SAFER: Social Capital-Based Friend Recommendation to Defend against Phishing Attacks

Zhen Guo[1], Jin-Hee Cho[1], Ing-Ray Chen[1], Srijan Sengupta[2], Michin Hong[3], Tanushree Mitra[4]
[1]Department of Computer Science, Virginia Tech, VA, USA
[2]Department of Statistics, North Carolina State University, NC, USA
[3]School of Social Work, Indiana University, IN, USA
[4]Information School, University of Washington, WA, USA
zguo@vt.edu, jicho@vt.edu, irchen@vt.edu, ssengup2@ncsu.edu, hongmi@iupui.edu, tmitra@uw.edu

Abstract

The tremendous growth of social media has been accompanied by highly advanced online social network (OSN) technologies. Such advanced technologies have been heavily utilized by perpetrators as convenient tools for deceiving people in online worlds. Social capital has been discussed as a powerful mechanism to leverage interpersonal relationships in social networks in order for an individual to achieve his/her goal. The beauty of social capital is the ability to materialize non-monetary, less costly, and non-economic resources into tools to solve social problems. In this paper, we aim to leverage *social capital* (SC) to minimize online users' vulnerabilities to online deception. In particular, we propose a Social cApital-based FriEnd Recommendation scheme, called SAFER, that can protect OSN users from phishing attacks. We quantify three dimensions of social capital, namely, structural, cognitive, and relational, based on user features obtained from real datasets and model a user's friending behavior based on their social capital. In addition, to model a user's behavior upon being attacked by a phishing attacker, we developed the so-called SER-SEIR (Susceptible, Exposed, Recovered-Susceptible, Exposed, Infected, and Recovered) model as a variant of the SEIR model. Via extensive simulation experiments based on two real datasets considering bot-based and human-based attackers performing phishing attacks, we demonstrate the performance of four SC-based friend recommendation schemes with three non-SC-based comparable counterparts in terms of the ratio of detecting attackers and the fraction of users in the states of S, E, I, and R. Based on the performance comparison, we analyze the overall trends of their performance in terms of the extent of resistance against phishing attacks by bot or human attackers.

Introduction

Motivation

Highly advanced social media technologies have been heavily utilized by cyber attackers as convenient tools for deceiving people in online worlds (Guo et al. 2020). In particular, the significant growth of phishing campaigns, including broad spear-phishing campaigns and targeted attacks, have introduced serious cybersecurity challenges (van de Weijer, Leukfeldt, and Bernasco 2018). The phishing campaigns has a growth rate of 15% from 2019 to 2020 among all emails, and a significant number of them are pandemic-related (Warburton 2020). The phishing attackers often distribute fake URLs to steal sensitive information or financial credentials. In addition, they compromise real accounts and use them to launch attacks using the real users' accounts. The cybercriminals exploit advanced phishing strategies to cause the loss of confidentiality, privacy, credibility, and financial loss of victims, including organizations.

The concept of *social capital* (SC) has been extensively studied in the social sciences to understand how and why entities (e.g., individuals or organizations/communities) participate in social networks. Social capital is a powerful mechanism by which an entity utilizes its interpersonal relationships in social networks to achieve its goals (Putnam 2000). Social capital's positive effects have also been leveraged in the computing and engineering domains (Blanchard and Horan 2000; Venkatanathan et al. 2012; Alaa, Ahuja, and Schaar 2018; Phung et al. 2013; Jung et al. 2013). However, the concept of social capital has not been applied to combat phishing attacks in online social networks (OSNs).

People often unconsciously use social capital as promising signals to make good friends in OSNs, who may help in various ways, such as being connected with valuable resources (e.g., for marketing or job hunting) or having critical help (e.g., quickly finding missing people or collecting donations to help someone). However, a had friend in in one's social network may perform various types of online social attacks, such as sending phishing or scam messages aiming to obtain private information or monetary benefits. Hence, when people decide whether to accept a new friend invite, they often check a number of (mutual) friends, a number of posts, and a number of likes from the inviter's friends (Xu, Zhou, and Ma 2019), which partially explain social capital as an indicator to maintain trustworthy social connections.

The major vulnerabilities to OSN attacks are closely related to how users make friending decisions, e.g., whether to accept or decline a friend invite, or whether to send a friend invite (i.e., accept/decline or send a friend invite). A user's friending decision process determines the characteristics and quality of that user's social network in various types of OSN attacks (Guo et al. 2020). The majority of friend recommendation systems (FRSs) (Guo, Zhang, and Fang 2015; Huang et al. 2016; Ning, Dhelim, and Aung 2019; Wang

et al. 2015) have been proposed to recommend friends to maximize users' satisfaction in making friends in OSNs. However, no FRSs have used the multidimensional concept of social capital, in terms of relational, structural, and cognitive capital, to combat phishing attacks in OSNs.

In particular, we are motivated to use a user's features representing social capital as the basis of making good friends, who can help users defend against phishing attacks. First of all, we will use social capital as a signal to select good friends. Second, even if an attacker becomes a friend of legitimate users by performing social capital manipulation attacks (e.g., an attacker can increase its social capital through connections with many other users and active interactions with other users), the attacker can be caught by other legitimate users who are friends of the attacked users. Finally, the OSN system can suspend the reported attacker accounts by legitimate users or legitimate friend users. We name this approach by Social cApital-based FriEnd Recommendation, called SAFER.

Our proposed SAFER can be applied on real OSN platforms like Twitter or Facebook. When user i receives a friend request from user j, SAFER can provide j's social capital in three dimensions, including structural, cognitive, and relational social capital (Nahapiet and Ghoshal 1998). We provide the details of SAFER in Section . The source codes and processed features are available on Github[1].

Research Questions

This study aims to answer the following research questions:

1. How can a user's social capital be quantified in its three key dimensions (i.e., relational, cognitive, and structural capital) based on the user's network and behavioral characteristics?
2. What dimension of social capital contributes more (or less) to defending against phishing attacks?
3. How does an attacker's type, such as attacks by humans or bots, affect the FRS's capability to defend against phishing attacks differently?

Key Contributions

In this paper, we made the following key contributions:

- To the best of our knowledge, our work is the first that leverages the concept of multidimensional social capital in order to model its defense capability against phishing attacks and evaluate its effectiveness.
- We quantify multiple dimensions of a user's social capital (i.e., structural, cognitive, and relational capital) based on the user's behavioral characteristics in OSN contexts derived from two real Twitter datasets where attackers are bots (Cresci et al. 2015) or humans (Yang et al. 2012; Yang, Harkreader, and Gu 2011), respectively.
- We develop a social capital-based friend recommendation system, called SAFER, and compare its performance with that of existing counterparts, such as social attribute-based (Guo, Zhang, and Fang 2015), topic-based (Wang et al. 2015), and trust-based (Cho, Alsmadi, and Xu 2016).

In addition, we conduct simulation-based extensive experiments to identify the key dimension of social capital contributing significantly to combating phishing attacks.

[1]https://github.com/zguo1010/osd/tree/main/social_capital

Background & Related Work

Concepts and Applications of Social Capital. The common role of social capital has been agreed as 'a vehicle to facilitate achieving individual or collective goals through personal relationships in social networks'. Although there have been many classifications of social capital (Putnam 2000), its most common concept is discussed in terms of bonding and bridging (Burt 2000). Bonding refers to being connected with people one can trust, while bridging is connecting with more people (Burt 2000). However, there have been more discussions on the concept of social capital in a broader sense. Nahapiet and Ghoshal introduced three key dimensions of social capital: structural, cognitive, and relational. *Structural capital* (STC) refers to the capital derived from social structure (e.g., network ties and configuration, roles, rules, precedents, and procedures). *Cognitive capital* (CC) indicates the benefit derived from shared understandings (e.g., shared language, codes, and narratives, shared value, attitudes, and beliefs). *Relational capital* (RC) is obtained from nature and indicates the quality of relationships (e.g., trust and trustworthiness, norms and sanctions, obligations and expectations, or identity and identification).

