Spectrum Based Fraud Detection in Social Networks

[Extended Abstract]

Xiaowei Ying, Xintao Wu
UNC Charlotte
9201 Univ. City Blvd
Charlotte, NC 28223
{xying,xwu}@uncc.edu

Daniel Barbará
George Mason University
4400 University Dr.
Fairfax, VA 22303
dbarbara@gmu.edu

ABSTRACT

Social networks are vulnerable to various attacks such as spam emails, viral marketing and the such. In this paper we develop a spectrum based detection framework to discover the perpetrators of these attacks. In particular, we focus on Random Link Attacks (RLAs) in which the malicious user creates multiple false identities and interactions among those identities to later proceed to attack the regular members of the network. We show that RLA attackers can be filtered by using their spectral coordinate characteristics, which are hard to hide even after the efforts by the attackers of resembling as much as possible the rest of the network. Experimental results show that our technique is very effective in detecting those attackers and outperforms techniques previously published.

Categories and Subject Descriptors

H.2.0 [**Database Management**]: General–Security, integrity, and protection

General Terms

Algorithms, Security, Theory

Keywords

Fraud Detection, Spectrum, Social Networks

1. INTRODUCTION

Social networks have always been vulnerable to various attacks including spam emails, annoying telemarketing calls, viral marketing, and individual re-identification in anonymized social network publishing. Recently, the authors in [4] provided a general abstraction, called the *Random Link Attack* (RLA), which identifies the collaborative nature of these attacks to evade detection. In an RLA, the malicious user creates a set of false identities and uses them to connect with a large set of victim nodes. To evade detection, the malicious user also creates various interactions among false identities, which make the subgraph formed by false identities similar to that formed by regular users. This property makes the discovery of the attack and the responsible entities a difficult task.

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

In this paper, we develop a spectrum based fraud detection framework to identify various attacks. Our approach, which exploits the spectral space of the underlying interaction structure of the network, is different from traditional topological analysis approaches [2, 4]. Traditional topology based detection methods explore the graph topology directly and discover abnormal connectivity patterns caused by attacks. Our approach is based on graph spectral analysis that deals with the analysis of the spectra (eigenvalues and eigenvector components) of the adjacency matrix. We study how to identify attackers by characterizing their distributions in the spectral space. Our theoretical results show that attackers locate in a different region of the spectral space from regular users. Specifically, the spectral coordinate of an attacker is mainly determined by that of its victims. The inner structure among collaborative attackers has little impact on attackers' distributions in the spectral space. By identifying fraud patterns in graph spectral spaces, we can detect various collaborative attacks that are hard to be identified from original topological structures.

Focusing on RLAs, we show how the identities of the perpetrators of the attack can be filtered using their spectral characteristics. We then show how to identify the RLAs from the set of suspects obtained by their spectral characteristics. We develop an efficient algorithm, SPCTRA, which utilizes a single measure called *node non-randomness*. We compare our spectrum-based attack detection approach with the topology based detection approach [4]. Empirical evaluations show that our approach significantly improves both effectiveness and efficiency especially when a mix of RLAs are introduced.

2. A SPECTRUM BASED FRAMEWORK FOR DETECTING ATTACKS

A network or graph G is a set of n nodes connected by a set of m links. It can be represented as the symmetric adjacency matrix $A_{n\times n}$ with $a_{ij} = 1$ if node i is connected to node j and $a_{ij} = 0$ otherwise. Graph spectral analysis deals with the analysis of the spectra (eigenvalues and eigenvector components) of the nodes in the graph. Let λ_j be the eigenvalues of the adjacency matrix A and $\boldsymbol{x}_j$ the corresponding eigenvectors, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. The spectral decomposition of A is $A = \sum_j \lambda_j \boldsymbol{x}_j \boldsymbol{x}_j^T$. Let x_{ju} denote the u-th entry of $\boldsymbol{x}_j$. The eigenvector $\boldsymbol{x}_j = (x_{j1}, x_{j2}, \cdots, x_{jn})^T$ is represented as a column vector. The row vector $(x_{1u}, x_{2u}, \cdots, x_{nu})$ represents the coordinate of node u in the n-dimensional spectral space.

In our framework, we exploit the spectral space and char-

acterize the difference between the spectral coordinates of regular users and that of attackers, rather than exploring the graph topology directly.

