

Stealing Reality: When Criminals Become Data Scientists (or Vice Versa)

Yaniv Altshuler, Nadav Aharony, and Alex Pentland, *Massachusetts Institute of Technology*

Yuval Elovici, *Ben Gurion University of the Negev, Israel*

Manuel Cebrian, *University of California, San Diego*

Stealing-reality attacks attempt to steal social network and behavioral information through data collection and inference techniques, making them more dangerous than other types of identity theft.

We live in the age of social computing. Social networks are everywhere, exponentially increasing in volume, and changing how we do business and how we understand ourselves and the world around us. The challenges and opportunities inherent in the social-oriented ecosystem have overtaken

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 increased demand for data scientists and data-mining experts, which derives from the increasing number of social data-driven start-up companies as well as the social-inference-related research sponsored by commercial entities and various nongovernmental organizations (NGOs).

This article explores a “what if” scenario. That is, history has shown that whenever something exhibits a tangible value, someone will try to steal it for profit. Along this line of thought—based on current trends in 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 criminals become data scientists?

We conjecture that the world will increasingly see malware-integrating tools and mechanisms from network science 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,”¹ we have reached the stage where “the network is the message.”

Specifically, a new type of information security threat involves a class of malware, the goal of which is not to corrupt and take control of the machines it infects or steal explicit information stored on them (such as credit card information and personal records). Rather, the goal is to steal social network and behavioral information through data collection and network science

Related Work in Reality Mining

The social sciences have been undergoing a digital revolution, heralded by the emerging field of *computational social science*, which combines the leading techniques from network science^{1–3} with new machine learning and pattern recognition tools specialized for the understanding of people’s behavior and social interactions.^{4,5} David Lazer and his colleagues described the potential of computational social science to increase our knowledge of individuals, groups, and societies, with an unprecedented breadth, depth, and scale.⁶ The pervasiveness of mobile phones has made them ubiquitous social sensors of location, proximity, and communications.

The phrase “reality mining” describes the collection of sensor data pertaining to human social behavior.⁷ Using call records, cellular-tower IDs, and Bluetooth proximity logs collected via mobile phones at the individual level, the subjects’ social network can be accurately detected as well as patterns in their daily activities.^{4,7} Mobile phone records from telecommunications companies have proven valuable for uncovering human-level insights. For example, researchers have used cell-tower location information to characterize human mobility.⁸ Nathan Eagle, Michael Macy, and Rob Claxton found that the diversity of individuals’ relationships is strongly correlated with the economic development of communities.⁹ Expanding on an earlier work,⁷ Anmol Madan and his colleagues showed how to use mobile social sensing to help measure and predict individuals’ health status based on mobility and communication patterns.¹⁰

Companies such as Sense Networks are already putting such tools to use in the commercial world to understand customer churn, enhance targeted advertisements, and

offer improved personalization and other services. The technical advancements in mobile phone platforms and the availability of mobile software development kits (SDKs) are making it easier than ever to collect reality-mining data.

References

1. A.-L. Barabasi and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, no. 5429, 1999, pp. 509–512.
2. D. Watts and S. Strogatz, “Collective Dynamics of ‘Small-World’ Networks,” *Nature*, vol. 393, no. 6684, 1998, pp. 440–442.
3. M. Newman, “The Structure and Function of Complex Networks,” *SIAM Rev.*, vol. 45, 2003, pp. 167–256.
4. N. Eagle, A. Pentland, and D. Lazer, “Inferring Social Network Structure Using Mobile Phone Data,” *Proc. Nat’l Academy of Sciences*, vol. 106, no. 36, 2009, pp. 15274–15278.
5. C. Au Yeung et al., “Measuring Expertise in Online Communities,” *IEEE Intelligent Systems*, vol. 26, no. 1, 2011, pp. 26–32.
6. D. Lazer et al., “Social Science: Computational Social Science,” *Science*, vol. 323, no. 5915, 2009, pp. 721–723.
7. N. Eagle and A. Pentland, “Reality Mining: Sensing Complex Social Systems,” *Personal and Ubiquitous Computing*, vol. 10, 2006, pp. 255–268.
8. M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabasi, “Understanding Individual Human Mobility Patterns,” *Nature*, vol. 453, no. 7196, 2008, pp. 779–782.
9. N. Eagle, M. Macy, and R. Claxton, “Network Diversity and Economic Development,” *Science*, vol. 328, no. 5981, 2010, pp. 1029–1031.
10. A. Madan et al., “Social Sensing for Epidemiological Behavior Change,” *Proc. 12th ACM Int’l Conf. Ubiquitous Computing*, ACM Press, 2010, pp. 291–300.

inference techniques. We call this a *stealing-reality attack*.

