53rd IEEE Conference on Decision and Control
December 15-17, 2014. Los Angeles, California, USA

Model-based Fraud Detection in Growing Networks

Pablo Moriano
School of Informatics and Computing
Indiana University
Bloomington, IN, USA
moriano@ieee.org

Jorge Finke
Department of Electrical Engineering and
Computer Science
Pontificia Universidad Javeriana
Santiago de Cali, Colombia
finke@ieee.org

Abstract— **People share opinions, exchange information, and trade services on large, interconnected platforms. As with many new technologies these platforms bring with them new vulnerabilities, often becoming targets for fraudsters who try to deceive randomly selected users. To monitor such behavior, the proposed algorithm evaluates structural anomalies that result from local interactions between users. In particular, the algorithm evaluates the degree of membership to well-defined communities of users and the formation of close-knit groups in their neighborhoods. It identifies a set of suspects using a first order approximation of the evolution of the eigenpairs associated to the continuously growing network. Within the set of suspects, the algorithm them locates fraudsters based on deviations from the expected local clustering coefficients. Simulations illustrate how incorporating asymptotic behavior of the structural properties into the design of the algorithm allows us to differentiate between the aggregate dynamics of fraudsters and regular users.**

I. INTRODUCTION

Every day hundreds of thousands of people interact on online platforms. Their scalable architecture offers a highly customizable experience for end users and a cost-effective solution for service providers. As the cost of technology drops, it is envisioned that even more users will engage in an even larger online exchange of information and services. And as the number of these transactions grow, attempts to exploit vulnerable platforms will have a profound impact on service providing businesses and government agencies.

For over two decades, data mining has provided different approaches for designing algorithms to visualize, organize, segment, and predict deceptive activities [1]. Substantial progress has been made in developing neural network and machine-learning techniques to mine vast arrays of data based on transactional signatures [2]. Generally speaking, signature-based detection techniques allow analysts to distinguish the characteristics of users who, in a similar manner to past instances, attempt to gain an unfair advantage over regular users [3]. Machine-learning algorithms generate graphical decision trees, which are a useful tool to identify the signatures of deceptive transactions (e.g., to extract profiles of suspects in the form of if-then rules) [4]. Nonetheless, if the design of signature-based algorithms ignores the growing complexity of the interaction between fraudsters and regular users, detection generally falls short (e.g., see [5] and [6]).

A more realistic assumption is to hold that fraudsters interact with a small set of other fraudsters (or false users). They mislead reputation mechanisms by mutually boosting their evaluation scores, eventually gaining trust and engaging in deceptive transactions with regular users. To tackle such schemes, detection techniques must examine the collective nature of fraud in relation to the expected aggregate behavior of regular users. Transactions can be represented as a growing network in which a link represents a service agreement between two users and the attachment of a node the addition of a new user.

Our premise is that fraudsters perform random link attacks (RLAs), which represent attempts to deceive randomly selected regular users and give rise to detectable anomalies in the structure of the network. It is envisioned that modeling the behavioral characteristics of regular users serves as a framework to provide a formal assurance for detection without the use of transactional signatures.

Past efforts along these lines include the work in [7], which introduces an algorithm that relies on measures of both (i) local clustering (to identify non-collaborative attacks) and (ii) neighborhood independence (to identify attacks by larger sets of common neighbors between the victims). Taking into account both properties ensures that fraudsters cannot easily resemble interaction patterns by simply forging strongly clustered neighborhoods. The work in [8]-[10] proposes an alternative approach, which focuses on the division of the network into modules or communities (i.e., densely connected groups of users, with only a sparse number of connections between them) [8]. In particular, the spectrum-based models in [9], [10] show that fraudsters who perform RLAs are located in a region that can be separated (to some extent) from regular users. The approach is based on (i) identifying suspects according to a measure of node non-randomness (to capture which users seem to unambiguously belong to a specific community); and (ii) filtering the resulting group of suspects (under the assumption that fraudsters are likely to form dense subgraphs). Unlike the work in [7]-[10], the approach presented in this paper focuses on evaluating local network properties that emerge as new users become part of the network and interact with each other over time.

The approach is novel in that it models the presumed behavior of regular users and fraudsters (based on connectivity trends found in empirical data for scenarios where RLAs lead to changes in the clustering properties of the network [11], [12]). Conditions for fraud detection depend on the

978-1-4673-6090-6/14/$31.00 ©2014 IEEE

measure of node non-randomness presented in [9], [10] and an analytical expression for the local clustering coefficient.