In a collaborative attack, the malicious user has complete control over the attacking nodes and uses them to attack (e.g., send emails) a large set of victim nodes. Assume there are c ($c \ll n$) attacking nodes and they form a subgraph with adjacency matrix $C = \{c_{ij}\}_{c\times c}$. The outgoing links from attacking nodes to regular nodes form the subgraph with adjacency matrix $B = (b_{ij})_{n\times c}$: $b_{ij} = 1$ if the j-th attacking node has a link to the i-th regular node, and $b_{ij} = 0$ otherwise. The graph after attacks $\tilde{G}$ has $N = n+c$ nodes, and we can arrange the nodes in the graph so that node 1 to c are attacking nodes and node $c+1$ to N are regular ones. We have:

$$A = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_n \end{pmatrix}, \tilde{A} = \begin{pmatrix} C & B^T \\ B & A_n \end{pmatrix}, E = \begin{pmatrix} C & B^T \\ B & \mathbf{0} \end{pmatrix} \quad (1)$$

Let $\boldsymbol{z}_j$ be the eigenvector of A associated to λ_j, and $\tilde{\boldsymbol{z}}_j$ be the eigenvector of $\tilde{A}$ associated to eigenvalue $\tilde{\lambda}_j$. Then,

$$\boldsymbol{z}_j = \begin{pmatrix} \mathbf{0}_{c\times 1} \\ \boldsymbol{x}_j \end{pmatrix}, \quad \tilde{\boldsymbol{z}}_j = \begin{pmatrix} \tilde{\boldsymbol{y}}_j \\ \tilde{\boldsymbol{x}}_j \end{pmatrix},$$

where $\tilde{\boldsymbol{y}}_j = (\tilde{y}_{j1}, \ldots, \tilde{y}_{jc})^T$ denotes the entries corresponding to the attackers in $\tilde{\boldsymbol{z}}_j$ and $\tilde{\boldsymbol{x}}_j = (\tilde{x}_{j1}, \ldots, \tilde{x}_{jn})^T$ denotes the entries corresponding to those regular nodes in $\tilde{\boldsymbol{z}}_j$. Since A is expanded by adding 0's into A_n, $\boldsymbol{x}_j$ is then the eigenvector of A_n along with the eigenvalue λ_j. Let $\bar{x}_j$ be the mean value of entries in $\boldsymbol{x}_j$: $\bar{x}_j = \frac{1}{n}\mathbf{1}_n^T \boldsymbol{x}_j$. To make the deduction simple, we choose the sign of $\boldsymbol{x}_j$ so that $\bar{x}_j \geq 0$.

We denote $\boldsymbol{\alpha}_u = (x_{1u}, x_{2u}, \cdots, x_{ku})$ as the spectral coordinate of regular node u in the original spectral space. Denote $\boldsymbol{\beta}_i = (y_{1i}, y_{2i}, \cdots, y_{ki})$ as the spectral coordinate of attacking node i. Since we assume there is no attack in the original graph, $\boldsymbol{\beta}_i$ is actually a zero vector. Similarly, we denote $\tilde{\boldsymbol{\alpha}}_u = (\tilde{x}_{1u}, \tilde{x}_{2u}, \cdots, \tilde{x}_{ku})$ and $\tilde{\boldsymbol{\beta}}_i = (\tilde{y}_{1i}, \tilde{y}_{2i}, \cdots, \tilde{y}_{ki})$ as the spectral coordinate of regular node u and attacking node i in the perturbed spectral space respectively.

Result 1. *In a graph $\tilde{G}$ under collaborative attacks, for attacking node i, $1 \leq i \leq c$, $\tilde{y}_{ji}$ can be approximated by:*

$$\tilde{y}_{ji} \approx \frac{1}{\lambda_j} \sum_{u\in\Omega_i} x_{ju} + \frac{1}{\lambda_j^2} \sum_{r=1}^{c} \left(c_{ir} \sum_{u\in\Omega_r} x_{ju} \right), \quad (2)$$

where Ω_r denotes the victim set of attacking node r. For any regular node u, $1 \leq u \leq n$, $\tilde{x}_{ju}$ is approximately unchanged: $\tilde{x}_{ju} \approx x_{ju}$.

Result 1 shows that the spectral coordinate of an attacking node can be approximated by the spectral coordinates of its victims. When attackers do not collaborate with each other ($C = \mathbf{0}_{c\times c}$), the second term of the right hand side of (2) disappears. We simply have $\tilde{y}_{ji} \approx \frac{1}{\lambda_j}\sum_{u\in\Omega_i} x_{ju}$, which indicates the attacker's spectral coordinate is fully determined by that of its victims. From (2) we can also observe that the inner structure C among the attackers only affects $\tilde{y}_{ji}$ in the order of λ_j^{-2}. When λ_j is large, the second term of the right hand side of (2) is already negligible, which means that the inner subgraph structure has little impact on the distribution of attackers in the spectral space.

