Appeared in "Workshop on Information in Networks", 2011

Stealing Reality: When Criminals Become Data Scientists

Yaniv Altshuler,[1,2] Nadav Aharony,[1] Yuval Elovici,[2,3] Alex Pentland,[1] and Manuel Cebrian[1,4]

[1]*The Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2]*Deutsche Telekom Laboratories, Ben Gurion University, Beer Sheva 84105, Israel*
[3]*Department of Information Systems Engineering, Ben Gurion University, Beer Sheva 84105, Israel*
[4]*Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093*

I. EXTENDED ABSTRACT

We live in the age of social computing. Social networks are everywhere, exponentially increasing in volume, and changing everything about our lives, the way we do business, and how we understand ourselves and the world around us. The challenges and opportunities residing in the social oriented ecosystem have overtaken the scientific, financial, and popular discourse. With the growing emphasis on personalization, personal recommendation systems, and social networking, there is a growing interest in understanding personal and social behavior patterns. This trend is manifested in the growing demand for "*data scientists*" and data-mining experts in the commercial ecosystem, which in turn is derived from the increasing number of social data-driven start-up companies as well the social inference related research sponsored by other commercial entities and various NGOs.

This work is somewhat of a 'what if' exploration: History has shown that whenever something has a tangible value associated with it, there will always be those who will try to steal it for profit. Along this line of thought — based on these current trends of the data ecosystem coupled with the emergence of advanced tools for social and behavioral pattern detection and inference — we ask the following : *What will happen when the criminals become data scientists?*

We conjecture that the world will increasingly see malware integrating tools and mechanisms from network science into its arsenal, as well as attacks that directly target human-network information as a goal rather than a means. Paraphrasing Marshall McLuhan's "*the medium is the message*" we have reached the stage where "*the network is the message*".

Specifically, we point out a new type of information security threat — a class of malware, the goal of which is not to corrupt the machines it infects, take control of them, or steal explicit information stored on them (e.g. credit card information and personal records). Rather, the goal of this type of attack is to steal social network and behavioral information through data collection and network science inference techniques. We call this type of attack a "*Stealing Reality*" attack.

After characterizing the properties of this new kind of attack, we analyze the ways it could be carried out — we show the optimal strategy for attackers interested in learning a social network and its hidden underlying social principles. Remarkably, our analysis shows that such an optimal strategy should follow in many cases an extremely slow spreading pattern. Counterintuitively, such attacks generate far greater damage in the long term compared to more aggressively spreading attacks. In addition, such attacks are likely to avoid detection by many of today's network security mechanisms, which tend to focus on detecting network traffic anomalies such as traffic volume increase. We demonstrate this surprising new discovery using several real world social networks datasets.

A preliminary version of this paper was presented in last year's WIN event. In the past year we have revalidated our model using extended mobile networks datasets, as well as developed a novel network measure for the assessment of social information that is encapsulated in a (social) network. The paper was accepted to publication in IEEE Intelligent Systems.

We shall model the social network as an undirected graph $G(V,E)$. A *Stealing Reality* attacker's first goal is to inject a single malware agent into one of the network's nodes. Upon such injection, the agent starts to 'learn' this node (and its interactions with its neighbors). Periodically, the agent tries to copy itself into one of the original node's neighbors. The probability that an agent tries to copy itself to a neighboring node at any given time step is called the "*aggressiveness*" of the attack, and is denoted as ρ. Namely, aggressive agents have higher value of ρ (and hence take shorter periods of time between each two spreading attempts), whereas less aggressive agents are less likely to try and spread at any given time, and will then wait on average longer between trying to copy themselves to one of the neighbors of their current host.

As the information about the network itself has become worthy cause for an attack, the attacker's motivation is stealing as much properties related to the network's social topology as possible. We shall denote the percentage of vertices-related information acquired at time t by $\Lambda_V(t)$ and the percentage of edges-related information acquired at time t is by $\Lambda_E(t)$.