The benefits of social capital have been discussed in individual capital (e.g., personal or micro) or collective capital (e.g., society or macro) (Putnam 2000; Yang 2007).The concept of social capital has been used for increasing virtual team productivity (Blanchard and Horan 2000), establishing 'bridging social capital' in OSNs (Venkatanathan et al. 2012), investigating mathematical models of online social activities (Alaa, Ahuja, and Schaar 2018), or examining its effect on mental health (Phung et al. 2013). However, to the best of our knowledge, social capital has not been applied to combat online deception, such as phishing attacks. In addition, no prior work has quantified three-dimensional social capital as a quantitative metric, although social science literature has extensively discussed it conceptually or measured it qualitatively to some extent (Nahapiet and Ghoshal 1998).

Social capital has been measured quantitatively in social sciences by conducting empirical studies with human subjects (Kim et al. 2021; Marcaletti and Oldani 2021) or under laboratory environments (Karlan 2005). Network scientists also measured structural social capital based on network topologies (Andersson 2021; Tsai 2021). Further, social capital in online platforms has been measured on Facebook and Twitter in the computer science domains (Phua, Jin, and Kim 2017; Cho, Alsmadi, and Xu 2016; Ye, Ho, and Zerbe 2021). However, none of them measured three-dimensional social capital nor employed it as a defense mechanism to handle phishing attacks. The key features in consideration for three-dimensional social capital are summarized in Table 1. Those features are also used in other works, showing the validity of those features in measuring social capital.

Friend Recommendation Systems (FRSs). An FRS has been extensively studied for matching similar friends.

	Structural capital	Cognitive capital	Relational capital
Description	Social structure (bridging)	Shared understandings	Quality of relationships (bonding)
Impact level	Collective (Macro)	Collective & Individual (Mezo)	Individual (Micro)
Measurements	Node/graph centrality metrics	Positive shared experiences	Trust, reputation, and homophily

Table 1: Key three dimensions of social capital and their key characteristics

A personality-trait-based FRS, called PersoNet, was proposed (Ning, Dhelim, and Aung 2019), and a similarity metric was considered to weigh the contribution from both Big Five personality traits similarity and harmony rating similarity. Their online experiment tested the friend recommendation accuracy of a small set of selected users for one month. A two-stage FRS was proposed by Huang et al. (2017), which generated a friend list in the first stage by the network alignment algorithm of a contact network and tag-similarity network. In the second stage, topic features from Flickr image information were used in a probabilistic topic model.

Wang et al. (2015) discussed life style events extraction as topics and words analysis by latent Dirichlet allocation (LDA) topic model and clustered by K-means. The similarity metric considered both of the topic vectors and the dominant topics based on an edge weight in a friend-matching graph. Zheng, Song, and Bao (2015) analyzed a user's temporal behaviors by a topic model to predict potential friends. A user's temporal similarity was calculated at different time intervals by the topics where the friend recommendation value was estimated based on the aggregation of the temporal similarity with a time decay function, addressing a more weight on the most recent similarity.

Cheng et al. (2018) and Guo, Zhang, and Fang (2015) focused on the privacy-preserving property of friend recommendation in the social network in terms of encryption and decryption directed by a central authority. The similarity between users was based on tag matching (Cheng et al. 2018), while the cosine similarity was estimated based on social attributes and trust level in a multi-hop friend recommendation chain (Guo, Zhang, and Fang 2015). Cho, Alsmadi, and Xu (2016) used the concept of social capital to investigate a tradeoff between social capital and privacy preserving. However, to the best of our knowledge, no prior work has investigated users' online social capital to combat phishing attacks.

Machine learning-based FRSs are also studied in the literature. Chen, Shih, and Lee (2016) used the preferences and behaviors of informative friends as key features and ranked them based on gradient descent to match those features to the preference of a target user. Ding et al. (2017) also proposed a model called *BayDNN* based on structural features and used a Bayesian ranking to recommend new friends. Chen et al. (2020) used a convolutional neural network (CNN) for user embedding in a graph convolutional network to process users and their neighbors' features. They recommended new friends based on the Bayesian ranking. However, our approach is different from ML-based approaches because its goal is not to predict an attacker accurately but to guide users to form their own social capital features and accordingly build a safe and trustworthy online social network against phishing attacks.

Preliminaries

An online social network (OSN) is defined by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{T})$ where $\mathcal{V}$ refers to a set of vertices, v_i's, representing users i's and $\mathcal{E}$ is a set of edges e_{ij} for users i and j, indicating users i and j are friends to each other. $\mathcal{T}$ is a set of edge weights T_{ij}, referring to user i's trust in user j (detailed in Section). How users are connected to each other is reflected in $\mathcal{E}$ and $\mathcal{T}$.

User Behavior Modeling

In this section, we discuss how each user's behavior in a given OSN is modeled based on the features obtained from the two real datasets (i.e., Cresci15 (Cresci et al. 2015) and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012)). Note that a user's friend list is not available due to the privacy issue. In addition, the goal of this work is to investigate how social capital-based friending decisions can ultimately help defend against phishing attacks. Therefore, instead of deriving user behaviors and their friending decisions in the real datasets, which is not feasible *per se*, we model a user's behaviors based on the behavior trends observed from the datasets, as commonly used in the simulation and modeling research (Cho et al. 2019; Gatti et al. 2014). A user in a given OSN is modeled by the following characteristics, which are used to dynamically interact with other users in our simulation model.

Feeding Information. A user can feed information to other users by: (i) *posting his/her own information* (e.g., opinions/thoughts, knowledge in professional domains, preferences, activities); (ii) *sharing a third-party's information* (e.g., news articles, posts by others, information from other third-party sources, such as blogs, websites); or (iii) *adding his/her own opinion* to a third-party's information. The information itself can be texts, images/photos, and/or videos. In this work, we derive a user's feeding behavior as behavioral seed probabilities from real social media datasets (Yang, Harkreader, and Gu 2011; Yang et al. 2012; Cresci et al. 2015). For example, based on the frequency of posting or sharing information, the probability of a user feeding information is calculated and used to exhibit his/her feeding behavior in our model (e.g., tweets in Twitter).

Providing Feedback. A user can show his/her preferences or express his/her opinions towards the posts by his/her friends (e.g., 'Likes' in Facebook or 'Favorites' in Twitter). In addition, the user can leave comments on his/her friends' posts or shared information. The frequency of leaving comments, 'Likes' or 'Favorites' can measure how often the user provides feedback to other users.

Inviting Friends. Depending on a user's propensity to make friends, the frequency of inviting friends can be different.

Metric	Features
h_i for STC	Account longevity, # of self-introduction words
cc_i for CC	Average # of retweets per post, average # of hashtags per post, average # of mentions per post, average # of URLs per post, # of followers
rc_i for RC	Average # of tweets per day ($\mathrm{post\_freq}$), average # of replies per day ($\mathrm{com\_freq}$), total # of favorite tweets (favorite frequency)
Friend network	Friend count ($\mathrm{friend\_num}$), a user's friends, followers count
T_{ij}	Average # of tweets per day ($\mathrm{post\_freq}$), average # of replies per day ($\mathrm{com\_freq}$)

Table 2: Features used to quantify dimensions of social capital, trust, and friend network in the Cresci15 (Cresci et al. 2015) and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012) datasets.

Based on real datasets that provide the number of friends per user, we derive the probability that a user invites a friend based on the number of friends the user has over the total number of users. However, it is difficult to know what types of friends a user prefers to be a friend with because users' privacy settings restrict the availability of datasets on friend-friend relationships. Hence, we evaluated a set of FRSs, including our proposed SAFER in Section , which governs the friend relationships.

