GANG: Detecting Fraudulent Users in Online Social Networks via Guilt-by-Association on Directed Graphs

Binghui Wang, Zhonghao Liao

December 13, 2017

1 Abstract

Detecting fraudulent users in online social networks is a fundamental and urgent research problem as adversaries can use them to perform various malicious activities. Global social structure based methods, which are known as guilt-by-association, have been shown to be promising at detecting fraudulent users. However, existing guilt-by-association methods either assume symmetric (i.e., undirected) social links, which oversimplifies the asymmetric (i.e., directed) social structure of real-world online social networks, or only leverage labeled fraudulent users or labeled normal users (but not both) in the training dataset, which limits detection accuracies.

In this project, we propose GANG, a guilt-by-association method on directed graphs, to detect fraudulent users in OSNs. GANG is based on a novel pairwise Markov Random Field that we design to capture the unique characteristics of the fraudulent-user-detection problem in directed OSNs. In the basic version of GANG, given a training dataset, we leverage Loopy Belief Propagation (LBP) to estimate the posterior probability distribution for each user and uses it to predict a users label. However, the basic version is not scalable enough and not guaranteed to converge because it relies on LBP. Therefore, we further optimize GANG and our optimized version can be represented as a concise matrix form, with which we are able to derive conditions for convergence. We compare GANG with various existing guilt-by-association methods on a largescale Twitter dataset and a large-scale Sina Weibo dataset with labeled fraudulent and normal users. Our results demonstrate that GANG substantially outperforms existing methods, and that the optimized version of GANG is significantly more efficient than the basic version. Our code is publicly available on the website "https://github.com/STAT430".

2 Background

Online social networks (OSNs) have become indispensable platforms for interacting with people, processing information, and diffusing social influence. However, a large number of users on OSNs are fraudulent, e.g., spammers, fake users, and compromised normal users. For instance, it was reported that 10% of Twitter users were fake. Adversaries use these fraudulent users to perform various malicious activities such as disrupting democratic election and influencing financial market via spreading rumors, distributing malware, as well as harvesting private user data. Therefore, detecting fraudulent users is an urgent research problem.

Indeed, this research problem has attracted increasing attention from multiple communities including data mining, cybersecurity, and networking. Depending on the used information sources, we classify existing approaches into two categories, global structure based methods and local feature based methods. Global structure based methods leverage the global structure of a social graph and are (or normal) if it is linked with other fraudulent (or normal) users. In order to stress their application to detecting fraudulent users, we also call these methods guilt-by-association. Existing guilt-by-association methods either assume symmetric (i.e., undirected) social links which oversimplifies the asymmetric (i.e., directed) social graph structure in real-world OSNs, or leverage only labeled fraudulent users or normal users (but not both) in the training dataset which limits their detection accuracies. Local feature based methods leverage a users local subgraph structure (e.g., ego-network), side information (e.g., IP address, behaviors, and content), and possibly combine them with features from the global social structure. A key limitation of these methods is that they are not adversarially robust, i.e., fraudulent users can evade detection via modifying their side information to mimic normal users and colluding to manipulate their local subgraph structures as desired. Indeed, in our experiments, we observe such fraudulent users on Sina Weibo, one of the largest OSNs in China (See Figure 5).

3 Our Contribution: GANG

In this work, we propose GANG, a guilt-byassociation method on directed graphs, to detect fraudulent users in OSNs. In GANG, we associate a binary random variable with each user to model its label, and then we design a novel pairwise Markov Random Field (pMRF) to model the joint probability distribution of all these random variables based on the directed social graph. Our pMRF incorporates unique characteristics of the fraudulent-user-detection problem. Specifically, we call an edge (u;v) unidirectional if the edge (v;u) in the reverse direction does not exist, otherwise we call the edge bidirectional. If two users are linked by bidirectional edges and have the same label, then our pMRF produces a larger joint probability. However, suppose u and v are linked by a unidirectional edge (u;v), e.g., on Twitter, this means that u follows v, but v does not follow back to u. If u is fraudulent or v is normal, then whether the unidirectional edge (u;v) exists or not does not influence the joint probability under our pMRF, otherwise the edge (u;v) makes the joint probability larger. This is because a fraudulent user can follow arbitrary users without being followed back, while a normal user can be followed by arbitrary users without following them back.