The remaining sections are organized as follows: Section 2 defines the problem and discusses some assumptions on the available data. Section 3 proposes a model that captures individuals tending to form triad junctions as they interact with other individuals. Theorem 1 characterizes the asymptotic behavior of the local clustering coefficient of the nodes of the undirected network of interactions. Section 4 introduces the detection algorithm based on Theorem 1. Section 5 presents the receiver operating curves (ROC) for different network conditions and detection parameters. In particular, we compare the proposed algorithm with the work in [9]. Finally, Section 6 draws some conclusions and future research directions.

II. Problem definition

Our objective is to identify fraudsters based on anomalies in the structure of the interactions between users. Specifically, let us define the detection problem as follows.

Given:

(i) A dynamic dataset that captures the transactions between regular users, who are naturally divided into two communities with highly clustered neighborhoods.

(ii) Some exposed fraudsters whose interactions with regular users can be characterized as RLAs.

We want to:

(iii) Model the evolution of the local clustering properties of regular users; and

(iv) Find out which individuals are performing RLAs.

Condition (i) requires that the structure of interaction between users follows some well-defined patterns. In particular, we assumed that users can be divided into two communities and the average local clustering distribution reaches a high stationary value. Condition (ii) requires that we know the origin of some fraudulent transactions (e.g., from historical data). Finally, the expected pattern of interaction between fraudsters and regular users must follow a uniform random process. Next, we focus on developing a dynamic model that we use to characterize the expected value of the local clustering of regular nodes as a function of their current degree.

III. A NETWORK MODEL

Let $H(t) = \{1, \ldots, n(t)\}$ be a finite set of interconnected nodes (users) at time index t. The adjacency matrix $A(t)$, with entries $a_{ij}(t) \in \{0, 1\}$, represents whether there has been a transaction (an exchange of information or services) between two users. In particular, $\forall i, j \in H(t)$, $i \neq j$, $a_{ij} = a_{ji} = 1$ (i.e., there exists an undirected link between the two nodes) if nodes i and j have engaged in a transaction up to time t. Let $G(t) = (H(t), A(t))$ be the network at time index t and $Q_i(t) = \{j \in H(t) : a_{ij} = 1\}$ represent all nodes connected to node i (i.e., the set of neighbors of node i). For node $i \in H(t)$, let $k_i(t) = |Q_i(t)|$ represent the degree of node i.

Under the premise that the network of regular users is naturally divided into two communities, it can be viewed as having three types of nodes. Nodes of type 1 or 2 represent regular users. Nodes of type 0 represent fraudsters. For node i, the variable $\delta_i \in \{0, 1, 2\}$ specifies its type. Let $R(t) = \{i \in H(t) : \delta_i \in \{1, 2\}\}$ and $F(t) = \{i \in H(t) : \delta_i = 0\}$ be the set or regular users and fraudsters, respectively. The topology of the network evolves based on a decision-making mechanism that involves two kinds of processes. The first represents the addition of new users and takes place the instant a new node is added to the network (called node attachment); the second represents the occurrence of new transactions and takes place at asynchronous instants of time thereafter (called node interaction).

A. Node attachment

Suppose that every time index t a newly added node attaches to m different nodes. The likelihood that this new user, node $j \notin H(t-1)$, is a fraudster is p_f (i.e., $\delta_j = 0$ with probability p_f). If $\delta_j \neq 0$, i.e., if node j is a regular user, there is a strong preference to attach to a node of the same type (i.e., nodes with similar characteristics are more likely to establish connections between them) [13], [14]. In particular, node j connects to node $j' \in H(t-1)$ of the same type with probability p_r (and with probability $1 - p_r$ to a regular node with $\delta_{j'} \neq \delta_j$). Note that there are no self-loops (i.e., $\forall i \in H(t)$, $a_{ii}(t) = 0$ for all $t \geq 0$).

Fraudsters do not differentiate between the two types of regular users. They connect randomly to m nodes of type 1 or 2, and to m_f nodes of type 0. Since fraudsters perform RLAs, each regular user has an equal probability to become the victim of an attack, independently of which users have been targeted in past instances.