Capability to Detect Phishing Attacks. We model user i's probability to detect attacker j based on the following factors: (i) an individual user's competence (c_i) to detect source credibility based on the number of followers. Based on Westerman, Spence, and Van Der Heide (2012), a user can best detect source credibility when the number of followers is not too high or not too small. To reflect this, we modeled the degree of a user's capability to judge source credibility by $c_i = -\lambda(f_i - f_{\mathrm{median}})^2 + 3$, in the range $[0, 3]$. For $c_i \geq 0$, λ is a constant to adjust the range of observed number of followers in all users of a given OSN; (ii) depending on the quality of deception skills a phishing attacker j uses, it may be easier or harder for user i to detect attacker j. We denote d_j as the degree of deception quality in a given phishing message by attacker j (see the detail on how d_j is modeled in Section); and (iii) how many times user i has experienced phishing attacks, denoted by g_i, also affects user i's ability to detect attackers. Considering these three factors, we model user i's probability to detect attacker j by:

$$P_{ij}^{\mathrm{crd}} = e^{-\frac{d_j}{(c_i \cdot g_i)}}. \quad (1)$$

Behavioral Features Related to Social Capital. In this work, we estimate a user's social capital (SC) in terms of structural capital (STC), cognitive capital (CC), and relational capital (RC). According to Nahapiet and Ghoshal (1998), we collected features to measure each dimension of social capital from two real datasets, Cresci15 (Cresci et al. 2015) and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012). The Cresci15 dataset has features of bots attackers while the 1KS-10KN dataset has features of humans attackers along with the majority of legitimate users' features. Features used to measure each dimension of social capital are described as follows:

- *Structural capital* (STC): User i's STC is measured based on the extent of human capital of user i's friends. Hence, a user's STC is mainly affected by (i) how many friends the user has; and (ii) the extent of human capital the user's friends have where the human capital is measured by the friend's individual capability. Due to the limited user features identifiable in various social media platforms (e.g., Facebook or Twitter), a user's human capital is measured based on the longevity of the user's account (i.e., the user's stable activity and/or relations with other users) and the length of self-introduction (i.e., how well the user self-introduces himself/herself) where these two features are identified as key features representing trustworthy users in the literature (Badri Satya et al. 2016; Inuwa-Dutse, Liptrott, and Korkontzelos 2018).
- *Cognitive capital* (CC): User i's CC is measured based on how much the friends support or like his/her posts or activities, such as: (i) the amount of feedback received (e.g., the average number of shares or retweets per post, likes per posts, or replies per post); (ii) the broadness of interest in posts (e.g., hashtags/URLs/mentions per post), and (iii) the number of followers.
- *Relational capital* (RC): User i's RC is measured based on how actively user i interacts with other users. The main activities to represent a user's RC are: (i) The total amount of posts; (ii) the frequency of posting, $\mathrm{post\_freq}$ (e.g., the number of uploading tweets per day); or (iii) the frequency of providing feedback, $\mathrm{com\_freq}$ (e.g., the number of comments, replies, likes, or favorites per day or month).

Table 2 summarized the features used to estimate the three social capitals from the real datasets (Yang, Harkreader, and Gu 2011; Yang et al. 2012; Cresci et al. 2015).

Adversary Model

A phishing attacker may trick users into revealing sensitive, private, or confidential information related to work, financial credentials, or even personal data in fraudulent activities, leading to the loss of confidentiality and/or privacy. We modeled random phishing attacks by an attacker simply selecting a set of users among its friends (i.e., its friend network) at random and disseminating phishing/scam messages. We assume that the attacker may send various types of phishing messages applying different levels of deception quality. A legitimate user may not be able to easily detect a phishing message which has a high quality of deception and *vice versa*. To model this, attacker j will apply a different level of deception quality, ranged in $[0, 3]$ as a real number, respectively, denoted by d_j. We model each attacker's deception quality following Gaussian distribution with a given mean and standard deviation. When a legitimate user receives the phishing message with deception level d_j from attacker j, the user's ability to detect this attack is estimated based on Eq. (1), implying that higher d_j decreases the user's attack detection ability. Furthermore, the attacker may send the same type of phishing messages to other users.

The attacker randomly selects new friends and invites

them to accept his/her friend invites. When the attacker receives any friend invite, it will always accept to maximize its influence with more friends in OSNs, leading to increasing its social capital. In addition, the attacker can invite many friends to increase his/her own social capital, leading other legitimate users to invite the attacker more frequently. Further, the attacker can actively interact with his/her friends by sharing, commenting, or posting information more frequently. These kinds of attack behaviors are to manipulate attackers' social capital. Further, the attacker may compromise legitimate users' accounts and let the compromised accounts perform phishing attacks, ruining the legitimate user's reputation. Those attacks may lower the user's social capital because some friends may terminate relationships with them when detecting their phishing attacks.

A legitimate user can make friends through FRSs. Now we discuss our proposed social-capital based FRS, namely SAFER, as detailed below.

SAFER

This section provides the detail of the proposed SAFER (Social cApital-based FriEnd Recommendation) scheme. We first describe our proposed multidimensional social capital metric. In addition, we provide the details of how social adversaries are detected in the SAFER and how a user's status is updated based on the proposed SER-SEIR (Susceptible, Exposed, Recovered-Susceptible, Exposed, Infected, and Recovered) model.

Quantification of Social Capital

We already described what features are considered to measure the three dimensions of social capital in 'Behavioral Features Related to Social Capital' of Section . In this section, we will focus on mathematically formulating each dimension of social capital as a metric used in this work.

A user's structural capital, STC_i, measures how well user i is connected to other users with high human capital. We denote node i's human capital by h_i. A user's cognitive capital, CC_i, measures how much user i's interactions with other users are supported (preferred) by his/her social network community. A user's relational capital, RC_i, is measured based on how actively user i interacts with other users (i.e., supports or activities provided to other users).

Finally, we measure user i's social capital, SC_i, by:

$$SC_i = w_{STC} \cdot STC_i + w_{CC} \cdot CC_i + w_{RC} \cdot RC_i, \quad (2)$$

where CC_i is user i's cognitive capital, which is measured based on the extent of the support the friends of user i have received from their friends. w_X is a weight to consider each social capital where $X = STC$ (structural capital), CC (cognitive capital), or RC (relational capital). Given cc_j representing the extent of the support friend user j received from its friend network, user i's cognitive capital, CC_i, is computed by:

$$CC_i = \frac{1}{|\mathrm{F}_i|} \sum_{j \in \mathrm{F}_i} T_{ij} \cdot cc_j, \quad (3)$$

where T_{ij} refers to user i's trust in user j. We discuss T_{ij} later in this section with Eqs. (6) and (7). F_i is the set of user i's friends. CC_i is scaled as a real number in $[0, 3]$ because cc_j considers three features, with each normalized as a real number in $[0, 1]$ based on Table 2. STC_i is estimated from the human capital of i's friends by:

$$STC_i = \frac{1}{|\mathrm{F}_i|} \sum_{j \in \mathrm{F}_i} T_{ij} \cdot h_j, \quad (4)$$

where h_j refers to user j's human capital described in Section and is derived based on three features, each being normalized as a real number in $[0, 3]$. Given rc_j indicating how actively friend user j feeds information and/or provides feedback in its social network (see Table 2), user i's relational capital, RC_i, is estimated by:

$$RC_i = \frac{1}{|\mathrm{F}_i|} \sum_{j \in \mathrm{F}_i} T_{ij} \cdot rc_j, \quad (5)$$

where RC_i is scaled as a real number in $[0, 3]$ as we consider three features in rc_j. Hence, SC_i is ranged in $[0, 3]$ as a real number based on these three components of social capital described in Table 2.