In the basic version of GANG, given a training dataset, we use Loopy Belief Propagation (LBP) to estimate the posterior probability distribution for each binary random variable and use it to predict label of the corresponding user. However, the basic version has two shortcomings: 1) it is not scalable enough because LBP needs to maintain messages on each edge, and 2) it is not guaranteed to converge because LBP might oscillate on loopy graphs. Therefore, we further optimize GANG to address these shortcomings. Our optimizations include eliminating message maintenance and approximating GANG by a concise matrix form. We also derive the conditions for our optimized GANG to converge. We evaluate GANG and compare it with various existing guilt-by-association methods using a large-scale Twitter dataset (42M users and 1.5B directed edges) and a large-scale Sina Weibo dataset (3.5M users and 653M directed edges). Both datasets have labeled fraudulent and normal nodes. Our results demonstrate that GANG substantially outperforms existing guilt-by-association methods. Via a case study on Sina Weibo, we found that GANG can detect a large amount of fraudulent users that evaded Sina Weibos detector. Moreover, we demonstrate that the optimized version of GANG is significantly more efficient than its basic version.

In summary, our key contributions are as follows:

- We propose GANG to detect fraudulent users in OSNs via guilt-by-association on directed graphs. GANG leverages a novel pMRF that captures the unique characteristics of the fraudulent-user-detection problem.
- We optimize GANG to make it scalable and convergent.
- We evaluate GANG and various existing guilt-byassociation methods using a large-scale Twitter dataset and a large-scale Sina Weibo dataset with labeled fraudulent and normal users. Our results demonstrate that GANG significantly outperforms existing methods, and that the optimized GANG is significantly more efficient than its basic version.

4 Experimental Results

4.1 Experimental Setup

Dataset description: We compare GANG with existing methods on two large-scale OSN datasets with labeled fraudulent and normal nodes.

First, we obtained a Twitter follower-followee graph with 41,652,230 users and 1,468,364,884 edges from Kwak et al.. In this graph, a directed edge (u, v) means that u follows v. We obtained ground truth labels for each node from Wang et al.. Specifically, 205,355 users were suspended by Twitter and we treated them as fraudulent users; 36,156,909 users were active and we treated them as normal users; and the remaining 5,289,966 users were deleted. As deleted users could be deleted by Twitter or by users themselves, we could not distinguish the two cases without accessing to Twitter's internal data. Thus, we treat them as unlabeled users. We sample 500,000 labeled users uniformly at random as a training set and treat the remaining labeled users as the testing set.

Second, we obtained a Sina Weibo dataset with 3,538,487 users and 652,889,971 directed edges. Like Twitter, a directed edge (u, v) means that u follows v. Fu et al. also manually labeled 2000 users sampled uniformly at random. Among them, 482 were fraudulent users, 1,498 were normal users, and 20 were unknown users (we do not consider these users in our experiments). We split the fraudulent and normal users into two halves; one is treated as the training set and the other is treated as the testing set. Table 1 shows some statistics of our datasets.

Compared methods: We compare GANG with both undirected and directed graph based methods. By default, we will use the optimized version of GANG.

Table 1: Dataset statistics.

Dataset	Twitter	Sina Weibo
#Nodes	41,652,230	3,538,487
#Edges	1,468,364,884	652,889,971
Ave. degree	71	369

1) Using undirected graphs. We consider the following undirected graph based methods: the well known graph-based semi-supervised learning method (SSL), SybilRank, SybilBelief, and SybilSCAR. SSL and SybilRank are based on random walks and SybilBelief is based on pMRF. SybilSCAR unifies random walk based methods and pMRF based methods as a local rule based framework. Moreover, under the framework, SybilSCAR designs a new local rule which outperforms existing random walk and pMRF based local rules. SSL, SybilBelief, and SybilSCAR can leverage both fraudulent users and normal users in the training dataset, while SybilRank is only able to leverage labeled normal users. These methods transform a directed graph into an undirected one via keeping an edge between two nodes if they are connected via bidirectional edges. This is more robust than keeping both bidirectional and unidirectional edges because fraudulent nodes can create arbitrary number of unidirectional edges with normal nodes, making them well embedded among normal nodes. Since these methods require connected graphs, we evaluate them on the largest connected component in the transformed undirected graph.