B. Node interaction

After attaching to the network, regular nodes try to form close-knit groups (triad junctions) based on conditions similar to [11], [12]. In particular, after node $j \notin H(t-1)$ attaches to node $j' \in H(t-1)$ (one of the total of m nodes that node j links to during node attachment), it may establish additional links to v regular neighbors of node j'. After attaching to the networks, node j establishes these additional links (forms triads) at asynchronous instants of time. Note that if the set of neighbors of node j' is a subset of the set of neighbors of node j, then there is no possibility of forming triads. In general, if $j \in Q_{j'}(t)$ and $j' \in Q_i(t)$ for some node i, then node j establishes an additional link to node i with probability $x_i(t)$. The value of $x_i(t)$ is influenced by δ_j and δ_i. In particular, a multivariate random variable X_t^δ with a positive expected probability $p_t^\delta = E[X_t^\delta] = f(\sigma_1, \ldots, \sigma_s)d\sigma_1 \ldots d\sigma_s$ captures the probabilities of establishing a link between nodes j and i, where $\sigma_1, \ldots, \sigma_s$ are independent factors that influence the affinity between the various types of nodes. The process of triad formation repeats for every link established during node attachment (m times).

As in [15], [16], here the probability of establishing additional links due to triad formation is given by

$$x_i(t) = \begin{cases} p_\Delta - \frac{\omega}{uk_i}, & \text{if } \delta_j = \delta_i \\ (1-p_\Delta) - \frac{\omega}{uk_i}, & \text{if } \delta_j \neq \delta_i \end{cases} \tag{1}$$

where u captures the compatibility between regular nodes (chosen from a uniformly random distribution with support on $[0,1]$). The parameter ω, $0 < \omega < u$, represents the cost of establishing additional links. The random variable X_t^δ takes values $x_i(t)$ and the expected value of X_t^δ at time t is given by

$$p_t^\delta = \begin{cases} \int_0^\infty \int_0^1 \left(p_\Delta - \frac{\omega}{uk_i(t)}\right) p_u p_k \; du \; dk_i, & \text{if } \delta_j = \delta_i \\ \int_0^\infty \int_0^1 \left((1-p_\Delta) - \frac{\omega}{uk_i(t)}\right) p_u p_k \; du \; dk_i, & \text{if } \delta_j \neq \delta_i \end{cases} \tag{2}$$

where p_u equals $\frac{1}{u}$, and p_k is the probability distribution of $k_i(t)$. Note that if $\delta_j = \delta_i$ the process of triad formation has a stationary mean p_Δ (because eq. (2) converges to p_Δ). Otherwise, if $\delta_j \neq \delta_i$ it has stationary mean $1-p_\Delta$. Let $X^\delta = \{X_t^\delta\}$ with stationary mean $0 \leq p_\Delta \leq 1$ be the random process associated to the formation of close-knit groups.

Unlike regular nodes, fraudsters do not tend to establish edges to nodes of the same type; instead they persistently perform RLAs. After attaching to the network, each fraudster chooses to attack, with probability p'_f, a total of m'_f regular nodes selected randomly at every time t (among nodes that have not been one of its past victims, i.e., are not part of its set of neighbors).

To ensure that the dynamics of the model is well-defined, the following assumptions are needed:

(A1) The network $G(0)$ is connected.

(A2) The network $G(0)$ has at least m nodes, each with at least v neighbors.

Assumption (A1) is satisfied when there exists a path between any pair of nodes. Assumption (A2) requires that the initial network has $n(0) \geq m$, and for every node $i \in H(0)$, $k_i(0) > v$ (which is required when $p_\Delta = 1$).

To characterize the evolution of the clustering properties of the network, we restrict our analysis to regular users. The following theorem describes their asymptotic behavior (due to space constraints, the complete proof is found in the online supplement for this paper, available at: homes.soic.indiana.edu/pmoriano/publications).

Theorem 1 *(local clustering coefficient): For all networks G(0) that satisfy* (A1)-(A2)*, the asymptotic behavior of the local clustering coefficient for a regular node with degree* k_i *satisfies*

$$c_i = \frac{2\left(k_i + pm + \varepsilon \ln\left(\frac{k_i+\varepsilon}{n+\varepsilon}\right)(p-1)\right)}{(k_i + p\varepsilon)(k_i + p\varepsilon - 1)} \tag{3}$$

with $p = \frac{v(p_r + p_\Delta - 2p_r p_\Delta + (p_\Delta - 1)p_r v - 1)}{(2p_r - 1)p_\Delta v - p_r v - 1}$ and $\varepsilon = \frac{2m(1+p)}{p}$.

Theorem 1 implies that the value of c_i does neither depend on the initial network $G(0)$ nor the size of $G(t)$ (note that the coefficient does not vanish as $n(t) \to \infty$). Theorem 1 allows us to estimate the clustering coefficient in the neighborhood of regular nodes. We will use this convergence result to design conditions that are useful to identify random attacks.