After exploring this new kind of attack, we analyze how 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, in many cases, such an optimal strategy should follow an extremely slow spreading pattern. Counterintuitively, such attacks generate far greater damage in the long term compared to attacks that spread more aggressively. In addition, such attacks will likely avoid detection by many of today’s network security mechanisms, which tend to focus on detecting network traffic anomalies, such as a traffic-volume increase. We demonstrate this discovery using several real-world social networks datasets.

Stealing Reality Threat Model

In our discussion, we refer to *reality information* as inferred information about personal and social behavior. This includes three elements:

- information about individuals, which we refer to as *node information* (including any parameter on a node that can be learned from available data, such as occupation, income level, health state, or personality type);
- dyadic information, which includes information on relationships and other parameters connecting two nodes (we call this *edge information*); and
- network-level information, or information on groups of nodes, communities, and general network properties and information.

The full network information also includes all data on nodes and edges. We do not refer here to explicitly stated information that can be found in (and stolen from) existing databases, such as names, social security, or credit card numbers. Whereas reality mining is the legitimate collection and analysis of such information, reality stealing is the illegitimate accrual of it. (See the “Related Work in Reality Mining” for previous research in this area.)

Motivation for Attackers

Secondary markets for the resale of stolen identities already exist, such as www.infochimps.com or black market sites and chat rooms for the resale of other illegal datasets.² It is reasonable to assume that an email address of a social hub would be worth more

to an advertiser than that of a social leaf, or that a person meeting a student profile might be priced differently than a person meeting a corporate executive profile. Stolen reality information could be used for several malicious goals:

- selling to the highest bidder (legitimate bidders, advertisers, and so on or in the black market to other attackers),
- bootstrapping other attacks (as part of a complex advanced persistent threats [APT] attack^{3,4}), and
- business espionage (for example, to analyze a competitor's customer base, profile high-yielding customers for targeted marketing,⁵ or produce high-quality predictions⁶).

Companies are already operating in this area, collecting email and demographic information with the intent to sell it. Methods for social network analysis and trends recognition have already been published in many leading venues.⁷ Why should attackers work hard when they can use automatic agents to collect the same data and possibly much higher-quality information?

Dangers of Reality-Stealing Attacks

Users can modify their communication network topologies and networked-device identifiers with the click of a button. In the event of a security breach, users can also change their passwords, usernames, and credit cards numbers; easily replace their email and other online accounts; and quickly warn their contacts. However, it is much harder to change one's social network, person-to-person relationships, friendships, or family ties. If a chronic health condition is uncovered through such an attack, there is no going back.

Victims of a behavioral-pattern theft cannot change their behaviors and life patterns. This type of information, once out, would be difficult to contain.

A second component accentuating this danger is that real-life information can be deduced from seemingly safe data, such as accelerometer and location information, which users already freely allow many mobile applications to access.

Because we believe this is a concrete threat, our research goal is to analyze potential attacks from the attackers' perspective to better understand them and develop proper defenses.

Past Attacks on Real-World Information

To help understand the risk of attacks on inferred real-world information, we reviewed prior attacks on explicit data. In 2008, real identity information of millions of Korean citizens was stolen in a series of malicious attacks and posted for sale. In 2007, the Israel Ministry of Interior's database, which contained information on every Israeli citizen, was leaked and posted on the Web.⁸ In the US, a court is ruling whether a database from a bankrupt gay-dating site for teenagers can be sold to raise money to help repay its creditors; the site includes the personal information of more than a million teenage boys.⁹

In all these cases, once the information is out, there is no way to get it back, and the damage might be felt for an extended period. In a recent *Wall Street Journal* interview, former Google CEO Eric Schmidt referred to the possibility that people in the future might choose to legally change their name to detach themselves from embarrassing "reality" information publicly exposed in social networking sites. This demonstrates the sensitivity

and challenges in recovering from leaked real-life information, whether by youthful carelessness or malicious extraction through an attack.¹⁰

Many existing viruses and worms use primitive forms of social engineering,¹¹ which attempt to gain the trust of their next victims and then convince them to click on a link or install an application. For example, Happy99 was one of the first viruses to attach itself to outgoing emails, thus increasing the chances of the recipient opening an attachment to a seemingly legitimate message from an acquaintance. (More information concerning security and privacy leakage in social networks is available elsewhere.^{12,13})

Social Attack Model

We model a 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 about 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 will try to copy itself to a neighboring node at any given time step determines the aggressiveness of the attack, ρ . Namely, aggressive agents have higher ρ values (and hence take less time between each two spreading attempts). Less aggressive agents are less likely to try and spread at any given time and generally will wait longer between trying to copy themselves to one of the neighbors of their current host.

