The Socio-monetary Incentives of Online Social Network Malware Campaigns

Ting-Kai Huang
Google
Mountain View, CA
tingkaih@google.com

Bruno Ribeiro
Carnegie Mellon University
Pittsburgh, PA
ribeiro@cs.cmu.edu

Harsha V. Madhyastha
University of California Riverside
Riverside, CA
harsha@cs.ucr.edu

Michalis Faloutsos
University of New Mexico
Albuquerque, NM
michalis@cs.unm.edu

ABSTRACT

Online social networks (OSNs) offer a rich medium of malware propagation. Unlike other forms of malware, OSN malware campaigns direct users to malicious websites that hijack their accounts, posting malicious messages on their behalf with the intent of luring their friends to the malicious website, thus triggering word-of-mouth infections that cascade through the network compromising thousands of accounts. *But how are OSN users lured to click on the malicious links?* In this work, we monitor 3.5 million Facebook accounts and explore the role of pure monetary, social, and combined socio-monetary psychological incentives in OSN malware campaigns. Among other findings we see that the majority of the malware campaigns rely on pure social incentives. However, we also observe that malware campaigns using socio-monetary incentives infect more accounts and last longer than campaigns with pure monetary or social incentives. The latter suggests the efficiency of an epidemic tactic surprisingly similar to the mechanism used by biological pathogens to cope with diverse gene pools.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human factors*

General Terms

Human Factors, Measurement

Keywords

OSN Malware; Social Incentives; Monetary Incentives; Labor Markets

COSN'14, October 1–2, 2014, Dublin, Ireland.

http://dx.doi.org/10.1145/2660460.2660478.

1. INTRODUCTION

In 1949 von Neumann first suggested the possibility of creating self-reproducing computer programs [42]. Thirty-five years later Cohen wrote one of the first computer malwares, which he named a *computer virus* [7]. Since then, the connection between computer malware and biological viruses has captivated the imagination of both researchers and the general public [10, 11, 12, 36], and the manner in which malware interacts with people and computers has evolved. The inception of e-mail in the 60's created a new medium for malware developers [5, 15, 30].

Today, online social networks (OSNs) offer another new medium for malware propagation that, as shown in this and some of other of our recent studies [19, 33], is profoundly changing the face of malware. In OSN malware, OSN users are lured into visiting malicious websites containing *clickjacking* attacks[1] or into installing malicious in-OSN apps (e.g., Facebook apps). Once infected, the victim is impersonated in the social network, unknowingly exposing his or her friends to the same campaign through bogus direct messages or broadcast posts, creating a word of mouth infection that cascades through the network [19, 33]. One of the key features of OSN malware (a.k.a. socware [19, 33]) is leveraging on the perceived "endorsement" of hijacked users from the in the eyes of that user's friends. OSN malware is more than just a nuisance, it enables identity theft and cybercrime with several reported cases resulting in financial losses for the victims [20, 35].

But how are OSN users lured to click on these malicious links in the first place? We leverage on our prior work on detecting malware posts through a combination of keywords, anomalous user behavior, and topological anomalies [19, 33] to study how OSN malware exploits psychological incentives. Following Heyman and Ariely [17] classification of behavioral incentives in labor markets, we divide incentives into *monetary*, *social*, and *socio-monetary* (the latter is a combination of monetary and social incentives). These social and monetary incentives are, for instance, a pure monetary posts that promises a "free iPad"; pure social posts related to improving or checking your social status such as "Can you beat me in this game ⟨*link*⟩?" and "OMG check

[1]Clickjacking and other attack mechanisms are described in Huang et al. [19]

if a friend has deleted you! Click ⟨*here*⟩, it works!", or of shared curiosity "This is shocking ⟨*link*⟩!"; and finally (socio-monetary) a combination of social and monetary incentives (e.g., a friend's challenge with the promise of a free iPad if you win).

Contributions

One of the main contributions of this work is to study the impact of distinct socio-monetary incentives in the size and duration of malware campaigns[2], covering the posts of nearly 3.5 million Facebook users collected over ten months between July 2011 and April 2012. With the help of MyPageKeeper malware post detection heuristics [19, 33] and 226 Mechanical Turk [3] volunteers we classify thousands of unique Facebook posts. We note, however, that our monitoring is restricted to both users that installed MyPageKeeper and the posts of their friends that are visible to these users. But while we are limited to users of one online social network (Facebook) that volunteer to have their accounts protected by MyPageKeeper, the data collected from this viewpoint (of 3.5 million users) is of great interest. Aside from truly random monitoring without user consent, data collection from volunteers is prone to unknown biases.

We observe that 67% of the malware campaigns in our dataset use pure social incentives. Interestingly, and despite Heyman and Ariely's observations that subjects exposed to socio-monetary incentives act like subjects exposed to monetary incentives [17], we observe that combined socio-monetary incentives are more effective – in fact, stochastically dominant with respect to number of infected users and campaign durations – than campaigns using pure social or pure monetary incentives. For instance, malware campaigns with socio-monetary incentives last on average 136% longer than pure monetary and pure social campaigns.

Relation to Biological Pathogens

A simple explanation for the effectiveness of combined socio-monetary incentives is a type of percolation effect observed in plant pathogen epidemics over mixed crops [28, 46]. Through simulations we show that even if the susceptibility to combined incentives is less than that of any one specialized incentive, combining incentives provide a tremendous advantage to percolate over the network. On Facebook the mix is the likely propensity of distinct users to be more attracted to either monetary or social incentives. While there are other plausible explanations to why campaigns with socio-monetary incentives infect more users and last longer than campaigns with pure incentives (e.g., posts with socio-monetary incentives could be more difficult to classify as spam), the percolation effect described does not need extra (unverifiable) assumptions.

Outline

This work is organized as follows. Section 2 presents the necessary background to understand Facebook malware and their incentives. Section 3 presents the related work. Section 4 introduces our dataset, the malware incentives together, and the definitions used throughout this work. Section 5 shows how we classified malware posts. Section 6 reports statistics of the observed incentives at the campaign level. Section 7 shows that simulations of epidemics on real OSN topologies using mixed incentives can indeed outperform pure incentives, even if mixed incentives are not as effective than pure incentives in infecting to the subpopulation of individuals susceptible to the pure incentive. Finally, Section 8 discusses our results and future work.

[2]Malware campaigns are precisely defined in Section 4 but can be roughly thought as the spread of a malware associated with a specific website.

2. PRELIMINARIES

Facebook is the largest online social network ever created in the Internet's short history. With over 1.11 billion active users as of March 2013 [9], Facebook is a prime source of online social network data. Through the analysis of posts of over 3.5 million Facebook users collected over ten months, we observe a new generation of computer malware that relies heavily on two factors to spread through word of mouth: (a) incentive mechanisms provided by the post and (b) people's cognitive capacity to distinguish between a legitimate request from a friend and a bogus request from an infected friend. Other security factors are also known to play a role in security threats [4, 18, 25, 45], such as users' belief that they are less at risk than others, the fact that privacy and security are abstract concepts, and that it is hard for non-experts to judge risk. We leave the analysis of the analysis of these other security factors as future work, so we can instead focus our analysis on the role of socio-monetary incentives.