2) Using directed graphs. We consider the following directed graph based methods: TrustRank, DistrustRank, CIA, and CatchSync. TrustRank, DistrustRank, and CIA are based on random walks, while CatchSync leverages HITS. TrustRank and DistrustRank were originally designed to detect fraudulent webpages based on hyperlinks, but they can be applied to detect fraudulent users in OSNs. TrustRank leverages only labeled normal nodes in the training dataset; DistrustRank and CIA are essentially the same, and they only leverage labeled fraudulent nodes; and CatchSync does not leverage the training dataset.

Table 2: AUCs of compared methods.

Methods		Twitter	Sina Weibo
Using undirected graphs	SSL	0.55	0.68
	SybilRank	0.57	0.61
	SybilBelief	0.61	0.65
	SybilSCAR	0.64	0.68
Using directed graphs	TrustRank	0.60	0.66
	DistrustRank	0.63	0.64
	CIA	0.63	0.64
	CatchSync	0.68	0.51
	GANG	**0.72**	**0.80**

4.2 Ranking Results

Each compared method essentially computes a score for each node. We rank the nodes in the testing dataset using the scores such that fraudulent nodes are supposed to rank higher than normal nodes.

Overall ranking performance: We first use AUC to measure the overall ranking performance of the compared methods. In our problem, AUC can be interpreted as the probability that a randomly sampled fraudulent node is ranked higher than a randomly sampled normal node in the testing dataset. The higher AUC, the better performance. Table 2 shows the AUCs of all compared methods on the Twitter and Sina Weibo datasets. We observe that GANG consistently outperforms all compared methods on both datasets. We note that CatchSync achieves a close AUC as GANG on the Twitter dataset. However, CatchSync's performance degrades substantially on the Sina Weibo dataset. CatchSync relies on node degrees and properties of a node's neighbors. Therefore,

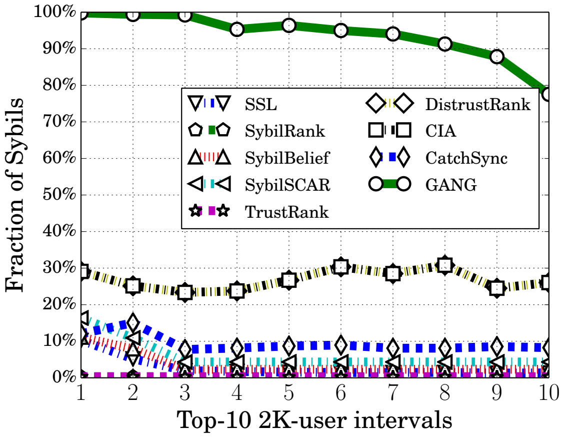


Figure 1: Fraction of fraudulent nodes in each top ranked interval on the Twitter dataset.

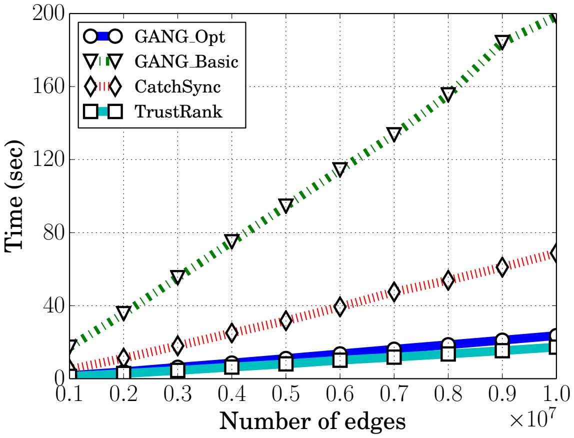
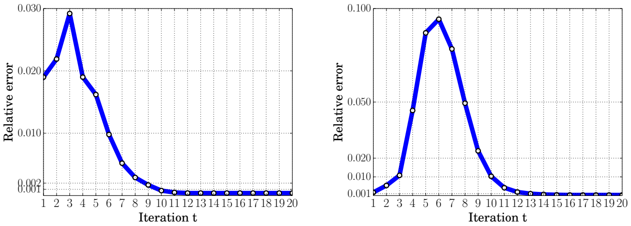


Figure 2: Running time of directed graph based methods on synthesized graphs with increasing number of edges. DistrustRank and CIA have almost identical results with TrustRank, and thus we omit their results for conciseness.

we suspect the reason for CatchSync's poor performance on Sina Weibo is that nodes in the Sina Weibo dataset have larger average node degrees and their neighbors have more diverse properties.