As the information about the network becomes worthy of an attack, the attacker's motivation is stealing as many properties related to the network's social topology as possible. We denote the percentage of vertices-related information acquired at time

t as $\Lambda_V(t)$ and the percentage of edge-related information acquired at time t is as $\Lambda_E(t)$.

The duration of the stealing-reality attack's learning process 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:

- The longer such an agent operates, the more precise its conclusions will be. We call this the *sigmoidal advancement*.
- The rate at which information is gathered is smallest at the start and end of the learning process.
- For any value of T , the amount of information gathered in the first T timesteps is greater than the amount of information gathered at the last T timesteps. We call this the *asymmetry of the asymptotes*.

Previous research demonstrated the applicability of the Gompertz function for the purpose of modeling the evolution of locally learning the preferences and behavior patterns of users.¹⁴ The authors attempted to predict which applications mobile users would install on their phones using an ongoing learning process. This experiment showed that this process can be best modeled using the function $1 - e^{-x}$. Because we know that $1 - t \leq e^{-t}$ (achieving tight results for most $t < 1$), we can clearly see that $1 - e^{-x} \approx e^{-e^{-x}}$, which is an instance of the Gompertz function (for $a = 1$, $b = c = -1$).

Users and administrators are more likely to detect an aggressive spreading

pattern, resulting in the subsequent blocking of the attack. On the other hand, attacks that spread slowly might evade detection for a longer period of time, but the amount of data they gather would be limited. To predict the detection probability of the attack at time t , we use Richard's curve¹⁵—a generalized logistic function often used for modeling the detection of security attacks:

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

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

Let $I_u(t)$ be the infection indicator of u at time t , T_u be the initial infection time of u , and $p(u, t)$ be the Gompertz function. Defining

$$\Lambda_V(t) = \frac{1}{|V|} \sum_{u \in V} I_u(t) \cdot p(u, t - T_u)$$

we get

$$\Lambda_V(\rho) = \int_0^{\infty} \left(\frac{\partial \Lambda_V(t)}{\partial t} \cdot (1 - p_{\text{detect}}(t)) \right) dt$$

Social Learnability

We defined a mathematical measure that predicts an attacker's ability to steal or acquire a given social network—what we call a network's *social learnability*. This measure reflects both the information contained in the network and the broader context from which the network was derived. Using this measure, we can sort real-world social networks according to their complexity (which is known) and even group two different social networks that were generated by the same group of people. The optimal

learning process with respect to this new measure in many cases involves nonaggressive attacks.

Information Complexity of Social Networks

A network's Kolmogorov complexity represents the basic amount of information contained in a social network.¹⁶ For example, a military organization's network has many homogeneous links and hierarchical structures. 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 heterogeneous network, consisting of many types of relationships such as work relationships, physical proximity, family ties, and other intricate types of social relationships and group affiliations.

Let K_E denote a network's Kolmogorov complexity, or the minimal number of bits required to "code" the network in such a way that it could later be completely restored.

Social Entropy of Social Networks

Every social-reality network belongs to one or more *social families*, each of which has its own consistency (or versatility). Some families might contain a great variety of possible networks, each having roughly a similar probability to occur, while another might consist of a limited number of possible networks. Each network's complexity, however, does not necessarily correlate with its entropy. For example, families of low variety might in fact be highly complicated networks, while other families might 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

$$\left\{x_1, x_2, \dots, x_{2^{\frac{1}{2}(|V|(|V|-1))}}\right\}$$

(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:

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

where $\zeta_{|V|}$ denotes the number of distinct nonisomorphic simple graphs of $|V|$ nodes. $\lambda(G)$ is then defined as $\lim_{n \rightarrow \infty} \lambda_n(G)$.