A representative example of the kind of incentive used on Facebook malware campaigns[3] is the campaign whose posts include the text "*OMG check if a friend has deleted you! Click ⟨here⟩, it works!*". This campaign simultaneously exploits the reader's incentive to know his or her social status in the group and the credibility (and *social capital* [37]) of the impersonated victim. The latter is remarkably different from messages seen in e-mail spam, an effect broadly felt on the use of keywords. For instance, "viagra" and "pills" are popular keywords in e-mail spam, but out of the hundreds of thousands of malicious posts we collected on Facebook, not a single one contains these keywords [33].

Once the post appears legitimate to the victim, a combination of the incentives in the post and the victim's susceptibility to the incentive drive the victim's decision as to whether or not to click on the malicious link. The data shows that social incentives often target the victim's social capital – increasing one's social capital is known as one of the reasons why people join online social networks [37] – or the victim's social insecurities about his or her social status in their social group.

3. RELATED WORK

While we are unaware of other works analyzing the impact of social and monetary incentives on online social network cascades, there is a rich literature on both percolation phenomena and computer viruses. Recent work on percolation phenomena over interdependent networks shows that a virus or a failure spreading over distinct but interconnected networks has a lower critical threshold than the same phenomenon spreading over a single network [14, 23, 43].

The connection between traditional computer viruses and biological pathogens has been extensively studied in the lit-

[3]Facebook actively combats malware alongside with third party application such as MyPageKeeper, the application we used to collect the data used in this study [19].

erature [7, 12, 36]. But these works do not consider the true networked nature of the infection. In the last decade there has been great interest in the connection between computer viruses and networks (e.g. [13, 32, 38]) but these works tend to look at worms and computer viruses that do not depend on human intervention to spread.

The role of diverse gene pools in the containment of biological pathogens has been studied since the early 70s [2, 6, 39]. Recent works in the literature make the connection between diversity of types and computer viruses. For instance, recently Wang et al. [44] conjectured that the spreading of viruses on mobile phones was hindered by the existence of two distinct popular (but incompatible) smartphone platforms. Even more recently, Newell et al. [29] proposed the use of a diverse set of network routers with distinct software (and possibly distinct vulnerabilities) as a way to deter malicious attacks that could compromise the network. The above works also disregard incentives and the effect of human intervention on the infection process. In business environments Yves et al. [8] study how to mitigate social engineering or insider attacks by using different administrative personnel.

In contrast to the above works, we study online social network malware that relies on user intervention to spread. Unlike worms and similar computer viruses, OSN malware must provide incentives to convince a user to perform actions that allow the malware to hijack that user's account. We analyze the incentives used by malware developers classifying these as social and monetary. Interestingly, we observe that malware campaigns with both monetary and social incentive are able to last longer and infect more users.

Adamic et. al. [22] study the evolution of memes of Facebook. They found that memes were copied or mutated by people during transmission. In contrast, our work focus on how malware developers conduct their malware campaigns with different incentives and how these incentives affect the duration and the number of infected users. Note that unlike Adamic et. al., victims of OSN malware often do not spread it willingly. The textual changes observed in the posts of a single malware campaign are overwhelmingly for text obfuscation purposes (likely to fool Facebook's malware detection tools). In this work we classify campaigns by the URL used in the attack rather than the text of the post as to avoid problems with obfuscation (in the following section we provide more details about our methodology).

4. DATASET AND DEFINITIONS

In this section, we describe some of the definitions used throughout this work. After presenting the definitions, we introduce our dataset.

Definitions. A *post* is a broadcast message that a Facebook user writes on her *timeline* (a.k.a. *wall*) that can be seen by all of her friends (modulo Facebook's post recommendation algorithm and privacy filters that users may have in place). We consider a *post text* to be the collection of all texts of posts that show small Levenshtein distance [24]. For instance, "Click here and win an iPad2" and "Click here and win an iPhone5" are considered posts with identical post text. A *campaign* is defined as the set of Facebook posts with the same unique URL, i.e., we are aggregating posts according to the website they advertise. Using URLs to fingerprint a campaign avoids the caveats of text-based classification or mutating post text. Some URLs are shortened using URL shortening services such as goo.gl and bit.ly. However, Huang et al. [19] shows that assigning posts to ' campaigns based on their URLs is similar to aggregating campaigns by the address of the final landing webpage.

Observation period	Jul'11–Apr'12
Number of observed posts with links	111 million
Number of malicious posts	164,304
Number of posts with text content	120,455
Number of campaigns with text	3,110

Table 1: Dataset summary.

A single post may contain one or more distinct incentives. For instance, 22.52% of the posts contain both social and monetary incentives. These posts are *combined socio-monetary incentive* posts. Otherwise the post has a *single incentive*. Our classification method is designed to find combined social and monetary incentives. A campaign may have both social and monetary incentives in two forms. Either the campaign sends messages with combined social and monetary incentives, or it uses two or more posts containing different incentives.

MyPageKeeper Summary: MyPageKeeper evaluates every URL that it sees on any user's Wall or News Feed to determine if the URL points to OSN malware. MyPageKeeper classifies a URL as malware if it points to a web page that 1) is known to spread malware, 2) attempts to "phish" for personal information, 3) requests the user to carry out tasks (e.g., fill out surveys) that profit the owner of the website, 4) promises false rewards, or 5) attempts to entice the user to artificially inflate the reputation of the page (e.g., forcing the user to "Like" the page to access a false reward). MyPageKeeper evaluates each URL using a classifier which leverages on hand-annotated and examples and post features that take into account the social reaction (context) of posts associated with the URL. For any particular URL, the features used by the classifier are obtained by combining information from all posts (seen across users) containing that URL. Example features used by MyPageKeeper's classifier include the similarity of text messages across posts and the number of comments and Likes on those posts. MyPageKeeper has false positive and false negative rates of 0.005% and 3%, respectively. For more details about MyPageKeeper's implementation and accuracy, please see Rahman et al. [33].

Datasets. Our data was collected through our MyPageKeeper Facebook app [1] from July 2011 to April 2012. We monitored the activity of 3.5 million Facebook users through the news feeds of 16,240 of MyPageKeeper's users[4]. In total, we identified 164,304 malicious posts, which contain at least one link out of 4,389 unique URLs. Within these posts we found 3,110 distinct posts that contain both text and a unique URL, the remaining 1,279 posts contain only URLs. Table 1 summarizes the dataset.