The duration of the learning process of the *Stealing Reality* attack refers to the time it takes the attacking agent to identify with high probability the properties of a node's behaviors, or of some of its social interactions. We model this process using a standard *Gompertz function* in the parametric form of $y(t) = ae^{be^{ct}}$ (for some parameters a, b and c). This model is flexible enough to fit various social learning mechanisms, while providing the following important features : (a) Sigmoidal advancement, namely — the longer such an gent operates, the more precise its conclusions will be. (b) The rate at which information is gathered is smallest at the start and end of the learning process. (c) Asymmetry of the asymptotes, implied from the fact that for any value of T, the amount of information gathered in the first T time steps is greater than the amount of information gathered at the last T time steps.

An aggressive spreading pattern is more likely to be detected by users or administrators, resulting in the subsequent blocking of the attack. On the other hand, attacks that spread slowly may evade detection for a longer period of time, however, the amount of data they gather would be limited. In order to predict the detection probability of the attack at time t we shall use *Richard's Curve* — a generalized logistic function often used for modeling the detection of security attacks [1] :

$$p_{detect}(t) = \frac{1}{\left(1 + e^{-\rho(t-M)}\right)^{\frac{1}{\rho}\sigma}} \tag{1}$$

where ρ — the *attack aggressiveness*, σ is a normalizing constant for the detection mechanism, and M denotes the normalizing constant for the system's initial state.

We shall now define a mathematical measure that predicts the ability of an attacker to "steal", or acquire, a given social network, we call the "*sociallearnability*" of a network. The measure reflects both the information contained in the network itself, as well as the broader context from which the network was derived. Once presenting the mathematical formulation of this measure, we demonstrate its importance by showing how it can sort several real world social networks according to their complexity (which is known), and even group two very different social networks that were generated by the same group of people. We conclude by showing that the optimal learning process with respect to this new measure involves in many cases extremely non-aggressive attacks.

Let us denote by K_E the *Kolmogorov Complexity* [2] of the network, namely — the minimal number of bits required in order to "code" the network in such a way that it could later be completely restored. The Kolmogorov Complexity of a network represents in fact the basic amount of information contained in a social network. For example, a military organization's network has very homogeneous links and hierarchical structures repeated many times over. We would expect it to require a much shorter minimal description than, say, the social network of the residents of a metropolitan suburb. In the latter, we would expect to see a highly heterogenous network, composed of many types of relationships (such as work-relationships, physical proximity, family ties, and other intricate types of social relationships and group affiliations).

At this point, let us recall that every social reality network belongs to (one or more) "*social family*", each of which having its own consistency (or versatility). Some families may contain a great variety of possible networks, each having roughly a similar probability to occur, while another may consist of a very limited number of possible networks.

Notice that the complexity of each network does not necessarily correlates with its entropy. There may exist families of low variety of highly complicated networks, while other families may contain a great variety of relatively simple networks.

Let us define $\mathcal{G}_n$ to contain n random instances of networks of $|V|$ nodes that belong to the same *social family* as G. Let X_n be a discrete random variable with possibility values $\{x_1, x_2, \ldots, x_{2^{\frac{1}{2}|V|(|V|-1)}}\}$ (corresponding to all possible graphs over $|V|$ nodes), taken according to the distribution of $\mathcal{G}_n$. The normalized social entropy of the network G would therefore be calculated by dividing the entropy of the variable X_n by the maximal entropy for graphs of $|V|$ nodes :

$$\lambda_n(G) \triangleq \frac{H(X_n)}{\log_2 \zeta_{|V|}} \tag{2}$$

where $\zeta_{|V|}$ denotes the number of distinct non-isomorphic simple graphs of $|V|$ nodes.

$\lambda(G)$ is then defined as : $\lim_{n\to\infty} \lambda_n(G)$.