Stealing a Network’s Social Essence

Reed’s law asserts that the utility of large networks (and particularly social networks) can scale exponentially with the size of the network. This is because the number of possible subgroups 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^I social principles behind a network, where I is the information 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. Because K_E is measured in bits, the appropriate normalization should maintain this scale. Multiplying by $\lambda(G)$, the normalized social entropy

of the network G , the network information can be written as follows:

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

After normalizing by the network’s overall social essence ($\Lambda_E = 1$), we achieve the following measurement for the social essence of the subnetwork acquired:

$$\begin{aligned} \Lambda_s(t) &= \frac{2^{\lambda(G) \cdot K_E \frac{\log_2(|E|\Lambda_E(t))}{\log_2|E|}}}{2^{\lambda(G) \cdot K_E}} \\ &= 2^{\lambda(G) \cdot K_E \frac{\log_2 \Lambda_E(t)}{\log_2|E|}} \end{aligned}$$

which yields

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

K_E represents the network complexity, whereas $\lambda(G)$ represents the complexity of the network’s social family.

At this point, we assert that our social-learnability measure is a valuable property for measuring network attacks. For this, we demonstrate the values of this measure for several real-world networks. Figure 1 presents an analysis of the networks derived from the Social Evolution experiment,¹⁷ the Reality Mining network,¹⁸ and the Friends and Family experiment.¹⁹ We can easily see the logic behind the predictions received using the social-learnability measure concerning the difficulty of learning each network.

Specifically, we determined the Social Evolution network is harder to steal than the Reality Mining network, but it is easier to steal than the Friends and Family networks. Whereas the Reality Mining experiment tracked people within a relatively static work environment, the Social Evolution experiment took place at Massachusetts Institute of

Technology undergraduate dorms, involving students with (apparently) much more complicated mobility and interactions patterns. The Friends and Family dataset involved even more complicated interactions because it includes a heterogeneous community of couples, increasing the amount of information encapsulated within the network.

In addition, the social-learnability 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 in terms of volume, meaning, and network information. Still, because the two networks essentially represent the same social group of people, their social-learnability measure has a similar value.

Figure 2 demonstrates the importance of a network’s social entropy, analyzing the Reality Mining network for various possible values of social entropy. We approximated the value for the network’s Kolmogorov complexity using an LZW compression.

Figure 3 demonstrates the progress of the network-essence-stealing process for various network-complexity values. As the amount of information contained in a network increases—that is, the network represents more complicated social structures—the network becomes much more difficult to acquire.

Experimental Results

We evaluated 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. The data consists of approximately 200,000 nodes and 800,000 edges. Figure 4 shows the attack efficiency (namely, the maximal amount of network information acquired) as a function of its aggressiveness—that is, the attack’s