Detailed Description of MyPageKeeper

In what follows we present details of the MyPageKeeper app. The reader uninterested in the inner workings of MyPageKeeper can safely skip to the next section. The key

[4]On any user's news feed, Facebook selectively shows only 12% of updates posted by the user's friends. Hence, we will not see all the updates of any user's friends.

novelty of MyPageKeeper lies in the classification module (summarized in Figure 1(b) of Rahman et al. [33]). The input to the classification module is a URL and the related social context features extracted from the posts that contain the URL. Our classification algorithm operates in two phases, with the expectation that URLs and related posts that make it through either phase with- out a match are likely benign and are treated as such. We use whitelists and blacklists. To improve the efficiency and accuracy of our classifier, we use lists of URLs and domains in the following two steps. First, MyPageKeeper matches every URL against a whitelist of popular reputable domains. We currently use a whitelist comprising the top 70 domains listed by Quantcast, excluding domains that host user-contributed content (e.g., OSNs and blogging sites). Any URL that matches this whitelist is deemed safe, and it is not processed further.

All the URLs that remain are then matched with several URL blacklists that list domains and URLs that have been identified as responsible for spam, phishing, or malware. Using machine learning algorithms (trained with social context features) we evaluate all URLs that do not match the whitelist or any of the blacklists are evaluated using a Support Vector Machines (SVM) based classifier. We train our system with a batch of manually labeled data, that we gathered over several months prior to the launch of MyPageKeeper. For every input URL and post, the classifier outputs a binary decision to indicate whether it is malicious or not.

Our SVM classifier uses the following features: (a) Spam keyword score: keywords such as "FREE", "Hurry", "Deal", and "Shocked". To compile a list of such keywords we collect words that 1) occur frequently in blacklisted URL posts, and 2) appear with a greater frequency in OSN malware as compared to their frequency in benign posts. (b) Timeline post count: the more successful a spam campaign, the greater the number of users will be infected. Therefore, for each URL, MyPageKeeper computes counts of the number of Facebook timelines that contain the URL. (c) Like and comment count. Facebook users can "Like" any post to indicate their interest or approval. Users can also post comments to follow up on the post, again indicating their interest. (d) Level of URL obfuscation: hackers often try to spread malicious links in an obfuscated form, e.g., by shortening it with a URL shortening service such as bit.ly or goo.gl. We store a binary feature with every URL that indicates whether the URL has been shortened or not; we maintain a list of URL shorteners. Further details on the inner workings of MyPageKeeper can be found in Rahman et al. [33].

5. CLASSIFYING INCENTIVES

We classify the different types of incentives into one of the fourteen different incentives shown as a hierarchical taxonomy tree in Figure 1. These categories help our volunteers define whether the post contains a social, a monetary, or a socio-monetary incentive. At the highest level, we divide incentives into two categories: *monetary incentives* and *social incentives*. Under these two main categories, we further divided them into subcategories based on the mechanisms that are used to attract new victims. Table 2 presents three of the most informative frequent words of some of our incentive categories.

We use Amazon's Mechanical Turk platform [3] to recruit volunteers to classify thousands of unique Facebook posts. We refers to these volunteers as *turkers* in the remaining of the study. Mechanical Turk's main advantage over automatic text classifiers is its ability to actually reflect the reactions of real users to the post text. Simultaneously, Mechanical Turk allows us to perform the classification in much larger scale than with in-campus volunteers; by recruiting 226 turkers on Mechanical Turk, we avoid potential biases resulting from the use of a few volunteers doing a tedious task.

5.1 Mechanical Turk Classification

A total of 26.88% of all malicious posts in our dataset are URLs that do not contain any text. Facebook automatically displays a snapshot of the webpage URL, and as a result, when users read the post, they are subject to an incentivized post even though there is no text content in the post. Unfortunately, we do not consider these posts in this study as these webpages are short-lived [19], making it challenging to collect the necessary information. We defer the analysis of the incentives of these kind of posts for future study.

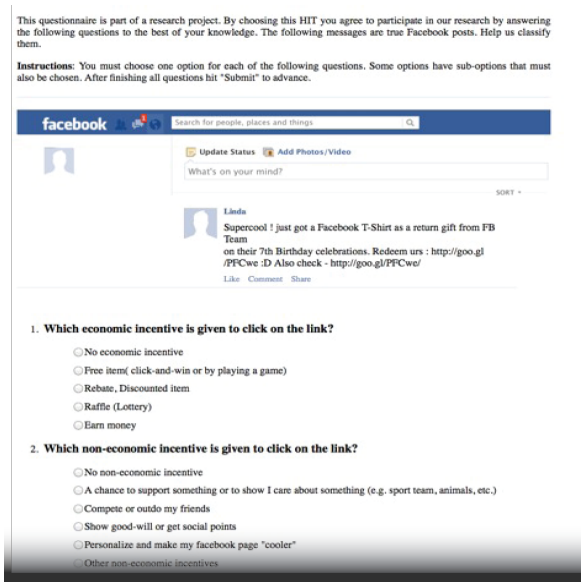


Figure 2: Screenshot of our Mechanical Turk survey.

From 120,455 malicious posts with text content, we identify 2,119 groups of posts, where posts in the same group contain text similar to each other with Levenshtein distance [24] less than 20. We select one post from each group to represent the group and ask workers two questions about the post:

1. Which monetary incentive is given to click on the link?

2. Which social incentive is given to click on the link?

On Amazon Mechanical Turk, tasks assigned to workers are called HITs (Human Intelligence Tasks). When we conduct our survey on Amazon Mechanical Turk, each HIT contains 19 different malicious posts and is assigned to 5 distinct *turkers*. To ensure the quality of the conducted survey, in each HIT, we add one control post with an obvious known incentive and one control question to each post. The results of a HIT completed by a *turker* are only valid if the

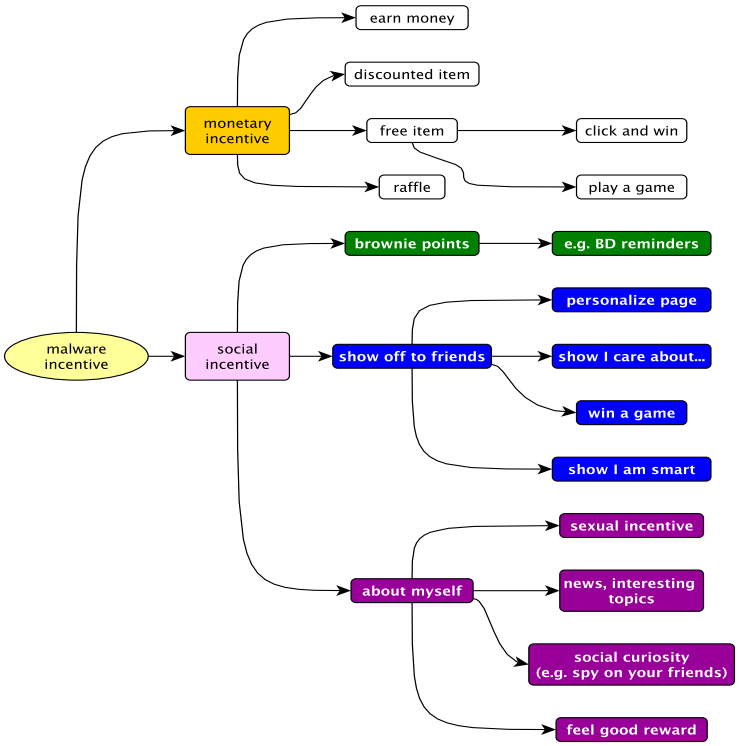


Figure 1: Taxonomy of malware post incentives.