IV. Detecting fraudsters

To detect fraudsters (i.e., nodes of type 0) the algorithm follows a two-step procedure. The first step identifies suspects (i.e., potential fraudsters) based on deviations from the location of nodes in the spectral space (following similar ideas as in [10]). The second step uses the expression given by eq. (3) to locate fraudsters within the set of suspects.

A. Spectral analysis

First, note that at time index t, there are $|H(t)| = n(0) + t$ nodes, of which the expected number $E[|R(t)|] = (1 - p_f)t + R(0)$ are expected to be regular users and $E[|F(t)|] = p_f t + F(0)$ are expected to be fraudsters. At this time t, fraudsters have, on average, engaged in $(mp_f + m'_f p'_f)t$ attacks. Second, define a time window $t_w > 0$ in which we measure the changes in the spectra of the adjacency matrix. Let the graph $\hat{G}(t) = (H(t), \hat{A}(t))$ with adjacency matrix $\hat{A}(t) = A(t - t_w)$ represent the network $G(t)$ without taking into account any link that was established after time $t - t_w$. We assume that perturbations are bounded by a small constant $\phi > 0$ (i.e., $\|A(t) - \hat{A}(t)\| < \phi$ for any $t \geq 0$). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_{n(t)}$ be the eigenvalues of the adjacency matrix $\hat{A}(t)$ and

$$\hat{\boldsymbol{z}}(t) = \begin{pmatrix} \hat{z}_{11}(t) & \cdots & \hat{z}_{j1}(t) & \cdots & \hat{z}_{n(t)1}(t) \\ \vdots & & \vdots & & \vdots \\ \hat{z}_{1i}(t) & \cdots & \hat{z}_{ji}(t) & \cdots & \hat{z}_{n(t)i}(t) \\ \vdots & & \vdots & & \vdots \\ \hat{z}_{1n(t)}(t) & \cdots & \hat{z}_{jn(t)}(t) & \cdots & \hat{z}_{n(t)n(t)}(t) \end{pmatrix}$$

be the matrix of eigenvectors. The eigenvector $\hat{\boldsymbol{z}}_j(t)$ is represented as a column vector and $\hat{z}_{jr}(t)$ denotes the r^{th} entry of $\hat{\boldsymbol{z}}_j(t)$. The row vector $(\hat{z}_{1i}(t), \hat{z}_{2i}(t), \cdots, \hat{z}_{n(t)i}(t))$ represents the spectral coordinates of node i. For the network $G(t)$, we use the leading $q \ll n(t)$ eigenpairs to form the set of suspects using the following procedure.

B. Identifying suspects in the spectral space

Every time window t_w, the algorithm estimates the spectra of $A(t)$, based on perturbations $\Delta A(t) = A(t) - \hat{A}(t) = A(t) - A(t - t_w)$. Using a first order approximation allows us to update the eigenpairs $\lambda_i(t)$ and $\boldsymbol{z}_i(t)$ of the matrix $A(t)$, without the need to recalculate the entire spectra (provided perturbations are sufficiently small, i.e., bounded by the constant ϕ) [17]. This approximation yields

$$\begin{aligned} \lambda_i(t) &= \hat{\lambda}_i(t) + \hat{\boldsymbol{z}}_i^\top(t)[\Delta A(t)]\hat{\boldsymbol{z}}_i(t) \\ \boldsymbol{z}_i(t) &= \hat{\boldsymbol{z}}_i(t) + \sum_{j \neq i} \frac{\hat{\boldsymbol{z}}_i^\top(t)[\Delta A(t)]\hat{\boldsymbol{z}}_i(t)}{\hat{\lambda}_i(t) - \hat{\lambda}_j(t)} \hat{\boldsymbol{z}}_i(t) \end{aligned}$$

Furthermore, the non-randomness of node i at time t is defined as

$$r_i(t) = \sum_{j=1}^{q} \lambda_j(t) z_{ji}^2(t) \tag{4}$$

where q is the value that maximizes $\hat{\lambda}_q - \hat{\lambda}_{q+1}$. The expected value and variance of the non-randomness of node i yields

$$\begin{aligned} E[r_i(t)] &= k_i^2(t) \sum_{j=1}^{q} \frac{E[\boldsymbol{z}_j^2(t)]}{\lambda_j(t)} \\ &+ \frac{k_i(t)}{n(t)} \left(1 - \frac{k_i(t)}{n(t)}\right) \sum_{j=1}^{q} \frac{1}{\lambda_j(t)} \end{aligned} \tag{5}$$