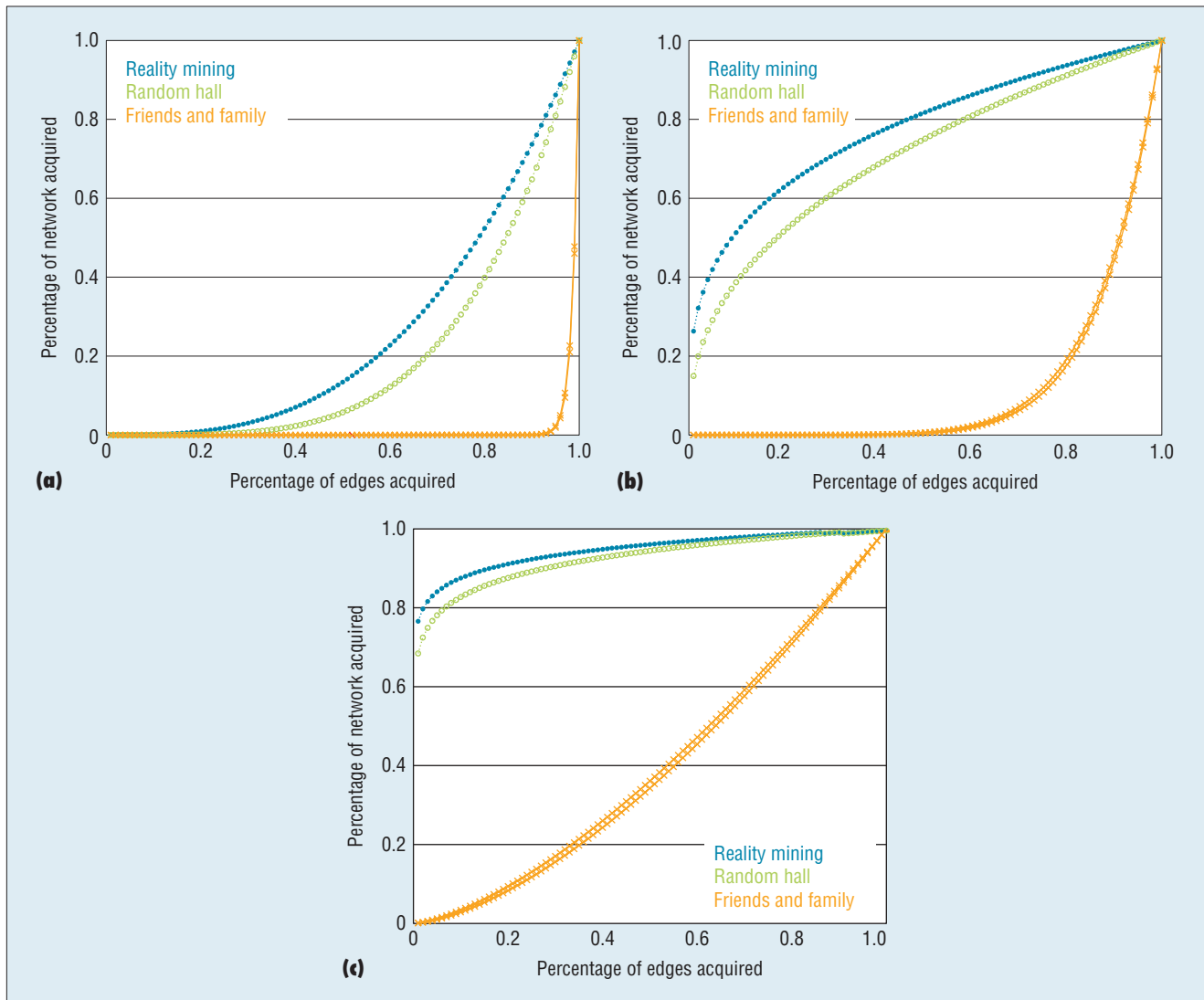


Figure 1. Reality-stealing process. We used three values of social entropy (a) $\lambda(G) = 1$, (b) $\lambda(G) = 0.1$, and (c) $\lambda(G) = 0.02$, for four networks: the Random Hall network,¹⁷ Reality Mining network,¹⁸ Friends and Family self-reporting network, and Friends and Family Bluetooth network.¹⁹ 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 than the Friends and Family networks.

infection rate. The two curves represent the amount of information (related edges and vertices) that can be obtained as a function of the aggressiveness value ρ . 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 $\rho = 0.04$.

To further validate our analytic model for predicting the success of

stealing-reality attacks, we simulated attacks for random subnetworks of our real-world 200,000-node mobile network using various attack-aggressiveness values. We used numerous sets of values for the attack properties, and for each, we empirically measured the overall expected amount of information stolen by the attack. Although the actual percentage of stolen information varied significantly between the various simulations, demonstrating the influence of changes made to the

attack's properties, many displayed the same interesting phenomenon: a global optimum for the attack's performance located around a low value of ρ . Figure 5 presents some of these scenarios.

To further validate our theoretical attack model, we used a small-scale real-world social network we obtained from the Friends and Family study containing data derived from a multitude of mobile mounted sensors (including call logs, accelerometers, Bluetooth, and WiFi interactions).

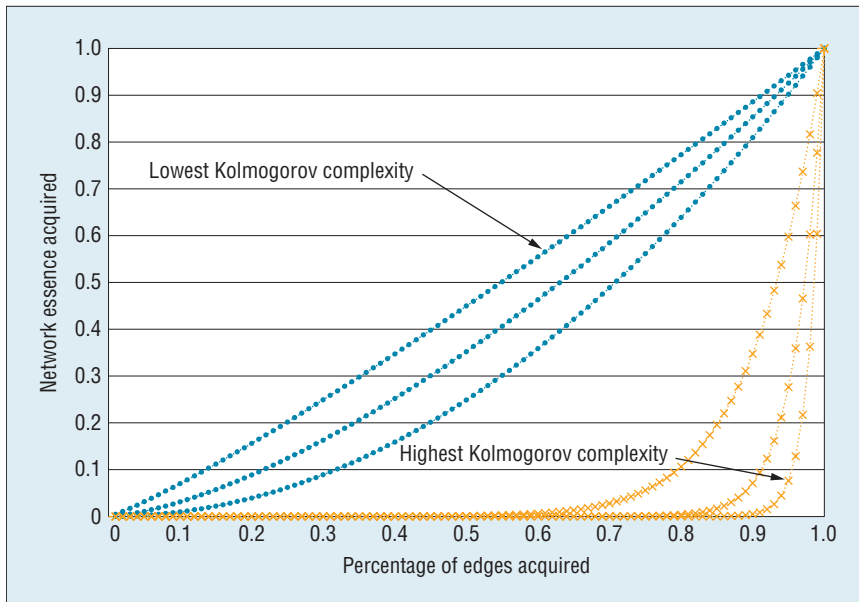


Figure 2. Network social entropy. This example illustrates a network’s social entropy $\lambda(G)$ for the Reality Mining network.¹⁸ The curves represent an approximation of the social essence measure calculated using an LZW compression of the Reality Mining network. If we assume that the network is 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.