Trait	Largest group	% of total
Age	25-34	39.42%
Gender	Male	55.29%
Education level	Some college	42.31%
Income	$25K - $37.5K	19.71%
Has FB account?	Yes	85.58%

Table 3: Demographical information of turkers

turker answers the control post and control questions correctly. Otherwise, we ignore it and assign the HIT to a new *turker*.

The content of a post is labeled social or non-social, monetary or non-monetary (socio-monetary are posts with both monetary and social labels) only if the majority of the *turkers* agree with the classification. In total, we successfully labeled 99.25% of all campaigns with a total disagreement rate of only 3%. Since these are malicious posts, we need to be careful with external links in the posts. To protect *turkers* we direct all links to our website explaining the nature of our study. To protect Facebook user's privacy, we do not include users' names on posts, replacing these names with some of the most common names and surnames in the U.S. population [40]. Figure. 2 shows a screenshot of our Mechanical Turk survey.

Each HIT consists of 19 malicious posts and one control post. Each post is rated by at least 5 *turkers*. We assign an incentive to a post by majority vote but to optimize labor costs we start assigning only 3 turkers to each HIT. We then create new HITs only for the posts where not all three

Control post	Get a free iPhone 4 Click here
Control question	$2 + 3 =?$

Table 4: Sample of a control post and a control question.

turkers agreed on the incentive. We stop creating new assignments when all posts have at least 5 turker ratings.

Qualification of turkers.

To ensure that turkers are reliable we filter turkers that can accept on our HITs by: (1) Only allowing turkers who have "Assignments Approval rate" higher than 95%. (2) Only allowing turkers who register their location as the United States. to ensure consistency, and a common linguistic background. (3) No turker can work on two HITs that share the same posts. In total, we get 226 turkers. Table 3 shows the self-reported information about the turkers that helped us classify campaign incentives.

Human subject verification.

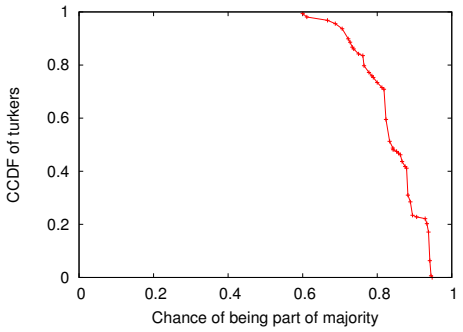
In order to ensure high quality in our classification results, we add one controlled message within each HIT and one controlled question in each post. Table 4 lists the control posts and sample of controlled questions. We design the incentive of the control post to be straightforward. If turkers cannot classify the controlled posts correctly the entire HIT is discarded.

We evaluate the quality of our survey by checking the design of the questions first. We want to make sure that turkers can understand our questions clearly. In the survey each question has 6-10 options. In total, $(85 - 90\%)$ of text

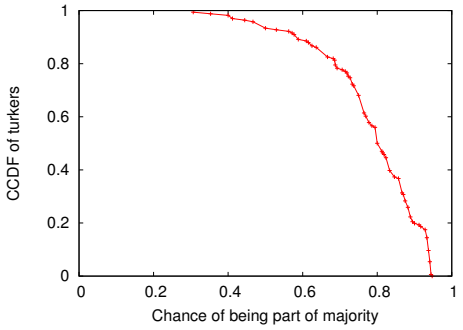
Category	Frequent Words
Click and win	Free, wow, offer
Play a game	contest, win, click
Discounted item	free, recharge, get
Best my friends in a game	context, win, click
Brownie points	just, got, miscrits
Raffle	cruise, win, holidays
Earn Money	make, moms, year
Sexual incentive	omg, shows, model
A chance to support something or to show I can about something (e.g. sport team, animals, etc.)	team, vote, favorite
Social curiosity (e.g., track profile visitors, monitor friend activities)	check, friend, deleted
News, interesting topics, etc. (e.g., Bin-Laden death video)	live, news, streaming
Feel-good reward (e.g., video of puppies)	freee, shirt, recharge
Personalize and make my Facebook page cooler	facebook, graphic, change

Table 2: Frequent words in each category

messages have at least 3 turkers voting the exact same answer according to our taxonomy in Figure 1 (e.g., the post contains a "Click-and-Win Rebate"). Second, we want to evaluate if our control questions are actually weeding out misbehaving turkers. Over the turkers that answered the control questions correctly, we select dissenting turkers who have at least one answer that disagrees with the majority of the other turkers of the same HIT. We found that $77-93\%$ of the remaining answers of the dissenting turkers agree with the majority, showing no consistent wrongdoing in the part of the dissenting turkers. Figure 3 lists the probability that the remaining answers of the dissenting turkers are the same as the majority. Based on the evaluation we just describe above, we believe that the results of our survey are trustworthy.



(a) Q: Which economic incentive is given to click on the link?



(b) Q: Which social (non-economic) incentive is given to click on the link?

Figure 3: Turker's classification performance.

6. USE OF INCENTIVES ON FACEBOOK MALWARE

In this section, we report the observed incentives at the campaign level. We also study these campaigns using metrics such as the number of infected accounts and duration of these campaigns. The duration of a campaign is defined as the period of time from the first day until the last day we observe newly infected users. Note that we consider a campaign to be alive only while it keeps infecting new users; we do not consider a campaign to be alive if it stops compromising new users. Figure 4 shows the evolution of the campaigns breaking them down into monetary, social, and socio-monetary incentivized campaigns relative to the total number of campaigns. The inset in Figure 4 shows the same plot with the absolute number of campaigns. Note that the overwhelming majority of the campaigns are social incentivized. Monetary and socio-monetary incentivized campaigns compete neck-in-neck.

Among campaigns with social incentives, we observe a sizable fraction of them lure users with social curiosity. For example, we observe 591 distinct campaigns—13% of all observed campaigns—with text content of the nature *"check if a friend has deleted you"*. Other representative text content of social curiosity campaigns—with observed 302 distinct campaigns—is *"Wow! l cant believe that you can check who is viewing your profile. l just checked my top profile lookers (sic) and l am shocked at who is seeing my profile! You can also see who viewed your profile ⟨here⟩"*.