$$\begin{aligned} Var[r_i(t)] &= \frac{4k_i^3(t)}{n(t)} \left(1 - \frac{k_i(t)}{n(t)}\right) \sum_{j=1}^{q} \frac{E[\boldsymbol{z}_j^2(t)]}{\lambda_j^2(t)} \\ &+ \frac{2k_i^2(t)}{n^2(t)} \left(1 - \frac{k_i(t)}{n(t)}\right)^2 \sum_{j=1}^{q} \frac{1}{\lambda_j^2(t)} \end{aligned} \tag{6}$$

where $E[\boldsymbol{z}_j(t)]$ denotes the mean of $\boldsymbol{z}_j(t)$.

Finally, given a probability $\rho \in [0, 1]$, let $\eta > 0$ denote the $\frac{1+\rho}{2}$ quantile of the standard normal distribution (i.e., the quantile denotes the interval $[-\eta, \eta]$ that covers the probability ρ). For a fixed time t, node i belongs to the suspect set if

$$\text{(C1)} \qquad r_i(t) \leq E[r_i(t)] + \eta Var[r_i(t)]^{1/2} \tag{7}$$

C. Filtering suspects based on clustering

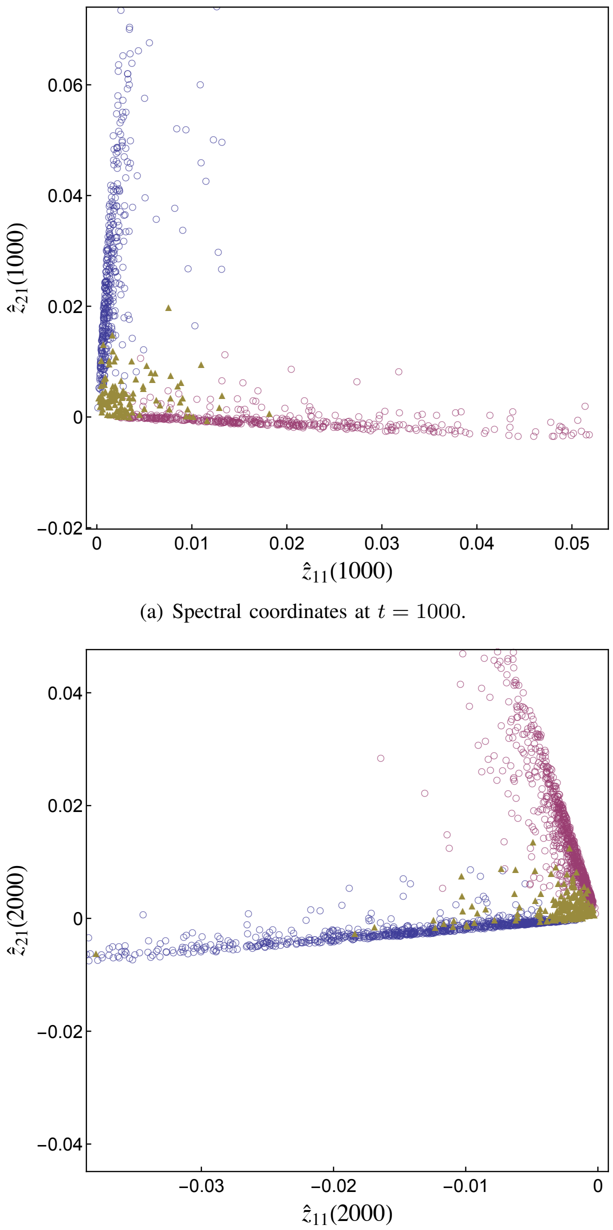
The second part of the algorithm filters the proposed set of suspects based on the average local clustering coefficient of a node. In particular, node i with local clustering coefficient $\bar{c}_i$ belongs to the set of fraudsters if

$$\text{(C2)} \qquad \left| \frac{\bar{c}_i - c_i}{\bar{c}_i} \right| > \theta \tag{8}$$

where $0 < \theta < 1$ is a sensibility parameter that quantifies the allowable deviations from eq. (3). Note that $\theta = 0$ means that any deviation from the theoretical local clustering coefficient mark a suspect as a fraudster. On the other hand, $\theta = 1$ means that a suspect will not be marked as a fraudster, even if its actual local clustering is far from the expected value. Because the condition is based on a local property, fraudsters who want to avoid detection by forging strongly clustered neighborhoods need to estimate the dynamic expression underlying close-knit group formation, which is more difficult because it evolves as a function of node degree. Note that in general, the parameters η, θ, and t_w need to be tuned based on the properties of exposed fraudsters in the historical data.