The above result is mathematically elegant. However, in practice users have no knowledge about which nodes are attackers (or victims). In the next section, we focus on RLAs and show that the distribution of attackers' spectral coordinates are determined by $\tilde{\boldsymbol{z}}_j$ and $\tilde{\lambda}_j$, which can be calculated directly from the observed graph $\tilde{A}$.

3. DETECTING RANDOM LINK ATTACK

In a RLA, the malicious user creates $c(\ll n)$ false identities (attacking nodes) and uses them to connect with a large set of victims. Attacking node i randomly attack v_i victims and each regular node has the same probability to be a victim. The total number of victims is $v = \sum_{i=1}^{c} v_i$. To evade detection, the malicious user also creates m_c links among attacking nodes, which may make the subgraph formed by attacking nodes similar to that formed by regular users.

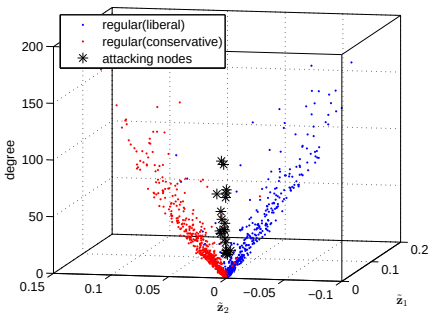

Figure 1: Spectral coordinates of political blogsphere data under a degree attack with 20 attackers.

Figure 1 plots the node spectral coordinates under a degree attack[1] with 20 attackers on the political blogosphere network. We also show node degrees in the z-axis. We can observe from the figure that the majority of nodes projected in the 2-D spectral space distribute along two straight and quasi-orthogonal lines while attacking nodes (denoted as black) locate between the two quasi-orthogonal lines in the spectral projection space.

In this section, we investigate how attackers distribute in the spectral space. By identifying the distribution of attackers' spectral coordinates, we expect to separate attacking nodes from regular ones in the spectral space. Our theoretical result shows that $\tilde{y}_{ji}$ follows the normal distribution whose mean and variance satisfy the following two inequalities:

$$\mathbf{E}(\tilde{y}_{ji}) \leq \frac{d_i \bar{x}_j}{\lambda_j}, \quad \mathbf{V}(\tilde{y}_{ji}) \leq \frac{d_i}{n}\left(1 - \frac{d_i}{n}\right)\frac{1}{\lambda_j^2}. \quad (3)$$

We can regard node i as a suspect if the corresponding entry $\tilde{y}_{ji}$ is within the confidence interval $[\mathbf{E}(\tilde{y}_{ji}) - \epsilon\sqrt{\mathbf{V}(\tilde{y}_{ji})}, \mathbf{E}(\tilde{y}_{ji}) + \epsilon\sqrt{\mathbf{V}(\tilde{y}_{ji})}]$ where $\epsilon > 0$ denotes the $\frac{1+p}{2}$ quantile of the standard normal distribution (i.e., interval $[-\epsilon, \epsilon]$ covers probability p).

Our spectrum based detection algorithm called SPCTRA consists of three steps. In the first step, we conduct node non-randomness test to identify suspects. In the second step, from G_{susp} we identify suspect groups as candidates of RLA groups. In the third step, we test whether each dense subgraph is a true RLA group. As a result, we can filter out dense subgraphs accidentally formed by regular nodes.

[1]Attacking nodes have the same degree distribution as the regular nodes. For attacking node i, it attacks $\frac{2*d_i}{3}$ victims.

Complexity. Our algorithm involves the calculation of the first k eigenvectors of a graph. In general, eigen-decomposition of an $n \times n$ matrix takes a number of operations $O(n^3)$. In our framework, we only need calculate the first k largest eigenvalues and their eigenvectors. We implemented the Arnoldi/Lanczos algorithm which generally needs $O(n)$ rather than $O(n^2)$ floating point operations at each iteration. The authors in [4] developed the GREEDY algorithm to catch RLAs from the suspect set. The GREEDY algorithm is to mine subgraphs satisfying the RLA-property, starting from the suspect nodes identified by two tests. It grows a potential attack cluster by iteratively adding nodes with a high degree of connectivity with the cluster. The time complexity of the neighborhood independence test is $O(\sum_i d_i^2) = O(m^2)$. To catch attacking groups among n_{susp} suspects, the GREEDY needs $O(m_{\text{susp}}^2)$ time.