We use two measures of campaign success: campaign duration and the number of infected users of a campaign. Figure 5(a) shows a semi-log plot of the complementary cumulative distribution (CCDF) of campaign durations separated by incentive. Note that socio-monetary campaigns are stochastically dominant over the duration of monetary and social campaigns. The average duration of a socio-monetary campaign is 17.8 days while the durations of pure social

Campaign type	(Total campaigns)	
Social campaigns	2075	68%
Monetary campaigns	540	17%
Socio-monetary campaigns	414	13%
Unclassified	81	3%

Table 5: Breakdown of malware campaigns for each incentive type (unclassified campaigns are campaigns where there was no agreement on the type of incentive provided).

Campaign type	duration	reach
All campaigns	7.1 days	28.1 users
Social campaigns	4.8 days	26.2 users
Monetary campaigns	7.5 days	22.3 users
Socio-monetary campaigns	17.8 days	43.2 users

Table 6: Average duration and reach (number of infected users) of malware campaigns.

and monetary campaigns are 4.8 and 7.5, respectively. The duration average by campaign incentive is summarized in Table 6.

The total number of infected accounts with socio-monetary campaigns is also stochastically dominant over the number infected with only monetary or social campaigns (see the semi-log CCDF of number of infected users in Figure 5(b)). Interestingly, we observe that the conditional distribution of the number of infected users given campaign durations (Figure 6) shows that campaign durations and the number of infected users in a campaign are mostly independent of each other. In other words, campaigns that last long do not necessarily infect more users. On average socio-monetary campaigns infect almost twice as many users as pure monetary campaigns (43.2 v.s. 22.3) and 165% more users than pure social campaigns. The average number of infected users (reach) by campaign incentive is summarized in Table 6.

To exemplify the dominance of socio-monetary campaigns, we select two similar "play/win a game" subcategories that belong to monetary and social incentives, denoted "play a game" and "win a game" in Figure 1, respectively. Table 7 shows the statistics of these game campaigns along with a sample post content. Observe that socio-monetary campaigns infect 4.4 times more users than campaigns with

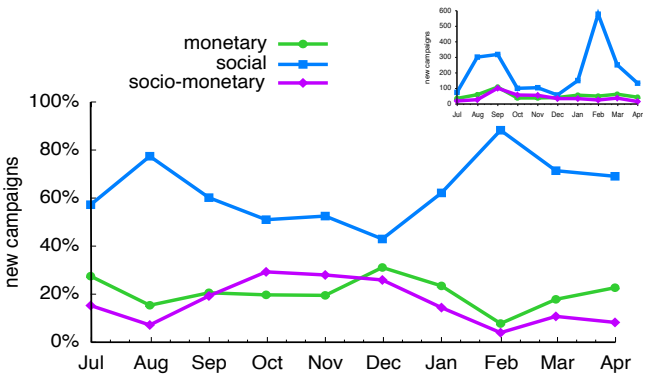


Figure 4: Evolution of the fraction (absolute numbers in the inset) of new campaigns per month per incentive.

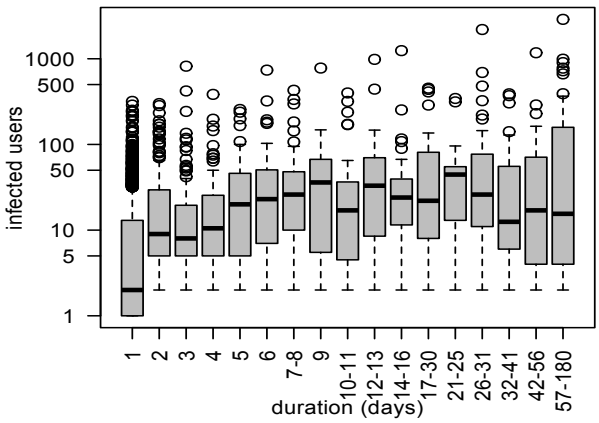


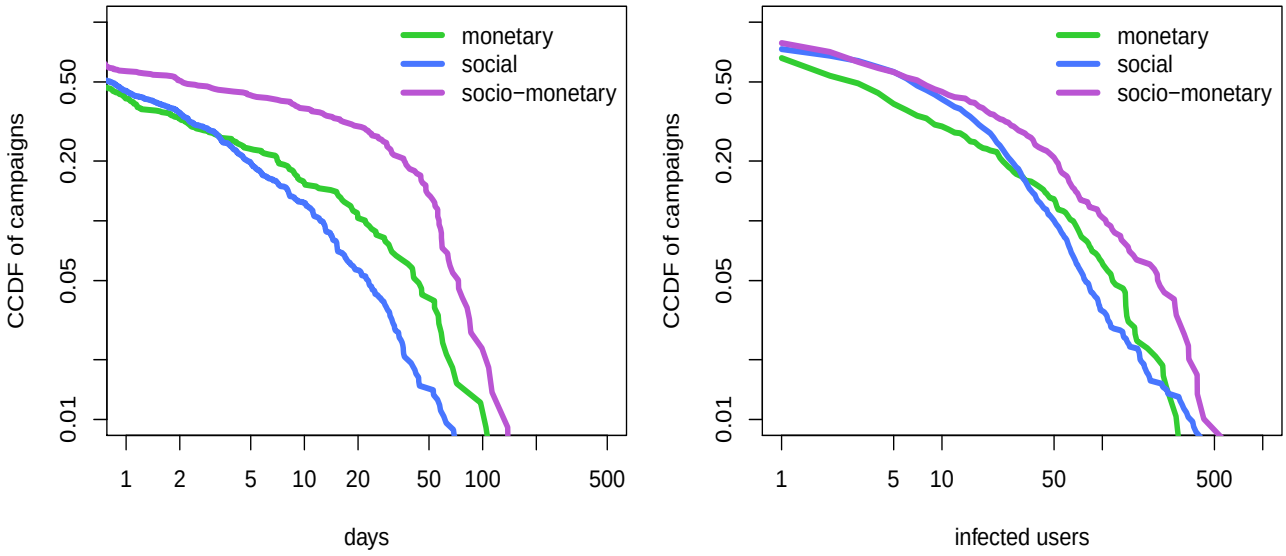
Figure 6: The number of infected users is loosely correlated with campaign durations. The graph shows the boxplot of the number of infected users of campaigns with a given duration. Duration values are sometimes binned to ensure that each bin has at least 31 samples.

pure incentives. Also note that, consistent with our general observations, there are more social game campaigns than monetary game campaigns, and that social campaigns are less effective than monetary campaigns.

These observations have remarkable real-world implications. For instance, stochastic dominance [16] implies that any rational agent betting on the outcome of a campaign whose utility function—for instance, the malware developer's revenue—increases according to the campaign duration or the number of infected (or any linear combination of the two) will choose socio-monetary campaigns. Another interesting observation is that while social incentivized campaigns are the majority of the observed campaigns, socio-monetary campaigns are superior and monetary campaigns are equally good if not better than social campaigns.