Fraudulent nodes in top-ranked nodes: In practice, the ranking of nodes can be used as a priority list to help OSNs' human workers manually inspect nodes and detect fraudulent nodes. Inspecting nodes according to their rankings could aid human workers to detect more fraudulent nodes than inspecting nodes picked uniformly at random, within the same amount of time. When ranking is used for such purpose, the number of fraudulent nodes in top-ranked nodes is important because human workers can only inspect a limited number of nodes.

AUC measures the overall ranking performance, but it cannot tell fraudulent nodes among the top-ranked nodes. Therefore, we further compare the considered methods using the fraction of fraudulent nodes in top-ranked nodes. In particular, we divide the top-20K nodes into 10 intervals, where each interval has 2K nodes. Figure 1 shows the fraction of fraudulent nodes in each interval for the Twitter dataset. Since the Sina Weibo dataset does not have enough labeled nodes to draw a similar graph, we omit its corresponding results. GANG achieves the best performance and substantially outperforms other methods. Specifically, the fraction of fraudulent nodes detected by GANG ranges from 77.5% to 99.8% in the top-10 2K-node intervals. The superiority of GANG comes from that GANG leverages unidirectional edges and GANG utilizes both labeled fraudulent and normal users.



(a) Twitter

(b) Sina Weibo

Figure 3: GANG's relative errors of residual posterior beliefs vs. number of iterations. GANG converges.

Table 3: Labeling results of the 1K nodes that are sampled from the top-ranked 100K nodes for Sina Weibo.

Category		**Percentage**	
Fraudulent users	Suspended users	41.5%	**92.0%**
	Spammers	42.5%	
	Compromised users	8.0%	
Normal users		**6.8%**	
Unknown users		**1.2%**	

4.3 Convergence

Figure 3 shows GANG's relative errors of residual posterior beliefs in two consecutive iterations, i.e., $\|\hat{\mathbf{p}}^{(t)} - \hat{\mathbf{p}}^{(t-1)}\|_1 / \|\hat{\mathbf{p}}^{(t)}\|_1$, as a function of the number of iterations t. We observe that the relative error first increases, then decreases, and finally converges on both datasets.

4.4 Scalability

We measure the scalability of compared directed graph based methods with respect to the number of edges in the graph. Since we need graphs with different number of edges, but the Twitter and Sina Weibo datasets have fixed number of edges, we synthesize graphs according to a Preferential Attachment (PA) model. We note that there are more advanced network models (e.g., the one proposed by Gong et al.) to synthesize more realistic graphs. However, since the scalability does not depend on the graph structures, we use the simple PA model to synthesize graphs. All the compared methods involve iterative computing processes, e.g., TrustRank, DistrustRank, and CIA iteratively compute random walks, while CatchSync relies on the iterative HITS algorithm. For fair comparison, we run the iterative processes with the same number of iterations. Figure 2 shows the running time used by the directed graph based methods (GANG_Basic and GANG_Opt are the basic and optimized versions of GANG, respectively) when we increase the number of edges in the synthesized graph.

First, GANG_Opt is slightly less efficient than random walk based methods TrustRank, DistrustRank, and CIA. This is because, in each iteration, these methods traverse each unidirectional edge once while GANG traverses twice. Second, GANG_Opt is more scalable than CatchSync. This is because CatchSync first uses HITS, which already has the same time complexity with GANG_Opt, to compute nodes' hubness and authoritativeness scores, and then CatchSync further computes each node's *synchronicity*, which involves going through node pairs between a node's outgoing neighbors, and *normality*, which involves going through node pairs between a node's outgoing neighbors and all nodes. Third, GANG_Opt is one order of magnitude more scalable than GANG_Basic.

4.5 Case Study on Sina Weibo

We apply our GANG to the Sina Weibo dataset and manually inspect the detected fraudulent nodes. Specifically, we use all the 1980 labeled nodes as a training dataset and produce a ranking list for the remaining nodes. Then we sample 1K nodes from the top-ranked 100K nodes uniformly at random, and we manually inspect them. Table 3 shows the labeling results of the 1K nodes.

Table 4: Manual labeling results of the 1,000 users that are sampled from the top ranked 100K users uniformly at random.

Category		**Percentage**	
Sybil	Suspended users	41.5%	**92.0%**
	Spammers	32.4%	
	Content-evading spammers	10.1%	
	Compromised users	8.0%	
Benign		**6.8%**	
Unknown		**1.2%**	

1. Suspended users: These users didn't exist any more at the time of our inspection. They could be suspended/deleted by Weibo or the users themselves.