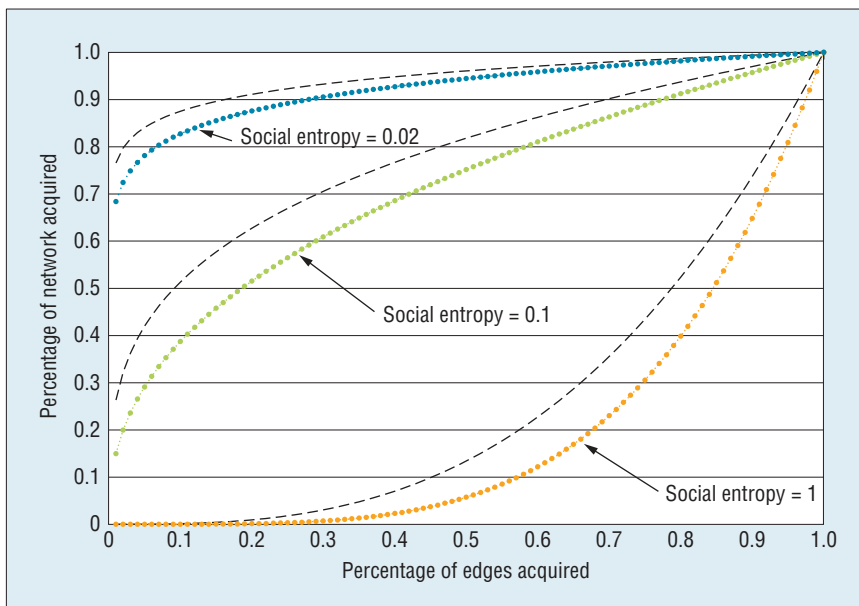


Figure 3. Evolution of Λ_5 as a function of the overall percentage of acquired edges. For networks of the same number of edges ($|E| = 1,000,000$), we assume the same social entropy $\lambda(G) = 0.1$, with different levels of Kolmogorov complexity.

Using this data, we confirmed our assumptions concerning the learning process.¹⁴ Our research currently

focuses on the empirical implementation and measurement of the model we presented here.

The new concept of stealing-reality attacks might provide an explanation for observed evidence in the process of investigating recent advanced persistent threats (APT) attacks as well as suggest that such attacks might have occurred in the past and gone undetected. Such attacks are difficult to detect because most existing network monitoring methods focus on detecting other, noisier attack attempts. Systems such as the Network Telescope²⁰ are designed to detect activity in IP segments that are supposed to contain no such activities. Other widely used methods rely on detecting anomalies in network activity,^{21,22} for which a considerable amount of data is required.

As a result, a nonaggressive attack might avoid detection. Finally, the attack is sensitive to the accuracy of the selection of the optimal aggressiveness value (Figure 4), which further hints at the usefulness of the attack for entities such as global hacking organizations or national defense agencies that have the resources needed to gather the information required for such accurate estimation. \square

Acknowledgments

This material is based upon work supported by the US National Science Foundation under grant number 0905645. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

1. M. McLuhan, *Understanding Media: The Extensions of Man*, Mentor, 1964.
2. C. Herley and D. Florencio, “Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy,” *Economics of Information Security and Privacy*, T. Moore, D. Pym, and C. Ioannidis, eds., Springer, 2010, pp. 33–53.