4. EXPERIMENTAL RESULTS

Data Set and Setting. We conducted experiments on the Web Spam Challenge 2007 data, which contains over 105 million pages in 114,529 hosts in the .UK domain. The number of links among these hosts is 1,836,228. We implemented our spectrum based detection algorithm and the topology based detection algorithm [4] (including two testing procedures, *clustering test* and *neighborhood independence test*, and the GREEDY algorithm) in Matlab. Our experiments were carried out on a Windows XP64 workstation with a 3.0 GHz Pentium-IV CPU and 2GB RAM.

Table 1: Evaluation results on Web Spam data set, 8 RLA attacking groups, 650 total attackers, and 56144 total victims.

RLA	setting		SPCTRA		GREEDY	
	size	$\bar{v}_i$ p_{in}	ssp	atck	ssp	atck
1	50	100 .3	50	50	49	47
2	50	100 .6	50	50	0	0
3	50	100 1	50	50	50	50
4	50	200 .3	50	50	79	47
5	100	100 .3	100	100	3	3
6	50	degree	49	49	20	20
7	100	degree	97	97	6	6
8	200	degree	188	188	27	27
final results (total)			634	634	4534	200

Accuracy of Detecting RLAs. We generated 8 RLAs with varied sizes and connection patterns (links between attackers and victims and internal links among the attackers), as shown in Table 1. The total number of attacking nodes is 650 and the size of victims is 56,144. Our goal is to test whether algorithms (SPCTRA and GREEDY) can catch them and how accurate they achieve. Each algorithm output a set of suspect groups (i.e., RLA candidates). Table 1 shows our detailed comparisons.

Table 2: Execution time (in seconds) of for different data sets

Data set	Alg.	Testing	Grouping	Total
polblogs (1222, 16714)	SPCTRA	0.037	0.041	0.078
	GREEDY	16.20	6.047	22.24
Web Spam (33%) (37562, 199406)	SPCTRA	0.702	0.239	0.941
	GREEDY	577.2	515.4	1093
Web Spam (114529, 1836228)	SPCTRA	4.017	29.68	33.69
	GREEDY	12728	83314	96043

Running Time. In this experiment, we compare the running times of the SPCTRA and GREEDY algorithms using three data sets, *polblogs*, Web Spam, and a sample of Web Spam data. In Table 2, we report the running time for both SPCTRA and GREEDY including the testing step (catching suspects) and the grouping step (catching RLAs). We can see that the time taken by GREEDY is 285, 1161, and 2851 times more than our SPCTRA algorithm.

5. CONCLUSIONS AND FUTURE WORK

We have presented a novel framework that exploits the spectral space of underlying network topology to identify frauds or attacks. Our theoretical results showed that attackers locate in a different region of the spectral space from regular users. By identifying fraud patterns in graph spectral spaces, we can detect various collaborative attacks that are hard to be identified from original topological structures. Focusing on RLAs, we presented an efficient algorithm, SPCTRA, and compared with the topology based detection approach [4]. Empirical evaluations show that our approach significantly improves both effectiveness and efficiency especially when a mix of RLAs are introduced. In our future work, we will explore various other attacking scenarios in both social networks and communication networks. Specifically, we will study how our spectrum based detection works when attackers choose victims purposely (e.g., *passive and active attacks*[1]) or only attack very few victims when they launch their collaborative attacks (e.g.,*Sybil attack* [3]). We will extend our approach to use the temporal information in dynamics networks to identify and catch potential attacks. We will explore matrix visualization and organization approaches that enable interactive navigation between network topology and its spectral spaces.

Acknowledgment

This work was supported in part by U.S. National Science Foundation CCF-1047621 and CNS-0831204. Refer to [5] for details on theoretical analysis, algorithm, and evaluations.

6. REFERENCES

[1] L. Backstrom, C. Dwork, and J. Kleinberg. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *WWW '07*, pages 181–190, 2007.

[2] D. H. Chau, S. Pandit, and C. Faloutsos. Detecting fraudulent personalities in networks of online auctioneers. In *PKDD*, pages 103–114, 2006.

[3] J. Newsome, E. Shi, D. Song, and A. Perrig. The sybil attack in sensor networks: analysis & defenses. In *Proceedings of the third international symposium on Information processing in sensor networks*, pages 259Ű–268, 2004.

[4] N. Shrivastava, A. Majumder, and R. Rastogi. Mining (social) network graphs to detect random link attacks. In *ICDE*, pages 486–495, 2008.

[5] X. Ying, X. Wu, and D. Barbará. Spectrum Based Fraud Detection in Social Networks. In *Technical Report, College of Computing and Informatics, UNC Charlotte*, 2010.