2. Spammers: These users post or share a large amount of advertisements, e.g., to promote their products or to sell pirated products. They often post a large amount of tweets, and almost all of them are advertisements. These users violate Weibo's policy,[1] and thus they should be suspended/deleted by Weibo according to its policy.

3. Content-evading spammers: These users post or share some advertisements. However, they also post a large number of *seemingly normal tweets*. We found that some of these users posted seemingly normal tweets at different random times; each time they posted, they posted multiple tweets within one minute. We call such behavior *clustered tweeting*, which is suspicious. Moreover, we randomly sampled some of these seemingly normal tweets, and used them as keywords to search on Baidu (the largest search engine in China). We found that these seemingly normal tweets were simply copied from Internet. Figure 6(a) shows a user who performs clustered tweeting, and Figure 6(b) shows the search results of one tweet of the user, which indicates that the tweet was simply copied from Internet.

Some users in this category posted seemingly normal tweets periodically. For instance, Figure 6(c) shows an example user who posted seemingly normal tweets every 9 hours and 13 minutes. We call such behavior *periodic tweeting*. Again, we randomly sampled some seemingly normal tweets posted by users through periodic tweeting, and we used them as keywords to search on Baidu. We found that these tweets were also simply copied from Internet. Figure 6(d) shows the search results of one tweet posted by the user shown in Figure 6(c).

We suspect these users are trying to evade content-based detection through posting seemingly normal tweets. Therefore, we call them *content-evading* users. Our observations imply that we could design a method based on tweeting behavior (e.g., clustered tweeting and periodic tweeting) to filter these spammers.

4. Compromised users: These users posted normal tweets about daily activities before a certain time point, and then they started to post or share a large amount of advertisements. We also randomly sampled some tweets of these users, but we could not find them on the Baidu search engine. Therefore, we suspect that these users could be compromised normal users.

5. Normal users: These users post normal tweets and comply with Weibo's policy.

6. Unknown users: These users did not have any tweets at the time of our inspection, so we were unable to characterize them through content.

4.6 Sybil Ranking Results

We use SybilDirect to produce a ranking list of the remaining unlabeled users in the Weibo dataset by taking the manually labeled training dataset as an input. We use the fraction of Sybils at top segments of the ranking list to evaluate SybilDirect. We do so by manually label some sampled users in each particular interval of the ranking list. Different methods will produce different ranking lists, and thus evaluating other methods requires manually inspecting sampled users in their ranking lists. Therefore, due to limited availability of human verifiers, we do not apply other methods to the Weibo dataset.

Manually inspecting top ranked users: We evaluate the performance of SybilDirect using the top ranked 100K users. Specifically, we divide the 100K users into 10 10K-user intervals. We sample 100 users from each interval uniformly at random, and thus we have 1,000 sampled users in the 10 intervals in total. We manually inspected Weibo pages of the 1,000 sampled users in April 2016, and we labeled them to be one of the six categories we defined. Specifically, if we cannot label a sampled user to be spammer, content-evading spammer, compromised user, or unknown, we labeled him/her to be normal. Table 4 shows the details of these 1,000 sampled users. We find that 92% of them are Sybils. Moreover, SybilDirect ranks a large number of Sybils that evaded Weibo's

[1] http://weibo.cn/dpool/ttt/h5/regagreement.php

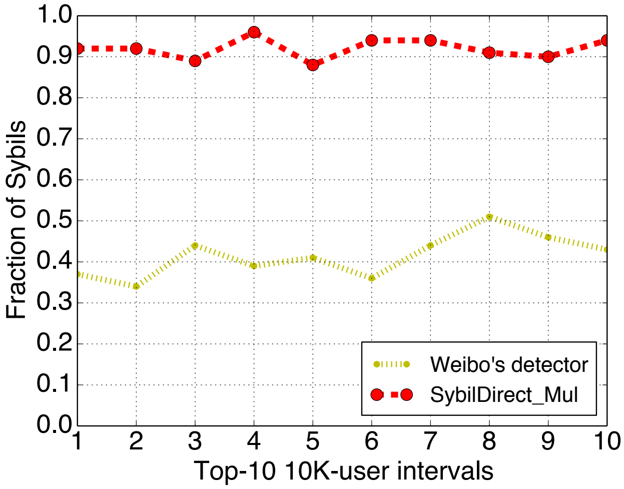


Figure 4: Fraction of Sybils in each of the top-10 10K-user intervals for SybilDirect on the Weibo dataset.