T_{ij} is user i's trust in user j (Cho, Alsmadi, and Xu 2016). A user's trust is generally estimated by sharing or feeding information and providing feedback such as liking or commenting, as described in Section . T_{ij} is estimated by:

$$T_{ij} = \sum_{x \in X} t_x \cdot T_{ij}^x, \quad (6)$$

where X is trust dimensions indicating feeding and feedback behaviors in this work. t_x is the weight for trust in x where $\sum_{x \in X} t_x = 1$. T_{ij}^x is calculated based on the number of positive interactions I_{ij}^x between user i and user j. The equation of T_{ij}^x is:

$$T_{ij}^x = \frac{I_{ij}^x}{\max(I_{ik}^x \text{ for } k \in \mathrm{F}_i)}, \quad (7)$$

where F_i is a set of user i's friends. The positive interactions are defined as any feeding and feedback activities between legitimate users i and j. However, when user i is an attacker sending a phishing message to user j, if user j cannot detect it, we treat it as a positive experience. However, if user j detects the attack, it is treated as a negative experience for the attacker i. For the details of what features measure h_i, cc_i, rc_i, T_{ij}, and a friend network (F_i), please refer to Table 2.

Phishing Attack Detection

When attacker j sends out a fraud message to its friend, a legitimate user i, this friend user i can detect attacker j based on P_{ij}^{crd}, which is discussed in Eq. (1). Legitimate users in one's OSN can help their friends detect phishing attacks. If user i's friend, k, can detect a phishing message by attacker j using P_{kj}^{crd}, then user i is immune to the attack or can be recovered from the attack with the help of user k. Then, user k's can increase their g_k's (i.e., the number of phishing attack experiences) like user i and use the indirect experience

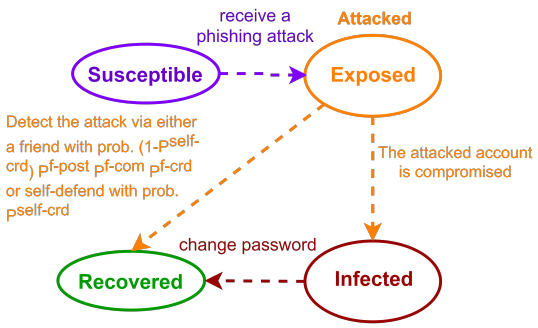


Figure 1: A user's status update in the SER-SEIR model.

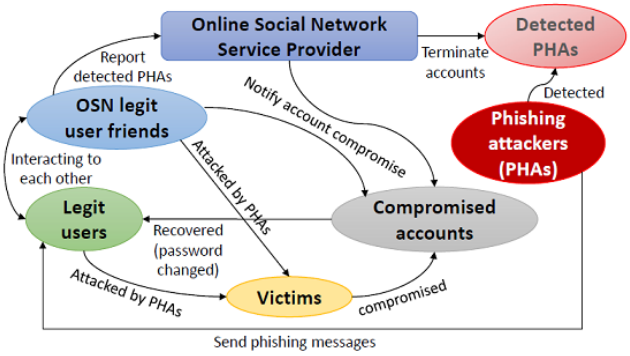


Figure 2: The overview of the proposed SAFER scheme.

to detect similar phishing attacks. This experience can naturally increase the probability that user i's friends can detect similar future attacks.

If legitimate user i receives a phishing message from someone j in his/her friend network but is unsure of whether the received message is from a phishing attacker based on his judgement ability in information credibility, P_{ij}^{crd} (see Eq. (1)), user i will share his/her information based on his/her posting frequency probability, $P_i^{\text{post}} = \text{post\_freq}_i$. If user i decides not to share it, he/she will not have a chance to get helped by his/her friends. When friend k sees the posting of user i on the phishing attack, he/she can help by leaving a comment based on $P_k^{\text{com}} = \text{com\_freq}_k$. If friend k can correctly detect the phishing message posted by user i, the attack from attacker j can be successfully detected with the probability of $P_i^{\text{post}} \cdot P_k^{\text{com}} \cdot P_{kj}^{\text{crd}}$. Otherwise, user i fails to detect the phishing attack from attacker j. This implies that as long as user i is active enough to reach out to other friends, who have a sufficient level of willingness and competence to help, user i's friends can be good assets to protect the user from phishing attacks, which is aligned with the core concept of social capital in terms of accessibility to resources (Portes 1998).

If more than three users detect an attacker and report it to the OSN provider, the attacker's account will be suspended, and all his/her social connections will be removed from this network. If a legitimate user's account is compromised by an attacker (with the probability P_{cp}), the compromised account can also propagate phishing messages to his/her friends. If the compromised account is detected, the OSN provider will require the original owner of the account to reset his/her password. After the password reset, this user can be back in R, but the original user of the compromised account may lose some extent of social capital because some friends detect the user as an attacker and terminate their friend relationships with the compromised account.

Updating a User's Status

The behaviors and activities of attacks and defenses are measured by the state update using our proposed SER-SEIR (Susceptible, Exposed, Recovered-Susceptible, Exposed, Infected, and Recovered) model, which is extended from (Li and Muldowney 1995), as depicted in Figure 1. The proposed SER-SEIR model allows users to move from S to E when the user receives phishing attacks. Then, the user can self-defend or defend against it with the help of friends, which allows the user to move from E to R directly without going through I. However, if the user cannot defend against the attack, its account is compromised by the attacker, moving from E to I. If the user in I changes the password of his/her account, it can move from I to R. A user can only interact with his/her friends. Each state is described as:

- A legitimate user, who has not experienced a given attack, is in S (Susceptible).
- A legitimate user is in E (Exposed) if the user receives a phishing attack and the user fails to detect an attack when: (i) The user received an attack but could not detect it without asking other friends for help; (ii) The user received an attack but could not detect it correctly even if the user asked other friends for help.
- A legitimate user is in I (Infected) if the user account that did not detect the attack is compromised and manipulated by the attacker. An attacker can infiltrate a legitimate user in I and transform this user account to a compromised account with the probability, P_{cp}.
- A legitimate user is in R (Recovered) if the user successfully detects an attack by: (i) Detecting the attacker correctly with his/her own detection ability, $P^{\text{self-crd}}$ (see Eq. (1)); (ii) If his/her friend(s) can help and correctly detect the attack when the user shares the incident with his/her friends as he/she cannot send-defend, with the probability $(1 - P^{\text{self-crd}})P^{\text{post}}P^{\text{com}}P^{\text{friend-crd}}$; or (iii) The compromised account is reported by at least three friends and the OSN provider asks the legitimate user to change his/her password.

We summarized the overview of the proposed SAFER scheme in Figure 2.

Experimental Setup

Datasets. We use two real datasets from Twitter: Cresci15 (Cresci et al. 2015) with bot attackers and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012) with human attackers. Each dataset is detailed as follows:

- *Cresci15* (Cresci et al. 2015): This dataset contains 1,481 normal Twitter accounts from a sociological study, 469

Parameter	Value	Parameter	Value
N	1,946	P_{AS}	0.1
P_a	20% (389 attackers)	w_{STC}, w_{CC}, w_{RC}	1/3
N_{sim}	100	t_x	0.5
λ for Cresci15	0.045	λ for IKS-10KN	0.0015
f_{median} for Cresci15	7.68	f_{median} for 1KS-10KN	37.18
P_{cp}	0.2	d_j	$N(1.5, 0.3)$

Table 3: Design Parameters and Their Default Values

certified humans accounts, and 845 fake bot accounts bought from online markets. After combining the three datasets and filtering out the accounts that posted no tweets, the network size is 2,664. There are 1,946 legitimate users with 398 friends and 718 attackers with 554 friends on average.

- *1KS-10KN* (Yang, Harkreader, and Gu 2011; Yang et al. 2012): This dataset is comprised of 10,000 legitimate users and 1,000 human attackers accounts crawled from Twitter. The spammers are accounts that are identified as posting malicious URLs. For meaningful result analysis, we filtered out the accounts that posted zero tweets and finally obtained the network with 10,766 users, comprising 9,766 legitimate users with 7,744 friends and 1,000 spammers with 2,520 friends on average.

Default Parameterization. The default attacker ratio is set to 20% with $P_a = 0.2$. We considered 1,946 legitimate users and 389 attackers to consider the same network size for both datasets. When the attackers perform attacks, the fraction of targeted victims from legitimate users is set to $P_{AS} = 0.1$ by default. The experimental results are based on the average data points from 100 simulation runs (N_{sim}). We used the ratios of susceptible, infected, and recovered users over the total number of legitimate users (see Sections) as the key metrics to evaluate the friending decision schemes. The default values of key parameters used in our experiment are summarized in Table 3.