The observation that socio-monetary campaigns are more effective than pure monetary or pure social campaigns is rather surprising. Heyman and Ariely [17] show that subjects exposed to a combined socio-monetary incentive treat the incentive mostly as monetary and, thus, are less susceptible to the social incentive component. However, it is possible that the virulence of the monetary and social incentives that go into the socio-monetary campaign is more pronounced alone for certain users than when in combination. For instance, consider the following hypothetical scenario. Alice sees that her good friend Bob is challenging her to a game that also offers an iPad prize but is suspicious of the monetary prize (likely a scam) but would love to play the game with Bob. Carol is friends with Bob but does not know him very well and is not interested in the personal challenge but would love to play to win an iPad. In a scenario where socio-monetary incentives are not as virulent as pure incentives for a given set of users, could using socio-monetary incentives be still advantageous?

Without the possibility to infect users with malware we are unable to perform an experiment to answer the above question. However, we find a likely explanation can be found in a percolation phenomena well known in biological pathogens. Consider an extreme toy problem scenario. Fig-



(a) campaign durations

(b) infected users per campaign

Figure 5: Impact of monetary, social, and socio-monetary incentives on the duration and size of malware campaigns.

incentive	infections per campaign	campaigns	sample post content
social (game)	2.46	24	Some People will dominate all the games and some are doomed to remain losers their whole life (*sic*): ⟨link⟩
monetary (game)	16.18	11	NEW GAME NOTICE! Come check out the awesome new Wild Wild Taxi contest that is available, you could win a Kindle Fire. Start playing⟨here⟩
socio-monetary (game)	71.33	51	CONTEST UPDATE: Currently in 10246th place in The Daily Addi's Gem Swap II contest to win a 16GB iPad2. Think you can do better? You should give it a try ⟨here⟩

Table 7: Statistics of social, monetary, and socio-monetary game campaigns.

ure 7(a) shows the original toy network where users have a pure incentive preference: social (blue) or monetary (green). While in real life people are susceptible to both social and monetary incentives with different intensities, in our toy example, we consider the extreme case where they have a strict preference for either social or monetary incentives. That is, a node with a social (monetary) incentive preference may not be infected by campaigns using pure monetary (social) incentive. Figs. 7(b-c) show the percolation paths that can be taken by pure incentive campaigns. In contrast, socio-monetary incentivized campaigns see the percolation paths as in Figure 7(d). However, despite the denser percolation paths, the virulence of a socio-monetary campaign may not be as high as the virulence of a pure incentive campaign. This is because users with social (monetary) incentive preference may be put off by the mix with a monetary (social) incentive.

While there is recent work on percolation phenomena over interdependent networks, showing that a virus spreading over distinct but interconnected networks is more virulent than the same virus spreading over a single network [14, 23, 43], the phenomenon of interest in this work is a different type of percolation. The percolation consists of a flexible "pathogen" (socio-monetary campaign) that can infect users with distinct incentive preferences, contrasting the latter with a more specialized "pathogen" (pure social or monetary campaigns) that is more virulent in infecting users of a specific incentive preference but less virulent to other users. We find a similar scenario looking at plant pathogens.

One of the most famous cases of the epidemiological impact of a single gene variant in crops is the blight disease that hit potatoes in Western Europe, particularly Ireland, in the middle of the 19th century (1845–1852). The widely used single variety of potato used in Europe allowed the disease to spread quickly through the continent [2]. Since then, some commercial farmers adopted mixed-strain crops.

The *E. graminis* is a fungus that causes powdery mildew on barley. *E. graminis* has genotypes that if specialized to attack one barley strand are known to be highly virulent on that strand but significantly less effective on other strands [6]. In the presence of mixed strands flexible *E. graminis* genotypes evolve to attack different barley strands, but this flexibility comes at the cost of limiting the pathogen's ability to adapt with increased virulence to individual strands. Despite the decreased virulence that comes with flexibility, *E. graminis* is known to suffer strong genetic pressure to

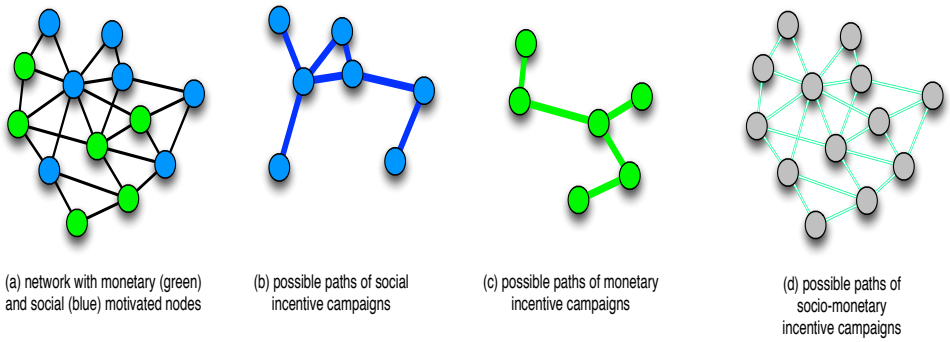


Figure 7: Malware spreading over a population with sharp incentive preferences such that users with one incentive preference may not be infected by the other incentive. Fig. (a) illustrates a network where users (nodes) prefer social (blue) or monetary (green) incentives. Figs. (b) and (c) show the possible edges that pure incentivized social and monetary campaigns can traverse, respectively. Fig. (d) shows the links that socio-monetary campaigns can traverse, however, with possibly lower virulence than pure incentive campaigns.

select flexible genotypes in mixed crops [6], an effect also extensively reported in simulations [28]. The percolation effect of specialized v.s. flexible pathogens on mixed crops is illustrated in Figure 8, showing why a more stable connectivity can tip the advantage towards flexibility. The reader is encouraged to see Mundt [28] for an interesting review on the effect of mixed crops on pathogen epidemics. This phenomenon is not limited to *E. graminis* [41] and provides an interesting mechanism to explain our observations: In populations with mixed susceptibilities to social and monetary incentives, flexible socio-monetary incentives are more effective due to percolation effects even if they are less virulent than pure incentives for some classes of users.

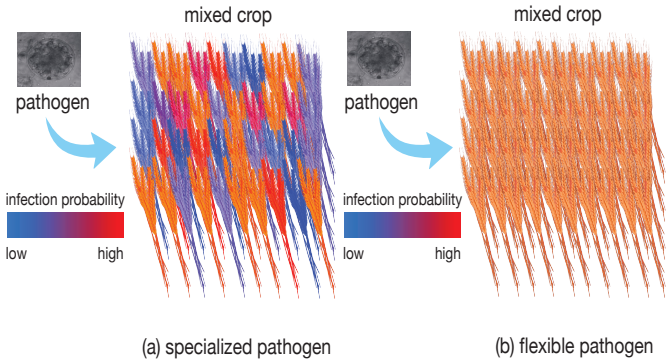


Figure 8: Percolation effect of specialized and flexible pathogens on mixed crops. In mixed crops, pathogen genotype flexibility reduces the infection probability variance, and possibly also its average. However, in most cases, the lower variability increases the percolation probability, facilitating the spread of the disease.