V. SIMULATIONS

To gain better insight into the conditions for detection, consider a network with $n(0) = 30$, $m = 2$, $m_f = 2$, $m'_f = 3$, $v = 3$, $p_r = 1.0$, $p_\triangle = 1.0$, $p_f = 0.1$, and $p'_f = 0.1$. Let the detection parameters be $\eta = 3$, $\theta = 0.01$, and $t_w = 200$. Figure 1(a) illustrates the node spectral coordinates at $t = 1000$. Regular users are represented by circles and fraudsters by triangles. They are divided into two communities (represented by the different colors). Most users are distributed along two quasi-orthogonal lines in the spectral space. Figure 1(b) shows the same plot for the network at $t = 2000$. Note that as the network grows it becomes harder to identify fraudsters based solely on spectral analysis.

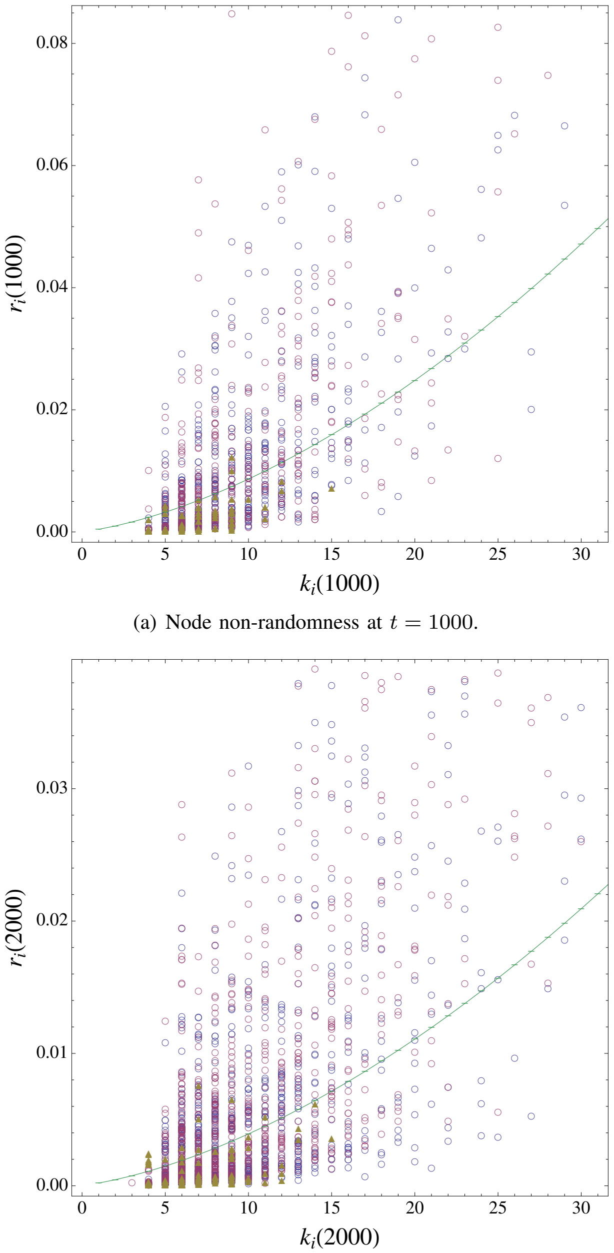


(a) Spectral coordinates at $t = 1000$.

(b) Spectral coordinates at $t = 2000$.

Fig. 1: Spectral coordinates for network data under collaborative attacks at (a) $t = 1000$; and (b) $t = 2000$.

Next, fig. 2(a) shows the value of the node non-randomness as a function of the degree. The solid line is the detection threshold represented by to the upper bound in eq. (7). The plot shows that it is hard to accurately identify fraudsters solely based on node non-randomness. Figure 2(b) illustrates the same plot for the network at $t = 2000$. As the network evolves, it actually becomes even more dificult to identify fraudsters.