Experiment Procedures. We take the following steps:

- Each user is assigned a set of features based on the normalized features in Table 2.
- An attacker will perform attacks where its deception quality is randomly selected as a real number ranging in $[0, 3]$ based on Gaussian distribution with 1.5 for mean and 0.3 for a standard deviation (i.e., $N(1, 5, 0.3)$).
- Given a set of behavioral seed probabilities for each user, including behaviors (i.e., feeding information, providing feedback, inviting friends, and judging credible information) modeled in Section , each user starts making friends and interacts with other users. We allow 100 times interaction steps (i.e., chances to interact with other users such as whether to invite a new friend or post/share information). Each user will exhibit his/her behavior based on the derived behavioral seed probabilities.
- Each user will start with friends selected based on individual features contributing to its social capital. Recall that a user's social capital is estimated based on the friends' social capital. When all users are not connected at all, an individual user's social network is started with himself/herself alone. However, since the features representing a different type of social capital are different, we can allow the estimation of a user's social capital without any friends based on cc_i, h_j, and rc_j in Eqs. (3), (4), and (5).
- For other non-social capital-based FRSs, we will initialize a user's features and select friends based on the given features. For example, for Trust-based FRS (TR), each user is initialized based on the average sum of posting and commenting behaviors (i.e., $T_i = w_{\text{post}} \cdot P_i^{\text{post}} + w_{\text{com}} \cdot P_i^{\text{com}}$) where $w_{\text{post}} + w_{\text{com}} = 1$. We weigh each trust component equally in our case study. We let a user select his/her first friend based on social attributes and topics of interest, respectively, for Social Attributes-based FRS (SA) and Topic model-based FRS (TM).
- Upon every interaction chance, calculate each user i's h_i, rc_i, cc_i, and T_i, respectively.
- Make friending decisions based on FRSs (see 'Comparing Schemes'), where each scheme needs the current interaction states in estimating the criteria features, such as social capital, trust, topic features, or social attributes.
- From the 101-st interaction step to 105-th step (i.e., 5 interaction steps), perform the one-time attack per step where each potential victim can respond based on his/her behavioral characteristics associated with social capital as described in Section . Update each node's status based on the proposed SER-SEIR model in Figure 1. The one-time phishing attack is explained in Section . We allow each attacker to apply its corresponding attack to a normal user friend who is currently in the state of S or E.
- Repeat the attacks from each attacker for 101-st to 105-th interaction step.
- Calculate $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}$, as described in the 'Metrics' of this section below, with seven FRSs under phishing attacks performed by either human or bot attackers based on the two datasets (Yang, Harkreader, and Gu 2011; Yang et al. 2012; Cresci et al. 2015).

Comparing Schemes. We will compare the following FRSs that determine how each user connects to other users where each user only knows the number of friends (i.e., friend_num) based on the used two datasets. Note that we allow each user to select five friends at the very beginning of the network deployment and use the corresponding FRS to create a user's social network.

- *RC-based FRS* (RC): A user selects new friends based on the top friend_num number of users using RC_i in Eq. (5).
- *STC-based FRS* (STC): A user selects new friends based on the top friend_num of users using STC_i in Eq. (4).
- *CC-based FRS* (CC): A user selects new friends based on the top friend_num number of users using CC_i in Eq. (3).
- *Multidimensional SC-based FRS* (MSC): A user selects new friends based on the top friend_num number of users using SC_i in Eq. (2).
- *Social Attributes-based FRS* (SA): A user selects new friends based on similar shared attributes (Guo, Zhang, and Fang 2015). The shared attributes are defined by a

vector of h_i, cc_i and stc_i. The top $\mathrm{friend\_num}$ number of users with the highest cosine similarity score of social attributes will be selected as new friends.

- *Topic model-based FRS* (TM): A user selects new friends based on the top $\mathrm{friend\_num}$ number of users with the highest topic similarity score (Wang et al. 2015). Under Cresci15 and 1KS-10KN, each user has a document of all of his posts. All the documents are processed by the LDA algorithm to generate the top 20 topics, along with the probabilities. The similarity between two users is defined by the cosine similarity of 20 topics probability scores.
- *Trust-based FRS* (TR): A user selects new friends based on the highest trust (Cho, Alsmadi, and Xu 2016). A user's trust is simply estimated based on how much other friends trust a given user by $T_i = \frac{1}{|\mathrm{F}_i|}\sum_{j\in \mathrm{F}_i} T_{ji}$. A user selects new friends based on the top $\mathrm{friend\_num}$ number of users with the highest trust.

Upon receiving a friend invite, a user will accept it as long as its key qualification in each FRS is no less than his/her own. For example, if the user uses MSC, it will accept the invite if the inviter has no less than MSC the user has.

Metrics. We use the ratios of $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}$ to evaluate each FRS. Note that the lower $\mathcal{E}$, $\mathcal{I}$ and the higher $\mathcal{R}$ represent higher robustness of a given FRS against phishing attacks.

Simulation Results & Analysis

Probability Distributions of Relational, Cognitive, and Human Capital

In this section, we discuss how the social capital in three dimensions (i.e., relational, cognitive, and structural social capital) can be differently identified for attackers and legitimate users. Figure 3 shows the histograms of social capital dimensions, including relational capital, cognitive capital, and human capital, for attackers and legitimate users, where the distributions are identified based on the best-fit probability density functions. Each social capital corresponds to rc_i, cc_i, or h_i. Here we note that because structural social capital is not known (it is to be derived after a user determines his/her social network structure based on his/her friend network features), we capture it by an individual user's human capital, h_i, which is the key input to estimate the structural capital as in Eq. (4).

In Figure 3, the fitting process applies 86 probability density functions using `scipy.stats` packages and picks the one that gives the minimum errors. From the two datasets (i.e., Cresci15, 1KS-10KN), legitimate users have higher social capital values, as shown in Figure 3. The pattern of human capital in the two datasets are similar, while the distributions of relational and cognitive capital are divergent in each dataset. All the spammer accounts under Cresci15 are bots bought from online markets. The spammer accounts under 1KS-10KN are human accounts crawled from random seeds where the human spammers posted malicious URLs.

Ratio of Reported Attackers

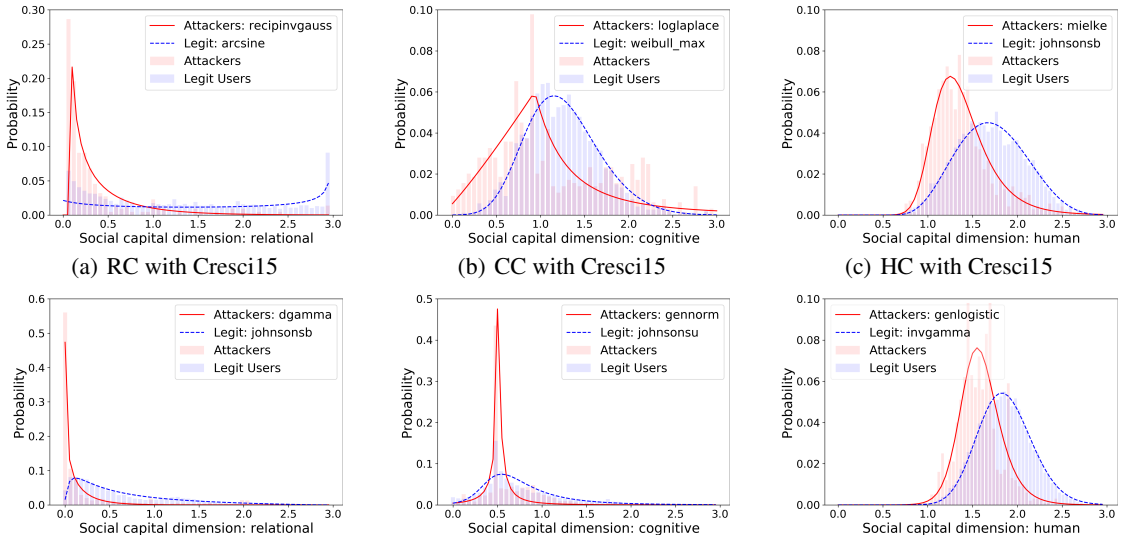
Figure 4 shows the ratio of attackers, whom at least three users detect, and the OSN provider suspends their accounts and cuts the friend connections. The reported attackers cannot deliver phishing attacks at later attack times. The four SC-based FRSs have more reported attackers than the non-SC-based schemes SA and TM for both Cresci15 and 1KS-10KN. However, Figure 4(a) and Figure 4(b) indicate that the growth rate of reporting attackers with more attack times in Cresci15 is much higher than in 1KS-10KN for four SC-based FRS. Reported attackers of TP are lower than the four SC-based FRS in Cresci15, but those of TP have the same scale with SC-based FRS in 1KS10KN. In Figure 4(a), SC-based FRSs can report 50% of attackers during the first attack time, and they can almost report all the attackers from three attack times. The three baseline approaches report around 10-20% attackers in the first attack and then gradually grow during the five attack times. In Figure 4(b), all the seven schemes report less than 30% of attackers from the first attack and then grow slowly with increasing attack times. These findings suggest that the SC-based FRSs are highly effective in detecting bot-related attackers even in the beginning of phishing attacks. Under both networks, SC-based FRSs effectively defend against phishing attacks better than the other three baseline FRSs.