However, crops live in a two dimensional world—akin to a lattice—and small world effects of social networks can greatly impact epidemic cascades [27]. In what follows, we perform epidemic simulations over real social networks to show that, even in the presence of small world effects, the scale can also be tipped in favor of flexibility (i.e., socio-monetary incentives in our setting).

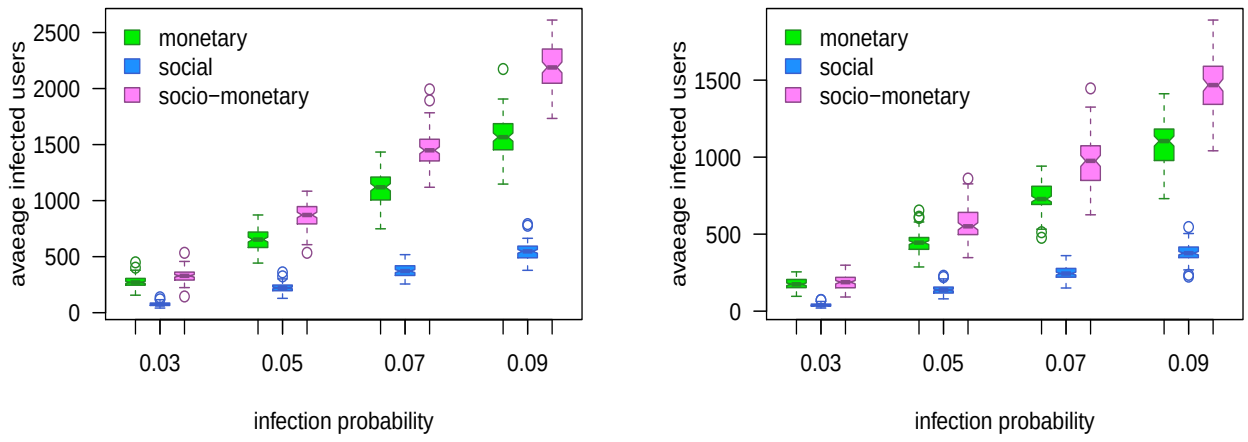
7. FLEXIBILITY V.S. VIRULENCE SIMULATIONS

We now use epidemic simulations over real social networks to show that, in the contest between flexibility (incentive combination) and virulence (pure incentive), the scale is tipped towards flexibility, thus providing a plausible explanation as to why socio-monetary campaigns outperform social and monetary campaigns. In other words, even if the combination of incentives reduces the virulence of each isolated incentive, incentive combination (flexibility) still produces larger cascades in most cases.

We simulate a variant of the Susceptible-Infected-Recovered (SIR) epidemic [31] that also takes into account user incentive preferences, a simple model to simulate the effects of different incentives over two distinct network datasets. The first network is the Enron e-mail network [21], which consists of 36,692 users. The Enron e-mail network is represented as an undirected graph between email senders and receivers. Malware campaign cascades are likely affected by trust relationships. To account for real (asymmetric) trust relationships, the second network dataset is the Epinions trust network [34], which consist of 75,879 users. The Epinion trust network shows a directed network of "who-trust-whom" among members of epinion.com. In our epidemic model, each user has a binary preference for social or monetary incentives, i.e., a purely socially incentivized campaign cannot infect monetary incentivized users and vice-versa. We randomly assign 60% of the users to prefer monetary incentives and the remaining 40% users to prefer social incentives.

At each time step, any infected user recovers with probability r. The recovery probability simulates the effect of the malware being discovered. Users that prefer social (monetary) incentives get infected with constant probability p by social (monetary) campaigns, but these same users cannot get infected by monetary (social) campaigns. The infection probability of socio-monetary campaigns is $p/2$ regardless of the incentive the user prefers. The latter captures the trade-off between flexibility and virulence.

Figure 9 shows box plots with the results of the average (over 300 campaigns) of number of infected nodes with different infection probabilities $p \in \{0.03, 0.05, 0.07, 0.09\}$. In



(a) The Enron network

(b) The Epinions trust network

Figure 9: Simulation results showing the number of infected users of distinct incentive types.

these simulations, we set the recovery probability r to be 0.5. The box plots have a "notch" around the median that offer a rough guide to the significance of difference of medians; if the notches of two boxes do not overlap, this offers evidence of a statistically significant difference between the medians [26]. As observed in Facebook, the number of infected users in socio-monetary campaigns is larger than that of pure monetary or social campaigns in both Enron (Figure 9(a)) and Epinions (Figure 9(b)). Note, however, that the advantage between socio-monetary incentives and pure monetary incentives is more prominent as we increase the infection probability p. Observe that typically socio-monetary campaigns infect more users than campaigns with pure incentives with or without taking into account asymmetric trust relationships.

8. DISCUSSION

In this work, we observe that post incentives play a major role on the spread of malware campaigns over online social networks. We observe that while social incentives are prevalent in OSN malware, socio-monetary incentives are more effective—with respect to the number of infected users and campaign durations—than social or monetary incentives when used in isolation.

The effectiveness of combined socio-monetary incentives can be explained by percolation effects resulting from mixed incentive preferences in the population of users. Through simulations we show that even if the susceptibility to combined incentives is less than that of any one specialized incentive, combining incentives provide a tremendous advantage to percolate over the network. This phenomenon is also observed on plant pathogens [28, 46], to which we make a variety of connections in our conclusions. In mixed crops, natural selection dislikes highly virulent but specialized pathogen to promote less virulent but unspecialized pathogens [28, 46], a mechanisms used by biological pathogens to cope with diverse gene pools [2, 28, 46]. Thus, the diversity of incentive preferences of OSN users may be a key factor in impeding the further spread of OSN malware, which should put selective pressure on OSN malware to evolve towards socio-monetary incentives but, interestingly, we found no evidence of such evolution over our 10 months of data.

Acknowledgements

This work was partially supported by NSF grant CNS-1065133 and ARL Cooperative Agreement W911NF-09-2-0053. The views and conclusions contained in this document are those of the author and should not be interpreted as representing the official policies, either expressed or implied of the NSF, ARL, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

9. REFERENCES

[1] MyPageKeeper. `https://apps.facebook.com/mypagekeeper/`.

[2] MW Adams, AH Ellingboe, and EC Rossman. Biological uniformity and disease epidemics. *BioScience*, pages 1067–1070, 1971.

[3] AWS. Amazon Mechanical Turk. `https://www.mturk.com/mturk/`, 2013.

[4] Kregg Aytes and Terry Connolly. Computer Security and Risky Computing Practices: A Rational Choice Perspective. *Journal of Organizational and End User Computing*, 16(3):22–40, 2004.

