to capture the similarity definition above, 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 definition by adding 1 each time to the denominator to account for the edge between the nodes. Also, we compute the similarity for adjacent nodes only by computing $\mathbf{S}$, the similarity matrix, as $\mathbf{S} = [s_{ij}]$ where $s_{ij} = S(v_i, v_j)$ if $v_j \in N(v_i)$ and 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 $\pi_i = (\sum_{z=1}^{n} s_{iz})(\sum_{j=1}^{n}\sum_{k=1}^{n} s_{jk})^{-1}$.

Interaction-biased random walks: recently, the interaction between nodes has been observed as one measure for determining the strength of social links between social actors especially in online social networks [9]. In its simple form, the interaction model captures activities between the different nodes in the graph (e.g., the posting between different users in the Facebook) and assigns weights, which translates into high trust value, to nodes that have higher interaction and lower weights to nodes with less interaction. Let $\mathbf{B}$ be the raw interaction measurements and $\mathbf{D}$ be a diagonal matrix with diagonal elements being the row norm of $\mathbf{B}$, the transition matrix $\mathbf{P}$ of the random walk based on interaction is computed as $\mathbf{P} = \mathbf{D}^{-1}\mathbf{B}$ and, similar to the similarity-biased walks, the stationary distribution derived from the interaction matrix is $\pi = [\pi_i]^{1\times n}$ where $\pi_i = (\sum_{i=1}^{n} b_{iz})(\sum_{j=1}^{n}\sum_{k=1}^{n} b_{jk})^{-1}$.

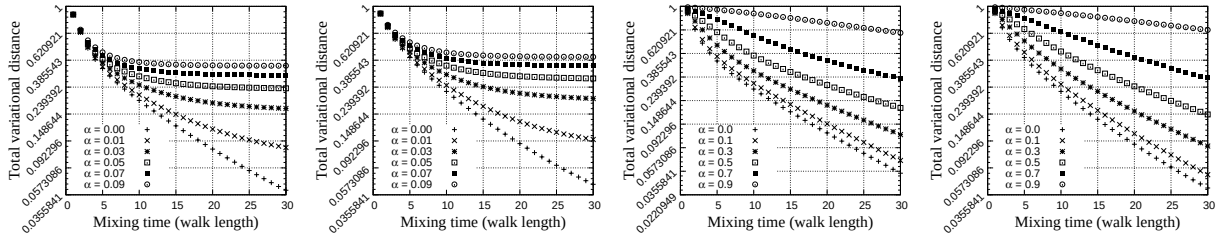
Mixed random walks: it is intuitive and natural to consider a hybrid design that constitutes more than one of the aforementioned random walks. In particular, the interaction and similarity-based models "rank" different nodes differently and "locally" assign weights to them. Though this limits the mixing 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 for the lazy and originator-biased models. It's hence technically promising and intuitively sound to consider combinations of these designs.

Table 1: Social graphs and their properties

Dataset	n/average degree/μ	Dataset	n/average degree/μ
Physics 1	4.2K /3.23/ 0.998133	Youtube	1.1M /2.63 / 0.997972
Wiki-vote	7.1K /14.256/0.899418	Livejournal B	1M / 27.56/ 0.999695
Slashdot 2	77.4K/7.06/0.987531	Livejournal A	1M /26.15 / 0.999387
Slashdot 1	82.2K/7.09/ 0.987531	Facebook B	1M /15.81 / 0.992020
Facebook	63.4K/12.87/ 0.998133	Facebook A	1M /20.35/ 0.982477
Physics 2	11.2K /10.50/0.998221	DBLP	615K /1.88/ 0.997494
Physics 3	8.6K/2.87/0.996879	Enron	33.7K /5.37/ 0.996473

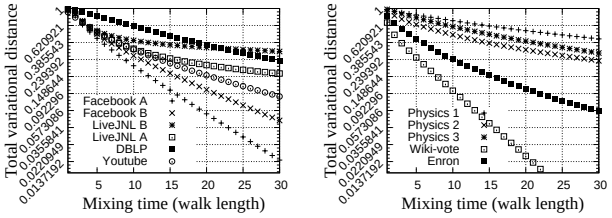
4. RESULTS AND BRIEF DISCUSSION

We first measure the mixing time of the social graphs used in this study — in Table 1 — using the definition of the mixing time



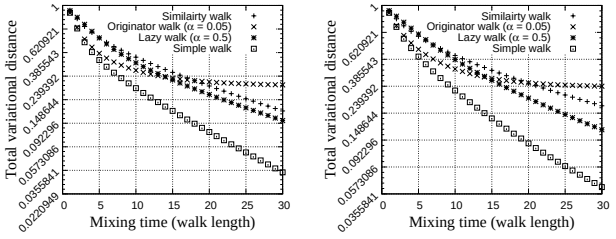
(a) Facebook A (b) Livejournal (c) Facebook A (d) Livejournal A

Figure 3: Preliminary measurements of modified random walks' impact on the mixing time — (a) and (b) are for originator-biased while (c) and (d) are for lazy random walks.