Figures 4(c) and 4(d) show the effect of varying the fraction of attackers when the number of attack is 1 in terms of the ratio of reported attackers. We observe that the detection ratio decreases particularly under non-SC based FRSs due to less power to deal with attacks upon the increase of the attackers. This implies that the SC-related FRSs outperform non-SC-based counterparts in defending against phishing attacks. In Figure 4(d), under human attackers, the trend of baseline methods, SA and TM, is similar to that under Cresci15 dataset. However, the four SC-related FRSs have lower report ratio than in Cresci5, and trust-based FRS has a similar performance compared to four SC-related FRSs. This pattern reveals that the defense capability against phishing attackers by SC-related and trust-based FRSs is insensitive to more numbers of attackers in the social network, in contrast to the trends observed under SA and TM. Among the SC-related FRSs, CC and RC show the best performance in defending against human attackers.

Comparative Performance Analysis Under Varying the Frequency of Attacks

Figure 5 demonstrates the performance of the seven FRSs in terms of the effect of phishing attacks on $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}$ when the attack strength varies from one-time attack to five-time attacks performed by each attacker. We set the percentage of attackers (P_a) to 20% among all users in the network. By increasing the attack frequency, the two networks under Cresci15 and 1KS-10KN have more recovered users ($\mathcal{R}$) and less susceptible users ($\mathcal{S}$). However, during the five-times phishing attacks, $\mathcal{E}$ and $\mathcal{I}$ reached a maximum point for some FRSs and then started decreasing, as shown for all four SC-based FRSs and TR in Figures 5(c), 5(f), and 5(g).

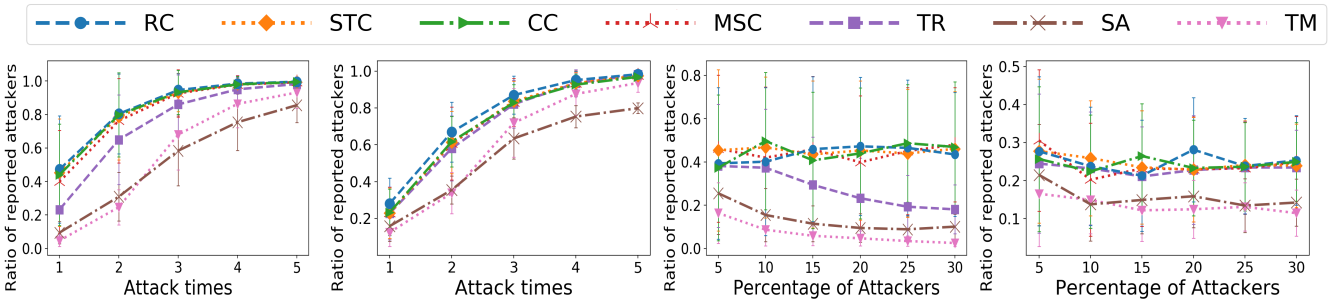
In Figures 5(a) and 5(d), under Cresci15 with bot attackers, we can clearly observe lower $\mathcal{S}$ and higher $\mathcal{R}$ in SC-based approaches (i.e., RC, CC, STC, and MSC) than three baseline methods (i.e., TR, SA, and TM). Under 1KS-10KN with human attackers in Figures 5(e) and 5(h), the perfor-



(a) RC with Cresci15 (b) CC with Cresci15 (c) HC with Cresci15

(d) RC with 1KS-10KN (e) CC with 1KS-10KN (f) HC with 1KS-10KN

Figure 3: Probability distributions of relational, cognitive, and human capital (rc_i, cc_i and h_i), namely RC, CC, and HC, using the datasets of Cresci15 (Cresci et al. 2015) (see (a)-(c)) and the 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012) (see (d)-(f)), respectively. The solid red lines and dotted blue lines show the best-fit probability density function curve for attackers and legitimate users, respectively. The red and blue bars are the histograms for attackers and legitimate users.



(a) Varying N_A under Cresci15 (b) Varying N_A under 1KS-10KN (c) Varying P_a under Cresci15 (d) Varying P_a under 1KS-10KN

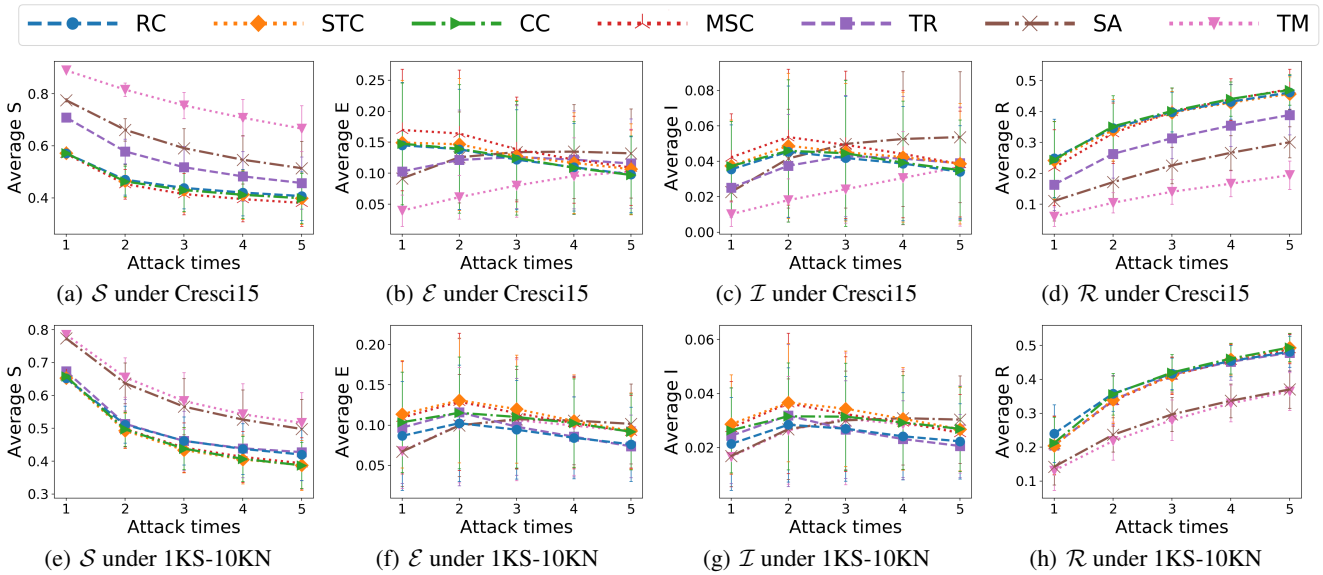
Figure 4: Comparison of the FRS schemes in terms of the ratio of detected attackers when varying the number of attacks (N_A) by a same attacker and the fraction of attackers (P_a).

mance gaps between SC-based and non-SC-based schemes are less than Cresci15. However, TR has comparable performance with SC-based FRSs for $\mathcal{R}$ and SC-based schemes still outperform non-SC-based counterparts, SA and TM.