[5] J Balthrop. Technological Networks and the Spread of Computer Viruses. *science*, 304(5670):527–529, April 2004.

[6] K M Chin and M S Wolfe. Selection on Erysiphe graminis in pure and mixed stands of barley. *Plant Pathology*, 33(4):535–546, December 1984.

[7] Fred Cohen. Computer viruses: theory and experiments. *Computers & security*, 6(1):22–35, 1987.

[8] Yves Deswarte, Karama Kanoun, and Jean-Claude Laprie. Diversity against accidental and deliberate faults. In *Computer Security, Dependability, and Assurance*, pages 171–171. IEEE Computer Society, 1998.

[9] Inc. Facebook. Facebook reports first quarter 2013 results. March 2013.

[10] R Ford and E H Spafford. Computer science: Happy birthday, dear viruses. *Science*, 317(5835):210–211, July 2007.

[11] Stephanie Forrest and Catherine Beauchemin. Computer immunology. *Immunological Reviews*, 216(1):176–197, 2007.

[12] Stephanie Forrest, Steven A Hofmeyr, and Anil Somayaji. Computer immunology. *Communications of the ACM*, 40(10):88–96, 1997.

[13] a. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. *IEEE INFOCOM*, 2(C):1455–1466, 2005.

[14] Jianxi Gao, Sergey V Buldyrev, H Eugene Stanley, and Shlomo Havlin. Networks formed from interdependent networks. *Nature Physics*, 8(1):40–48, December 2011.

[15] Jacob Goldenberg, Yuval Shavitt, Eran Shir, and Sorin Solomon. Distributive immunization of networks against viruses using the 'honey-pot' architecture. *Nature Physics*, 1(3):184–188, December 2005.

[16] Josef Hadar and William R Russell. Rules for Ordering Uncertain Prospects. *The American Economic Review*, 59(1):25–34, January 1969.

[17] James Heyman and Dan Ariely. Effort for payment a tale of two markets. *Psychological Science*, 15(11):787–793, 2004.

[18] A E Howe, I Ray, M Roberts, M Urbanska, and Z Byrne. The Psychology of Security for the Home Computer User. In *Security and Privacy (SP), 2012 IEEE Symposium on IS* -, pages 209–223. IEEE, 2012.

[19] Ting-Kai Huang, Md Sazzadur Rahman, Harsha Madhyastha, Michalis Faloutsos, and Bruno Ribeiro. An analysis of socware cascades in online social networks. In *WWW*, 2013.

[20] Hyphenet. Facebook phishing scam costs victims thousands of dollars, `http://goo.gl/4uVME4`.

[21] Bryan Klimt and Yiming Yang. Introducing the enron corpus. In *First conference on email and anti-spam (CEAS)*, 2004.

[22] Eytan Adar Lada Adamic, Thomas Lento and Pauline Ng. The evolution of memes on facebook, `http://goo.gl/JysRpD`.

[23] E A Leicht and Raissa M D'Souza. Percolation on interacting networks. *arXiv preprint arXiv:0907.0894*, 2009.

[24] V Levenshtein. Binary codes capable of correcting spurious insertions and deletions of ones. *Problems of Information Transmission*, 1(1):8–17, 1965.

[25] Hsi-Peng Lu, Chin-Lung Hsu, and Hsiu-Ying Hsu. An empirical study of the effect of perceived risk upon intention to use online applications. *Information Management & Computer Security*, 13(2):106–120, 2005.

[26] Robert McGill, John W Tukey, and Wayne A Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.

[27] Cristopher Moore and Mark EJ Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000.

[28] C C Mundt. Use of multiline cultivars and cultivar mixtures for disease management. *Annual Review of Phytopathology*, 40(1):381–410, September 2002.

[29] Andrew Newell, Daniel Obenshain, Thomas Tantillo, Cristina Nita-Rotaru, and Yair Amir. Increasing network resiliency by optimally assigning diverse variants to routing nodes. In *IEEE/IFIP International Conference on Dependable Systems and Networks*, pages 1–12. IEEE, June 2013.

[30] M Newman, Stephanie Forrest, and Justin Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101, September 2002.

[31] Mark Newman. *Networks: An Introduction*. Oxford University Press, Inc., May 2010.

[32] B Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos, and Christos Faloutsos. Virus Propagation on Time-Varying Networks: Theory and Immunization Algorithms. In *Machine Learning and Knowledge Discovery in Databases*, volume 6323, pages 99–114. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[33] Md Sazzadur Rahman, Ting-Kai Huang, Harsha V Madhyastha, and Michalis Faloutsos. Efficient and scalable socware detection in online social networks. In *USENIX Security*, 2012.

[34] Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *The Semantic Web-ISWC 2003*, pages 351–368. Springer, 2003.

[35] Matt Russell. Facebook scam involves money transfers to the philippines, `http://goo.gl/SMLDyh`.

[36] Eugene H Spafford. Computer viruses as artificial life. *Artificial Life*, 1(3):249–265, 1994.

[37] Charles Steinfield, Nicole B Ellison, and Cliff Lampe. Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, 29(6):434–445, November 2008.

[38] Hanghang Tong, B. Aditya Prakash, Charalampos Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, and Duen Horng Chau. On the Vulnerability of Large Graphs. In *ICDM*, pages 1091–1096. IEEE, December 2010.

[39] AJ Ullstrup. The impacts of the southern corn leaf blight epidemics of 1970-1971. *Annual Review of Phytopathology*, 10(1):37–50, 1972.

[40] U.S. Census Bureau. Genealogy Data: Frequently Occurring Surnames from Census 2000. `http://www.census.gov/genealogy/www/data/2000surnames/Top1000.xls`, 2000.

[41] Lorys M M A Villaréal and Christian Lannou. Selection for Increased Spore Efficacy by Host Genetic Background in a Wheat Powdery Mildew Population. *Phytopathology*, 90(12):1300–1306, December 2000.

[42] John Von Neumann and Arthur Walter Burks. *Theory of self-reproducing automata*. University of Illinois press Urbana, 1966.

[43] Huijuan Wang, Qian Li, Gregorio D'Agostino, Shlomo Havlin, H Eugene Stanley, and Piet Van Mieghem. Effect of the interconnected network structure on the epidemic threshold. *Physical Review E*, 88(2):022801, August 2013.

[44] P Wang, M C Gonzalez, C A Hidalgo, and A L Barabasi. Understanding the Spreading Patterns of Mobile Phone Viruses. *Science*, 324(5930):1071–1076, May 2009.

[45] Ryan West. The psychology of security. *Communications of the ACM*, 51(4):34–40, April 2008.

[46] X M Xu and M S Ridout. Stochastic simulation of the spread of race-specific and race-nonspecific aerial fungal pathogens in cultivar mixtures. *Plant Pathology*, 49(2):207–218, April 2000.