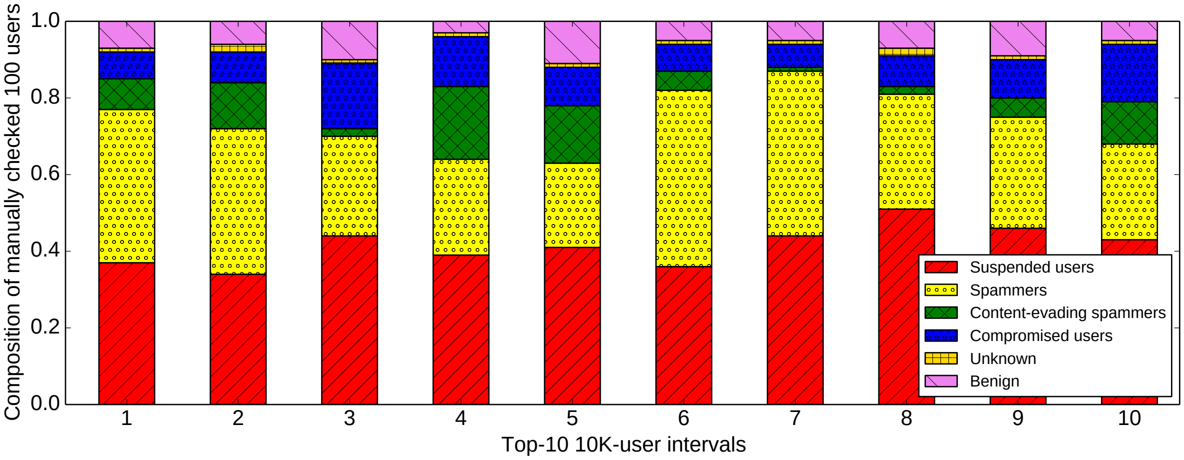
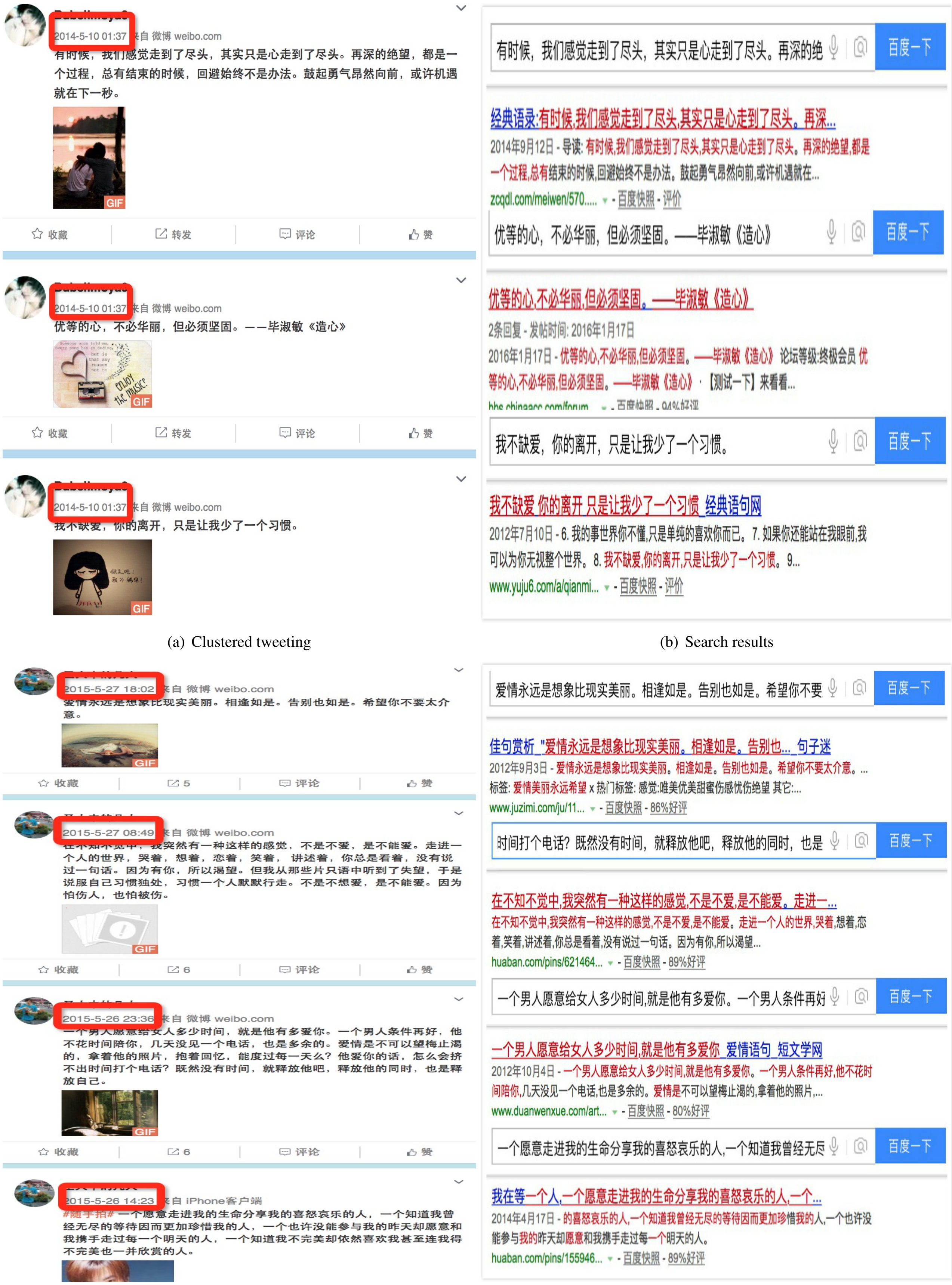