Under 1KS-10KN with human attackers, unlike what we observed in Cresci15, which has bot attackers, the degrees of human attackers' social capital are less distinct than those of legitimate users. Due to this reason, it seems more human attackers are able to penetrate into friend networks of legitimate users (being friends of more legitimate users). In addition, the network in 1KS-10KN with the mean degree of about 73 is much denser than the network in Cresci15 with the mean degree of about 36. Higher network density is more likely to make each user's social capital less distinct due to more users with many friends, leading to less distinctive characteristics of users' friend networks. Therefore, compared to the results under Cresci15, the results under 1KS-10KN are less sensitive to varying the attack strength. However, we can still observe SC-based and TR-based (which is highly similar to RC-based in dense networks) FRSs perform better than SA and TM due to high attack detection by friend users of a user with high social capital, as shown in Figures 4(c) and 4(d).

Comparative Performance Analysis Under Varying the Percentage of Attackers

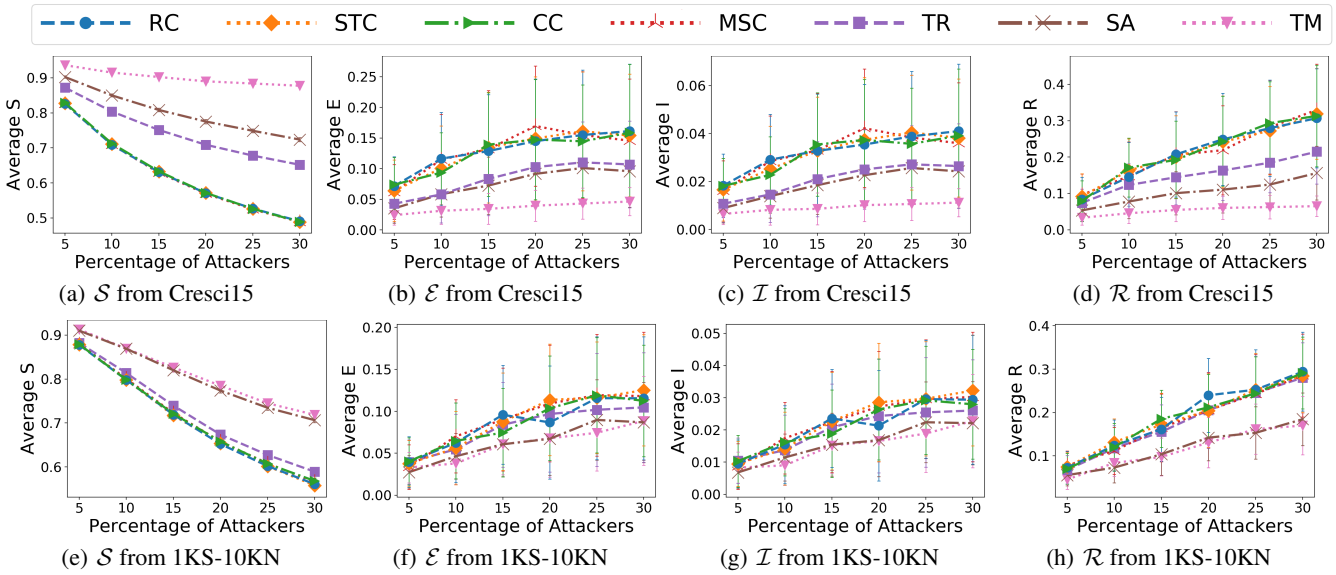
Figure 6 shows the performance of the seven FRSs under varying the percentage of attackers (P_a) on $\mathcal{S}$, $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}$ where attackers perform one-time phishing attacks. As overall trends, more attackers in a network introduce more exposed users $\mathcal{E}$ and accordingly increase more recovered users $\mathcal{R}$, resulting in the reduced $\mathcal{S}$.



(a) $\mathcal{S}$ under Cresci15 (b) $\mathcal{E}$ under Cresci15 (c) $\mathcal{I}$ under Cresci15 (d) $\mathcal{R}$ under Cresci15

(e) $\mathcal{S}$ under 1KS-10KN (f) $\mathcal{E}$ under 1KS-10KN (g) $\mathcal{I}$ under 1KS-10KN (h) $\mathcal{R}$ under 1KS-10KN

Figure 5: $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{R}$ comparisons of seven FRSs varying one to five phishing attacks by 20% attackers from the Cresci15 (Cresci et al. 2015) and the 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012), respectively.



(a) $\mathcal{S}$ from Cresci15 (b) $\mathcal{E}$ from Cresci15 (c) $\mathcal{I}$ from Cresci15 (d) $\mathcal{R}$ from Cresci15

(e) $\mathcal{S}$ from 1KS-10KN (f) $\mathcal{E}$ from 1KS-10KN (g) $\mathcal{I}$ from 1KS-10KN (h) $\mathcal{R}$ from 1KS-10KN

Figure 6: The comparisons of seven FRSs varying the fraction of attackers (P_a) from Cresci15 (Cresci et al. 2015) and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012), respectively.

Under Cresci15 with bot attackers, the performance of SC-based FRSs clearly outperforms non-SC-based counterparts, TR, SA, and TM, showing the highest $\mathcal{R}$. Interestingly, all SC-based schemes perform very similarly without showing much difference in all SC-based FRSs. During the first attack time, if more attackers exist in the network from 5% to 30%, the users who are attacked increase from 6% to 15% while the recovered users increase more from 9% to 30% in SC-related FRSs. However, still, the outperformance of SC-based FRSs over non-SC-based counterparts (i.e., TR, SA, and TM) is clear. Both are demonstrated in Figures 6(b) and 6(d). Under 1KS-10KN, the gap between SC-related FRSs and non-SC related FRSs is closer in $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}$ from Figures 6(f), 6(g), and 6(h); especially, $\mathcal{R}$ in TR is close to SC-based FRSs.

Under both Cresci15 and 1KS-10KN, more users from $\mathcal{S}$ move to $\mathcal{R}$, $\mathcal{E}$, and $\mathcal{I}$ in the network built by SC-based FRSs than three baseline FRSs. The increased $\mathcal{R}$ implies that more

phishing targets can verify the phishing attacks. This is the benefits of leveraging friend resources: The target can verify the attacker by his/her own judgment, and the target has a chance to help the friends to increase experience g_i, or the target cannot verify the targets, but his/her friends have the chance to detect the phishing attacks and reply to the target.

Conclusion & Future Work

We proposed the SAFER (Social cApital-based FriEnd Recommendation) scheme to combat phishing attacks in OSNs. We quantitatively measured social capital by three dimensions in terms of structural, cognitive, and relational capital based on nine behavioral features, which are collected from the two real datasets, Cresci15 (Cresci et al. 2015) and 1KS-10KN (Yang, Harkreader, and Gu 2011; Yang et al. 2012). We analyzed and demonstrated how the social capital in three dimensions is different between legitimate users and attackers.

We developed four different friend recommendation schemes (FRSs) based on four different types of social capital (i.e., relational, cognitive, structural, and multidimensional social capital) and compared their performance with three different non-social capital friending decision schemes, which are trust-based (Cho, Alsmadi, and Xu 2016), social attributes-based (Guo, Zhang, and Fang 2015) and topic-based (Wang et al. 2015). We investigated the performance of these seven FRSs in the extent of resistance against phishing attacks using two real datasets with phishing attacks by bots and humans, respectively. The users were modeled as legitimate users, while spammers were modeled as phishing attackers. A user's resistance against phishing attacks is evaluated in terms of three states, susceptible, infected, and recovered, based on the proposed SER-SEIR model. We examined the performance of the seven FRSs under varying attack severity and portion in a given OSN.