3. "The Advanced Persistent Threat (APT)," white paper, Solutionary, 2011.
4. B.E. Binde, R. McRee, and T.J. O'Connor, *Assessing Outbound Traffic to Uncover Advanced Persistent Threat*, tech. report, Sans Inst., 2011.
5. M. Brunner et al., *Infiltrating Critical Infrastructures with Next-Generation Attacks*, tech. report, Fraunhofer-Inst. for Secure Information Technology, 2010.
6. L. Tang and H. Liu, "Toward Predicting Collective Behavior via Social Dimension Extraction," *IEEE Intelligent Systems*, vol. 35, no. 5, 2010, pp. 19–25.
7. D. Barbieri et al., "Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics," *IEEE Intelligent Systems*, vol. 25, no. 6, 2010, pp. 32–41.
8. N. Jeffay, "Israel Poised to Pass National I.D. Database Law," *The Jewish Daily Forward*, 2009; www.forward.com/articles/112033.
9. D. Emery, "Privacy Fears over Gay Teenage Database," BBC News, 2010; www.bbc.co.uk/news/10612800.
10. R.M. Stana and D.R. Burton, *Identity Theft: Prevalence and Cost Appear to Be Growing*, tech. report GAO-02-363, US General Accounting Office, 2002.
11. S. Granger, "Social Engineering Fundamentals, Part I: Hacker Tactics," Symantec, 2001; www.symantec.com/connect/articles/social-engineering-fundamentals-part-i-hacker-tactics.
12. R. Gross and A. Acquisti, "Information Revelation and Privacy in Online Social Networks," *Proc. 2005 ACM Workshop Privacy in the Electronic Society (WPES)*, ACM Press, 2005, pp. 71–80.
13. A. Korolova et al., "Link Privacy in Social Networks," *Proc. 17th ACM Conf. Information and Knowledge Management (CIKM)*, ACM Press, 2008, pp. 289–298.
14. W. Pan, N. Aharoni, and A. Pentland, "Composite Social Network for Predicting Mobile Apps Installation," *Proc. 25th Conf. Artificial Intelligence*, 2011, pp. 821–827.
15. N.A. Christakis and J.H. Fowler, "Social Network Sensors for Early Detection of Contagious Outbreaks," *PLoS ONE*, vol. 5, 2010, p. e12948.
16. A. Kolmogorov, "Three Approaches to the Quantitative Definition of

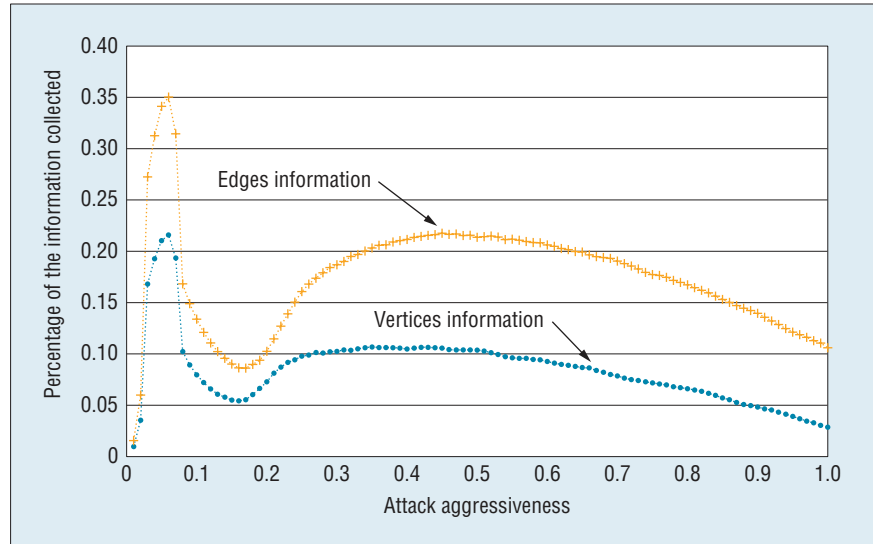


Figure 4. Overall amount of data that can be captured by a stealing-reality attack. This example illustrates the phenomenon where the most successful attack possible (namely, an attack capable of stealing the maximal amount of information) is produced by a low attack aggressiveness ρ value. The upper curve represents $\Lambda_E(\rho)$, the overall percentage of edge-related information stolen. The lower curve represents $\Lambda_V(\rho)$, the overall percentage of vertices-related information stolen. The local maximum around $\rho = 0.5$ is outperformed by the global maximum at $\rho = 0.04$.