At this point let us recall *Reed's Law* which asserts that the utility of large networks (and particularly social networks), can scale exponentially with the size of the network. This observation is derived from the fact that the number of possible sub-groups of network participants is exponential in N (where N is the number of participants), stretching far beyond the N^2 utilization of *Metcalfe's Law* (that was used to represent the value of telecommunication networks).

Extending this notion we assert that a strong value emerges from learning the $2^{\mathcal{I}}$ "social principles" behind a network, denoting by $\mathcal{I}$ the *information* that is encapsulated in a network.

Assuming that at time t an attacker has stolen $|E|\Lambda_E(t)$ edges, then taking K_E as the maximal amount of information that can be coded in the network G, we normalize it by the fraction of edges acquired thus far. As K_E is measured in bits, the appropriate normalization should maintain this scale. Multiplies by $\lambda(G)$, the *normalized social entropy* of the network G, the network information can be written as follows :

$$\mathcal{I} = \lambda(G) \cdot K_E \cdot \frac{\log_2\left(|E|\Lambda_E(t)\right)}{\log_2 |E|}$$

After normalizing by the overall "social essence" of the network (received for $\Lambda_E = 1$) the following measurement for the social essence of the sub-networked acquired is achieved :

$$\Lambda_S(t) = \frac{2^{\lambda(G)\cdot K_E \cdot \frac{\log_2(|E|\Lambda_E(t))}{\log_2 |E|}}}{2^{\lambda(G)\cdot K_E}} = 2^{\lambda(G)\cdot K_E \cdot \frac{\log_2 \Lambda_E(t)}{\log_2 |E|}}$$

which after some arithmetics yields :

$$\Lambda_S(t) = \Lambda_E(t)^{\frac{\lambda(G)\cdot K_E}{\log_2 |E|}} \tag{3}$$

Note that K_E represents the *network* complexity, whereas $\lambda(G)$ represents the complexity of its *social family*.

At this point we assert that our *socialearnability measure* presented above is indeed a valuable property for measuring network attacks. For this, we demonstrate the values of this measure for several different real world networks. Figure 1 presents an analysis of the networks derived from the *Social Evolution* experiment [3], the *Reality Mining* network [4], and the *Friends and Family* [5] experiment. One can easily see the logic behind the predictions received using the *socialearnability measure* concerning the difficulty of learning each of the networks. Specifically, the *Social Evolution* network is predicted to be harder to steal compared to the *Reality Mining* network, however easier to steal compared to the networks of *Friends and Family*. This can be explained when looking closely at the details of the three experiments. Whereas the *Reality Mining* experiment tracked people within a relatively static work environment, the *Social Evolution* experiment took place at an MIT Undergraduate dorms, involving students with (apparently) much more complicated mobility and interactions patterns. The *Friends and Family* dataset involved even more complicated interactions as it includes a heterogeneous community of couples, increasing the amount of information encapsulated within the network.

In addition, notice how the *socialearnability measure* places the two *Friends and Family* networks directly on top of each other, despite the fact that the two networks contain significantly different information (of volume, meaning and network information). Still, as the two networks essentially represent the same social group of people, their *socialearnability measure* has a very similar value.

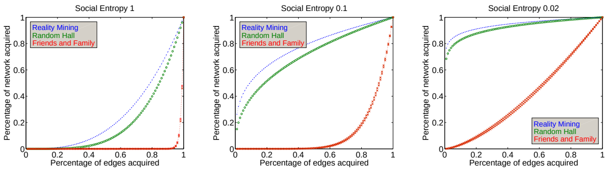

FIG. 1. An illustration of the reality stealing process for three different values of social entropy $\lambda(G)$ (0.02, 0.1, and 1), for four different networks — the *Random Hall* network [3], *Reality Mining* networks [4], *Friends and Family* [5] self-reporting network and *Friends and Family* Blue-Tooth network [5]. Using this example we can see that the *Reality Mining* network is easier to steal than the Random Hall network, which in turn is easier to steal compared to the Friends and Family networks.