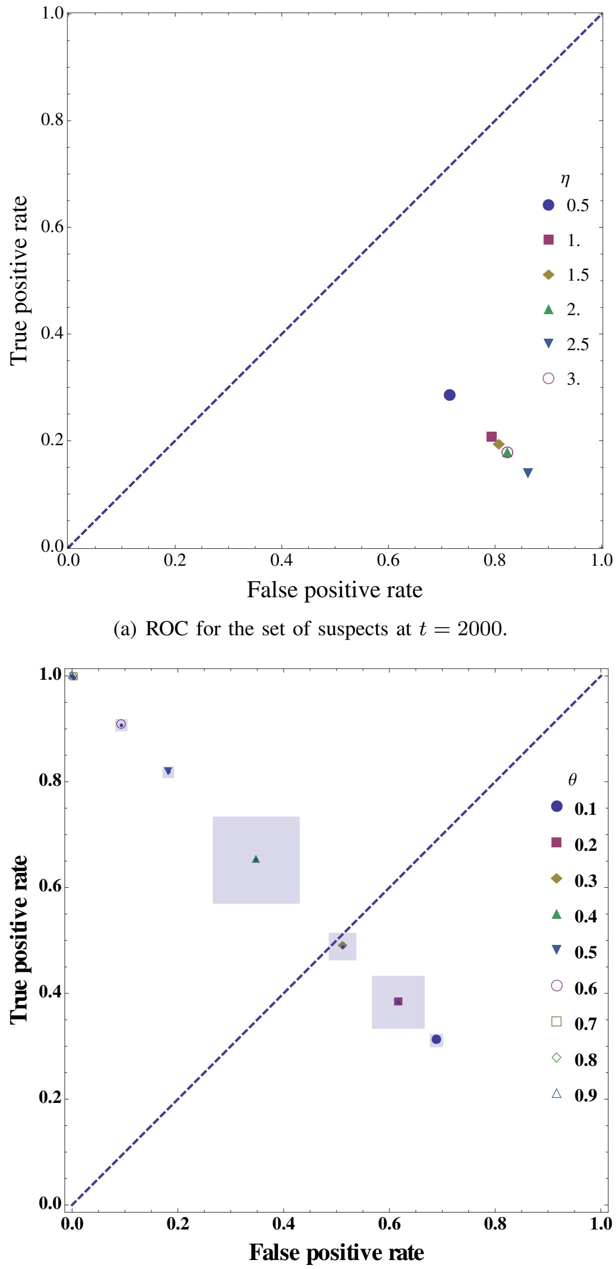


(a) Node non-randomness at $t = 1000$.

(b) Node non-randomness at $t = 2000$.

Fig. 2: Node non-randomness for the network at (a) $t = 1000$; and (b) $t = 2000$.

Figure 3(a) illustrates the corresponding ROC plot for suspects for different values of η (i.e., the average rate of true positives and false negatives over 10 different networks of size $t = 2000$). It shows that the first step of the algorithm has high rates of false positives, which means that fraudsters are able to hide among regular users. The value $\eta = 0.5$ minimizes the rate of false positives. Figure 3(b) illustrates the ROC plot for the identified fraudsters for different values of θ for $\eta = 0.5$. The performance of the algorithm improves as the value of θ increases, which verifies that for the clustering properties allow us to accurately identify fraudsters in the second step of the algorithm.



(a) ROC for the set of suspects at $t = 2000$.

(b) ROC for the set of fraudsters at $t = 2000$. Shaded regions represent standard deviations.

Fig. 3: ROC plot for (a) suspects; and (b) fraudsters.

Finally, Table I shows the performance of the algorithm presented in [9] using the network substrate generated by our model. Note that the percentage of false negatives is about twice that of false positives. Compared to Table II, note that the percentage of false positives based on (C1) and (C2) is similar to the results in [9]. However, the percentage of false negatives vanishes as the network grows.

TABLE I: Results for the algorithm in [9]. The variable μ represents the rounded mean and σ the standard deviation of 10 simulation runs.

t	Actual fraudsters		Potential fraudsters		Detected fraudsters		% of false positives		% of false negatives	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
200	21	2.8	115	5.9	49	17.1	22	7.2	81	34.9
400	40	3.0	219	10.3	97	38.2	24	8.0	92	16.6
600	60	7.3	315	9.9	195	74.6	30	9.6	83	17.7
800	81	10.0	427	16.7	222	106.8	27	11.1	81	21.3
1000	104	10.0	551	19.7	312	83.4	31	7.3	87	12.1
1200	122	14.0	641	26.7	360	98.4	30	7.0	87	11.9
1400	145	12.1	734	25.4	461	110.0	33	7.0	86	16.8
1600	159	15.2	883	31.3	503	140.7	32	8.0	87	10.1
1800	178	9.7	994	20.8	537	160.2	31	8.6	89	11.43
2000	197	15.6	1107	23.5	752	224.8	38	10.1	82	12.2

TABLE II: Results for the proposed algorithm. The variable μ represents the rounded mean and σ the standard deviation of 10 simulation runs.