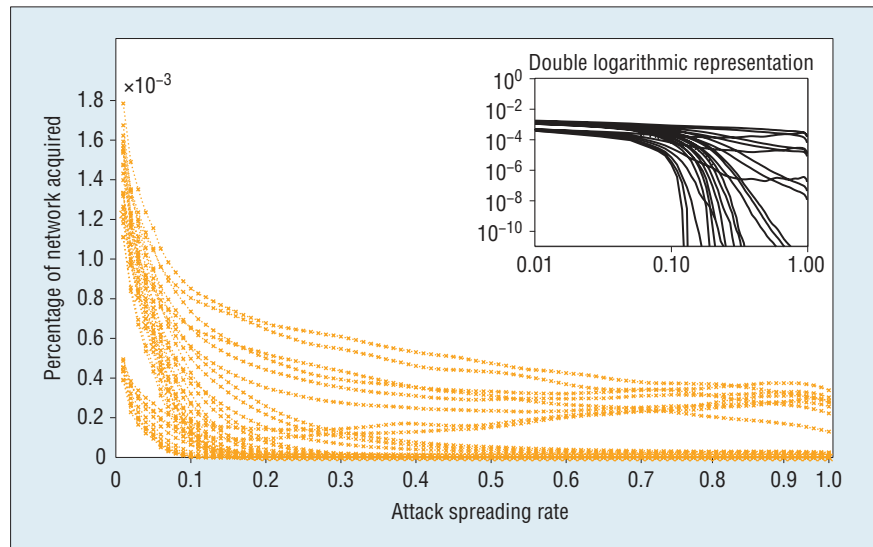


Figure 5. Extensive study of a real-life mobile network, simulating stealing-reality attacks. The performance of each scenario is measured as the percentage of information acquired, as a function of the infection rate ρ . The scenarios that are presented in this figure demonstrate a global optimum of the attack performance for low ρ values, stressing the fact that in many cases an extremely nonaggressive attack achieves the maximal amount of stolen information.

THE AUTHORS

Yaniv Altshuler is a postdoctoral associate at the Massachusetts Institute of Technology Media Lab and an adjunct faculty member in the Computer Science Department at the Technion, Israel Institute of Technology. His research interests include information flow dynamics, social network analysis, and swarm algorithms in dynamic communities. Altshuler has a PhD in computer science from Technion. Contact him at yanival@media.mit.edu.

Nadav Aharony is a doctoral student in the Human Dynamics Group at the MIT Media Lab. His research interests include revolves around the intersection between communication networks, social dynamics, and systems that learn. Aharony has a BS in electrical engineering from Technion. Contact him at nadav@media.mit.edu.

Alex "Sandy" Pentland is the Toshiba Professor of Media, Arts, and Sciences at MIT and directs the MIT Human Dynamics Lab. He also directs the Media Lab Entrepreneurship Program, spinning off companies to bring MIT technologies into the real world. He is a pioneer in computational social science, organizational engineering, and mobile information systems and is developing computational social science and using this new science to guide organizational engineering. Pentland has a PhD in AI and psychology from MIT. Contact him at sandy@media.mit.edu.

Yuval Elovici is an associate professor in the Department of Information Systems Engineering and the director of the Deutsche Telekom Research Laboratories at Ben Gurion University of the Negev, Israel. His research interests include computer and network security, privacy and anonymity in an electronic society, social network security, complex networks, and detection of malicious code using machine learning techniques. Elovici has a PhD in information systems management from Tel Aviv University. Contact him at elovici@bgu.ac.il.

Manuel Cebrian is a research scientist in the Department of Computer Science and Engineering at the University of California, San Diego. His research focuses on the advancement of computational social science, an emerging field that collects and analyzes data at a scale capable of revealing patterns of individual and group behaviors. Cebrian has a PhD in computer science from the Autonomous University of Madrid. Contact him at mcebrian@ucsd.edu.

Information," *Problems Information Transmission*, vol. 1, no. 1, 1965, pp. 1–7.

17. A. Madan et al., "Social Sensing for Epidemiological Behavior Change," *Proc. 12th ACM Int'l Conf. Ubiquitous*

Computing, ACM Press, 2010, pp. 291–300.


18. N. Eagle, A. Pentland, and D. Lazer, "Inferring Social Network Structure Using Mobile Phone Data," *Proc. Nat'l Academy of Sciences*, vol. 106, no. 36, 2009, pp. 15274–15278.

19. N. Aharony et al., "The Social fMRI: Measuring, Understanding, and Designing Social Mechanisms in the Real World," *Proc. 13th ACM Int'l Conf. Ubiquitous Computing (UbiComp)*, ACM Press, 2011, pp. 445–454.

20. D. Moore et al., "Inside the Slammer Worm," *IEEE Security & Privacy*, vol. 1, no. 4, 2003, pp. 33–39.

21. F. Apap et al., "Detecting Malicious Software by Monitoring Anomalous Windows Registry Accesses," *Recent Advances in Intrusion Detection*, Springer, 2002, pp. 36–53.

22. R. Moskovitch et al., "Host Based Intrusion Detection Using Machine Learning," *Proc. IEEE Int'l Conf. Intelligence and Security Informatics (ISI)*, IEEE Press, 2007, pp. 107–114.

 Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.