Figure 5: Composition of the 100 sampled users in each top ranked 10K-user interval of the ranking list produced by SybilDirect.

detector at the top. In particular, 32% are spammers, 10% are content-evading spammers, and 8% are compromised users, all of which evaded Weibo's detector (because they were still active at the time of our inspection). Figure 4 shows the fraction of Sybils in the sampled 100 users in each 10K-user interval. Figure 5 in Appendix shows the details about each of the six user categories in each 10K-user interval. Even if the training dataset is small, about 90%+ of users in each top ranked interval are Sybils.

Analyzing the structure of the top-100K users: We find that there are 18 weakly connected components in the subgraph consisting of the top-100K users and edges between them. The largest weakly connected component includes 98904 (98.9%) users. In other words, the top-ranked Sybils (together with some benign users) form a densely connected component, which is the reason why our method can rank them at the top. Via analyzing the compromised users in the sampled 1,000 users, we find that the compromised users link to spammers/content-evading spammers to forward their spams, which make them be ranked at the top by our method.

Comparing with Sina Weibo's detector: When Sina Weibo's detector detects a fraudulent user, the user will be suspended or deleted. In other words, the number of fraudulent nodes detected by Sina Weibo's detector is upper bounded by the category *suspended users*, and the users in the category *spammers* and *compromised users* have evaded Sina Weibo's detector. However, our method GANG can detect these fraudulent users.


收藏
转发
评论
赞
GIF
Someone once told me, Every song has an ending, but is that any reason not to enjoy the music?
百度一下
百度快照
评价

2014-5-10 01:37 来自 微博 weibo.com
有时候，我们感觉走到了尽头，其实只是心走到了尽头。再深的绝望，都是一个过程，总有结束的时候，回避始终不是办法。鼓起勇气昂然向前，或许机遇就在下一秒。

2014-5-10 01:37 来自 微博 weibo.com
优等的心，不必华丽，但必须坚固。——毕淑敏《造心》

2014-5-10 01:37 来自 微博 weibo.com
我不缺爱，你的离开，只是让我少了一个习惯。

(a) Clustered tweeting

有时候，我们感觉走到了尽头，其实只是心走到了尽头。再深的绝

经典语录:有时候,我们感觉走到了尽头,其实只是心走到了尽头。再深...
2014年9月12日 - 导读: 有时候,我们感觉走到了尽头,其实只是心走到了尽头。再深的绝望,都是一个过程,总有结束的时候,回避始终不是办法。鼓起勇气昂然向前,或许机遇就在...
zcqdl.com/meiwen/570..... - 百度快照 - 评价