t	Actual fraudsters		Potential fraudsters		Detected fraudsters		% of false positives		% of false negatives	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
200	21	3.2	113	9.0	73	9.5	31	4.2	63	11.4
400	39	7.3	211	9.0	158	8.4	34	2.3	34	8.3
600	60	9.6	323	10.6	238	13.1	34	2.9	39	5.4
800	83	9.0	427	15.8	326	13.8	35	1.5	25	6.3
1000	99	10.7	537	19.6	412	19.2	35	1.9	22	3.7
1200	126	8.5	653	16.6	508	21.4	36	1.7	20	3.5
1400	141	14.4	756	28.8	576	31.1	35	2.5	20	3.5
1600	158	15.3	874	24.4	664	33.9	36	2.4	21	3.3
1800	179	8.9	999	27.8	785	32.8	38	2.2	19	2.8
2000	202	10.1	1107	31.7	857	30.2	37	1.6	17	1.7

VI. CONCLUSIONS

This paper introduces a detection algorithm that uses both spectrum-based and direct topological measures to differentiate between the aggregate dynamics of fraudsters and regular users. The generic notion of random link attacks characterizes the behavior of fraudsters, while regular users tend to form close-knit groups (highly clustered neighborhoods). Taking into account the evolution of community structures and clustering properties in the design of the algorithm, allows us to identify fraudsters with a negligible percentage of false negatives. The approach offers a novel perspective on how fraud detection could exploit structural properties of dynamic networks, and should be of interest to analysts trying to detect structural anomalies in networks representing a wide class of information and service exchanges. Extending the detection procedure to evaluate networks with directed links provides an important direction for future research.

REFERENCES

[1] J. Charles, "AI and law enforcement," *IEEE Intelligent Systems*, vol. 13, no. 1, pp. 77–80, 1998.

[2] D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting fraudulent personalities in networks of online auctioneers," in *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (PKDD)*, pp. 103–114, Berlin, Germany, September 2006.

[3] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.

[4] S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning," *Information Fusion*, vol. 10, no. 4, pp. 354 – 363, 2009.

[5] P. O. Boykin and V. P. Roychowdhury, "Leveraging social networks to fight spam," *Computer*, vol. 38, no. 4, pp. 61–68, 2005.

[6] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, Washington, D.C., August 2003.

[7] N. Shrivastava, A. Majumder, and R. Rastogi, "Mining (social) network graphs to detect random link attacks," in *Proceedings of the IEEE International Conference on Data Engineering*, pp. 486–495, Cancun, Mexico, April 2008.

[8] K. Leung and C. Leckie, "Unsupervised anomaly detection in network intrusion detection using clusters," in *Proceedings of the Twenty-eighth Australasian Conference on Computer Science*, vol. 38, pp. 333–342, 2005.

[9] X. Ying, X. Wu, and D. Barbará, "Spectrum based fraud detection in social networks," in *Proceedings of the IEEE 27th International Conference on Data Engineering*, pp. 912–923, Hannover, Germany, April 2011.

[10] X. Ying, L. Wu, and X. Wu, "A spectrum-based framework for quantifying randomness of social networks," *IEEE Trans. on Knowl. and Data Eng.*, vol. 23, no. 12, pp. 1842–1856, 2011.

[11] P. Moriano and J. Finke, "On the formation of structure in growing networks," *J. Stat. Mech. Theor. Exp.*, vol. 6, p. P06010, 2013.

[12] P. Holme and B. J. Kim, "Growing scale-free networks with tunable clustering," *Phys. Rev. E*, vol. 65, no. 2, p. 026107, 2002.

[13] P. Moriano and J. Finke, "Structure of growing networks with no preferential attachment," in *Proceedings of the American Control Conference*, pp. 1090–1095, Washington, DC, June 2013.

[14] P. Moriano and J. Finke, "Characterizing the relationship between degree distributions and community structures," in *Proceedings of the American Control Conference*, pp. 2383–2388, Portland, OR, 2014.

[15] M. O. Jackson and B. W. Rogers, "Meeting strangers and friends of friends: How random are social networks?," *Am. Econ. Rev.*, vol. 97, no. 3, pp. 890–915, 2007.

[16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw Hill Higher Education, 4th ed., 2002.

[17] Y. Shmueli, G. Wolf, and A. Averbuch, "Updating kernel methods in spectral decomposition by affinity perturbations," *Linear Algebra Appl.*, vol. 437, no. 6, pp. 1356–1365, 2012.