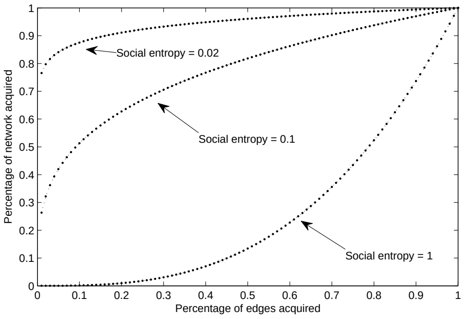

FIG. 2. A demonstration of the importance of a network's social entropy $\lambda(G)$, illustrated for the *Reality Mining* network [4]. It can be seen that if we assume that the network is derived from a family of the maximal entropy (namely, having a uniform distribution of all possible networks) the evolution of the *Stealing Reality* attack differs significantly than for networks that were derived from a family of a lower social entropy. In fact, even for $\lambda(G) = 0.1$ stealing the network would be materially easier, having additional information out of any edge acquired.

The importance of the social entropy of a network is demonstrated in Figure 2, analyzing the *Reality Mining* network [4] for various possible values of social entropy. The value for the Kolmogorov Complexity of the network was approximated using an *LZW* compression of the network.

We evaluate our model on data derived from a real-world cluster of mobile phone users drawn from the call records of a major city within a developed western country, comprised of approximately $200,000$ nodes and $800,000$ edges.

Figure 3 demonstrates the attack efficiency (namely, the maximal amount of network information acquired) as a function of its "aggressiveness" (i.e. the attack's infection rate). The two curves represent the overall amount of information (edges related and vertices related) that can be obtained as a function of the aggressiveness value ρ. It can be seen that although a local optimum exists for an aggressiveness value of little less than $\rho = 0.5$ (namely, a relatively aggressive attack), it is preceded by a global optimum achieved by a much more "subtle" attack, for an aggressiveness value of $\rho = 0.04$.

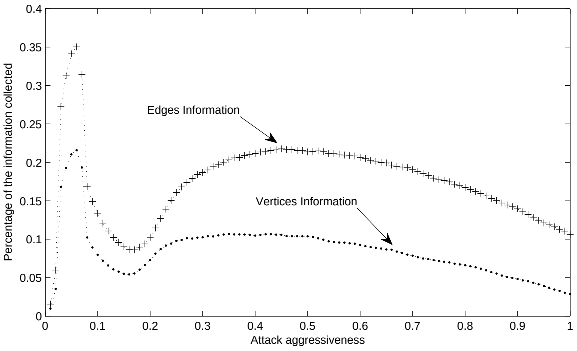

FIG. 3. An analytic study of the overall amount of data that can be captured by a *Stealing Reality* attack, illustrating the phenomenon where the most successful attack possible (namely, an attack that is capable of stealing the maximal amount of information) is produced by a very low value of the *attack aggressiveness* ρ. The upper curve represents $\Lambda_E(\rho)$, the overall percentage of edges related information stolen. The lower curve represents $\Lambda_V(\rho)$, the overall percentage of vertices related information stolen. Notice the local maximum around $\rho = 0.5$ that is outperformed by the global maximum at $\rho = 0.04$.

[1] N. A. Christakis and J. H. Fowler, PLoS ONE, **5**, e12948 (2010).
[2] A. Kolmogorov, Problems Information Transmission, **1**, 1 (1965).
[3] A. Madan, M. Cebrian, D. Lazer, and A. Pentland, in *Proceedings of the 12th ACM international conference on Ubiquitous computing* (New York, NY, USA).
[4] N. Eagle, A. Pentland, and D. Lazer, Proceedings of the National Academy of Sciences (PNAS), **106**, 15274 (2009).
[5] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland, in *Proceedings of the 13th ACM international conference on Ubiquitous computing (to appear)*, Ubicomp '11 (ACM, 2011).