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

Figure 1: The average mixing time of a sample of 1000 initial distributions in each graph in Table 1 using the sampling method for computing the mixing time by its definition over P.



(a) Facebook A (b) Livejournal

Figure 2: The mixing time of the different graphs when using simple vs. lazy, originator, and similarity-biased walks.

in section 2, highlighting the variability of the algorithmic property and relating that to graph nature (see datasets below for details). We follow this by examining the impact of adapting the different trust characterization methods on the mixing time. In all measurements we examine the mixing time and quantifying the impact of degraded mixing time on the actual performance of each defense become a secondary issue. We leave this part to the complete work.

Social graphs — datasets: the datasets used in our experimentation are in Table 1. These datasets are carefully selected so to feature (hypothetically) different models of knowledge of the social actors among each other in the social graph. These graphs are categorized into: (1) Social graphs of networks that exhibit concrete knowledge between social actors and are good for the trust assumptions of the Sybil defenses — e.g., co-authorship datasets, such as physics co-authorships and DBLP which are shown to be slower mixing (see Figure 1). (2) Graphs of networks that may not require face-to-face knowledge but require the effort of interaction. — e.g., Youtube, Livejournal, and Enron, which are shown for slow mixing. (3) Datasets that may not require prior knowledge between the social actors and are known for exhibiting less strict social model such as those of the online social networks (e.g., Facebook).

Implication of the pragmatic designs on the mixing time: we implement three of the proposed designs: lazy, originator, and similarity biased random walks and examine their impact on the mixing time of social graphs in Table 1. For feasibility reasons, we sample only 10K nodes, using the breadth-first search algorithm, from each graph larger than 10K. The results are shown in Figure 2 and Figure 3. 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 instance, as shown in Figure 2(a), $\epsilon \approx 1/4$ is realizable at $w = 6$ (for 99% admission of non-Sybil nodes) with the simple random walk, $w = 17$ for both lazy and originator-biased random walk. However, this is realized with $\alpha = 0.5$ in the lazy against $\alpha = 0.05$ in the originator-biased walk.

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 parameters to model trust and pragmatically evaluate Sybil defenses based on the "real value" of social networks.

Acknowledgement: This research was supported by the NSF under grant no. CNS-0917154 and a grant from Korea Advanced Institute of Science and Technology (KAIST). We thank Alan Mislove and Ben Y. Zhao for providing the data used in this study.

5. REFERENCES

[1] 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 Y. Li, B. Liu, and S. Sarawagi, editors, *KDD*, pages 160–168. ACM, 2008.

[2] G. Danezis and P. Mittal. Sybilinfer: Detecting sybil nodes using social networks. In *NDSS*. The Internet Society, 2009.

[3] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney. Statistical properties of community structure in large social and information networks. In J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, editors, *WWW*, pages 695–704. ACM, 2008.

[4] C. Lesniewski-Lass and M. F. Kaashoek. Whānau: A sybil-proof distributed hash table. In *7th USENIX Symposium on Network Design and Implementation*, pages 3–17, 2010.

[5] A. Mohaisen, A. Yun, and Y. Kim. Measuring the mixing time of social graphs. In *ACM SIGCOMM Conference on Internet Measurements*. ACM, 2010.

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

[7] N. Tran, B. Min, J. Li, and L. Subramanian. Sybil-resilient online content voting. In *NSDI'09: Proceedings of the 6th USENIX symposium on Networked systems design and implementation*, pages 15–28, Berkeley, CA, USA, 2009. USENIX Association.

[8] B. Viswanath, A. Post, K. P. Gummadi, and A. Mislove. An analysis of social network-based sybil defenses. In *SIGCOMM (to appear)*, pages 00–00, 2010.

[9] C. Wilson, B. Boe, A. Sala, K. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In W. Schröder-Preikschat, J. Wilkes, and R. Isaacs, editors, *EuroSys*, pages 205–218. ACM, 2009.

[10] 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*, pages 3–17, 2008.

[11] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. SybilGuard: defending against sybil attacks via social networks. In *SIGCOMM*, pages 267–278, 2006.