优等的心，不必华丽，但必须坚固。——毕淑敏《造心》

优等的心,不必华丽,但必须坚固。——毕淑敏《造心》
2条回复 - 发帖时间: 2016年1月17日
2016年1月17日 - 优等的心,不必华丽,但必须坚固。——毕淑敏《造心》 论坛等级:终极会员 优等的心,不必华丽,但必须坚固。——毕淑敏《造心》 · 【测试一下】来看看...
bbs.chinaacc.com/forum ... - 百度快照 - 94%好评

我不缺爱，你的离开，只是让我少了一个习惯。

我不缺爱 你的离开 只是让我少了一个习惯_经典语句网
2012年7月10日 - 6. 我的事世界你不懂,只是单纯的喜欢你而已。7. 如果你还能站在我眼前,我可以为你无视整个世界。8. 我不缺爱,你的离开,只是让我少了一个习惯。9...
www.yuju6.com/a/qianmi... - 百度快照 - 评价

(b) Search results

2015-5-27 18:02 来自 微博 weibo.com
爱情永远是想象比现实美丽。相逢如是。告别也如是。希望你不要太介意。
5

2015-5-27 08:49 来自 微博 weibo.com
在不知不觉中，我突然有一种这样的感觉，不是不爱，是不能爱。走进一个人的世界，哭着，想着，恋着，笑着， 讲述着，你总是看着，没有说过一句话。因为有你，所以渴望。但我从那些片只语中听到了失望，于是说服自己习惯独处，习惯一个人默默行走。不是不想爱，是不能爱。因为怕伤人，也怕被伤。
6

2015-5-26 23:36 来自 微博 weibo.com
一个男人愿意给女人多少时间，就是他有多爱你。一个男人条件再好，他不花时间陪你，几天没见一个电话，也是多余的。爱情是不可以望梅止渴的，拿着他的照片，抱着回忆，能度过每一天么？他爱你的话，怎么会挤不出时间打个电话？既然没有时间，就释放他吧，释放他的同时，也是释放自己。
6

2015-5-26 14:23 来自 iPhone客户端
#随手拍# 一个愿意走进我的生命分享我的喜怒哀乐的人，一个知道我曾经无尽的等待因而更加珍惜我的人，一个也许没能参与我的昨天却愿意和我携手走过每一个明天的人，一个知道我不完美却依然喜欢我甚至连我得不完美也一并欣赏的人。

(c) Periodic tweeting

爱情永远是想象比现实美丽。相逢如是。告别也如是。希望你不要

佳句赏析_"爱情永远是想象比现实美丽。相逢如是。告别也..._句子迷
2012年9月3日 - 爱情永远是想象比现实美丽。相逢如是。告别也如是。希望你不要太介意。...
标签: 爱情美丽永远希望 x 热门标签: 感觉:唯美优美甜蜜伤感忧伤绝望 其它:...
www.juzimi.com/ju/11... - 百度快照 - 86%好评

时间打个电话？既然没有时间，就释放他吧，释放他的同时，也是

在不知不觉中,我突然有一种这样的感觉,不是不爱,是不能爱。走进一...
在不知不觉中,我突然有一种这样的感觉,不是不爱,是不能爱。走进一个人的世界,哭着,想着,恋着,笑着,讲述着,你总是看着,没有说过一句话。因为有你,所以渴望...
huaban.com/pins/621464... - 百度快照 - 89%好评

一个男人愿意给女人多少时间,就是他有多爱你。一个男人条件再好

一个男人愿意给女人多少时间,就是他有多爱你_爱情语句_短文学网
2012年10月4日 - 一个男人愿意给女人多少时间,就是他有多爱你。一个男人条件再好,他不花时间陪你,几天没见一个电话,也是多余的。爱情是不可以望梅止渴的,拿着他的照片,...
www.duanwenxue.com/art... - 百度快照 - 80%好评

一个愿意走进我的生命分享我的喜怒哀乐的人,一个知道我曾经无尽

我在等一个人,一个愿意走进我的生命分享我的喜怒哀乐的人,一个...
2014年4月17日 - 的喜怒哀乐的人,一个知道我曾经无尽的等待因而更加珍惜我的人,一个也许没能参与我的昨天却愿意和我携手走过每一个明天的人。
huaban.com/pins/155946... - 百度快照 - 89%好评

(d) Search results

Figure 6: (a) A user performing clustered tweeting. (b) Search results of one of the user's tweet on Baidu. (c) A user performing periodic tweeting. (d) Search results of one of the user's tweet on Baidu.