We identified the following **key findings** from our study:

- Users having friends with high social capital (SC) can self-defend against phishing attacks better than users having friends with the same topic interests or social attributes.
- SC-based FRSs can enable users to combat phishing attacks better than non-SC-based counterparts because the friends of the users with high social capital can help them defend against the phishing attacks.
- Bot-based phishing attacks can be more easily detected and defended than human-based phishing attacks under all FRSs because bot attackers show more distinctive characteristics than human attackers in social capital.
- Although SC-based FRSs can allow more attackers to infect (engage) users in the attacks due to users with high social capital attracting more attacks, even the infected users can be easily recovered with the help of their friends in their social networks.
- All SC-based FRSs perform comparably in detecting phishing attacks, while the cognitive SC has shown the best performance among all FRSs with a slightly better performance. This suggests that if a weighted linear model is used, cognitive SC can be assigned with a higher weight than relational SC and structural SC.

In future work, we plan to: (1) collect real datasets with more user behavioral features and personality traits based on text information generated by users in OSNs; (2) study the influences of personality traits on social capital quantification and the behaviors of legitimate users and human attackers types; (3) study more online social deception attacker types and the effect of social capital on the defense capability against them; and (4) address the effectiveness of using social capital in terms of recommending 'useful friends', emphasizing the resourcefulness which is well-aligned with the core concept of social capital.

Ethical Statement

We use publicly available datasets collected from Twitter API in the existing research (Yang, Harkreader, and Gu 2011; Yang et al. 2012; Cresci et al. 2015) to evaluate our proposed approach. The datasets were all anonymized by hiding identity information by their publishers.

Broader Impact. This work can introduce the potential broader impact to build a safe, trustworthy cyberspace by defending online social networks against phishing attacks through intelligent user interactions based on the proposed friending recommendation framework.

Funding and Competing Interests. This work is funded by The Virginia Tech and has no competing interests with financial activities outside this paper.

References

Alaa, A. M.; Ahuja, K.; and Schaar, M. 2018. A micro-foundation of social capital in evolving social networks. *IEEE Trans. Network Science and Engineering*, 5(1): 14–31.

Andersson, A. B. 2021. Social capital and leaving the nest: Channels and housing tenures. *Social Networks*, 65: 8–18.

Badri Satya, P. R.; Lee, K.; Lee, D.; Tran, T.; and Zhang, J. 2016. Uncovering fake likers in online social networks. In *Proc. the 25th ACM CIKM*, 2365–2370.

Blanchard, A.; and Horan, T. 2000. *Social Dimensions of Information Technology: Issues for the New Millennium*, 6–22. Hershey, PA: IGI Global.

Burt, R. S. 2000. The network structure of social capital. *Research in Organizational Behavior*, 22: 345–423.

Chen, C. C.; Shih, S.-Y.; and Lee, M. 2016. Who should you follow? Combining learning to rank with social influence for informative friend recommendation. *Decision Support Systems*, 90: 33–45.

Chen, L.; Xie, Y.; Zheng, Z.; Zheng, H.; and Xie, J. 2020. Friend Recommendation Based on Multi-Social Graph Convolutional Network. *IEEE Access*, 8: 43618–43629.

Cheng, H.; Qian, M.; Li, Q.; Zhou, Y.; and Chen, T. 2018. An efficient privacy-preserving friend recommendation scheme for social network. *IEEE Access*, 6: 56018–56028.

Cho, J.-H.; Alsmadi, I.; and Xu, D. 2016. Privacy and social capital in online social networks. In *2016 IEEE Global Communications Conf. (GLOBECOM)*, 1–7.

Cho, J.-H.; Rager, S.; O'Donovan, J.; Adali, S.; and Horne, B. D. 2019. Uncertainty-Based False Information Propagation in Social Networks. *Trans. Soc. Comput.*, 2(2).

Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; and Tesconi, M. 2015. Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems*, 80: 56–71.

Ding, D.; Zhang, M.; Li, S.-Y.; Tang, J.; Chen, X.; and Zhou, Z.-H. 2017. BayDNN: Friend recommendation with Bayesian personalized ranking deep neural network. In *Proc. the 2017 ACM CIKM*, 1479–1488.

Gatti, M.; Cavalin, P.; Neto, S. B.; Pinhanez, C.; dos Santos, C.; Gribel, D.; and Appel, A. P. 2014. Large-Scale Multi-agent-Based Modeling and Simulation of Microblogging-Based Online Social Network. In Alam, S. J.; and Parunak, H. V. D., eds., *Multi-Agent-Based Simulation XIV*, 17–33.

Guo, L.; Zhang, C.; and Fang, Y. 2015. A trust-based privacy-preserving friend recommendation scheme for online social networks. *IEEE Trans. Dependable and Secure Computing*, 12(4): 413–427.

Guo, Z.; Cho, J.-H.; Chen, I.-R.; Sengupta, S.; Hong, M.; and Mitra, T. 2020. Online social deception and its countermeasures: A survey. *IEEE Access*, 9: 1770 – 1806.

Huang, S.; Zhang, J.; Schonfeld, D.; Wang, L.; and Hua, X. 2017. Two-stage friend recommendation based on network alignment and series expansion of probabilistic topic model. *IEEE Trans. Multimedia*, 19(6): 1314–1326.

Huang, S.; Zhang, J.; Wang, L.; and Hua, X. 2016. Social friend recommendation based on multiple network correlation. *IEEE Trans. Multimedia*, 18(2): 287–299.

Inuwa-Dutse, I.; Liptrott, M.; and Korkontzelos, I. 2018. Detection of spam-posting accounts on Twitter. *Neurocomputing*, 315: 496–511.

Jung, Y.; Gray, R.; Lampe, C.; and Ellison, N. 2013. Favors from facebook friends: Unpacking dimensions of social capital. In *Proc. the ACM CHI*, 11–20.

Karlan, D. S. 2005. Using experimental economics to measure social capital and predict financial decisions. *American Economic Review*, 95(5): 1688–1699.

Kim, Y.; Schneider, T.; Faß, E.; and Lochbaum, M. 2021. Personal social capital and self-rated health among middle-aged and older adults: a cross-sectional study exploring the roles of leisure-time physical activity and socioeconomic status. *BMC Public Health*, 21(1): 1–11.

Li, M. Y.; and Muldowney, J. S. 1995. Global stability for the SEIR model in epidemiology. *Mathematical Biosciences*, 125(2): 155–164.

Marcaletti, F.; and Oldani, R. C. 2021. An exploratory study on family social capital and its relations with other forms of social capital in Spain. *REIS: Revista Española de Investigaciones Sociológicas*, (173): 47–68.

Nahapiet, J.; and Ghoshal, S. 1998. Social capital, intellectual capital and the organizational advantage. *Academy of Management Review*, 23(2): 242–266.

Ning, H.; Dhelim, S.; and Aung, N. 2019. PersoNet: Friend recommendation system based on Big-Five personality traits and hybrid filtering. *IEEE Trans. Computational Social Systems*, 6(3): 394–402.

Phua, J.; Jin, S. V.; and Kim, J. J. 2017. Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat. *Computers in Human Behavior*, 72: 115–122.

Phung, D.; Gupta, S. K.; Nguyen, T.; and Venkatesh, S. 2013. Connectivity, online social capital, and mood: A Bayesian nonparametric analysis. *IEEE Trans. Multimedia*, 15(6): 1316–1325.

Portes, A. 1998. Social capital: Its origins and applications in modern sociology. *Annual Rev. Sociology*, 24(1): 1–24.

Putnam, R. D. 2000. *Bowling Alone*. New York, USA: Simon and Schuster.

Tsai, M.-C. 2021. Kin, friend and community social capital: Effects on well-being and prospective life conditions in Japan, South Korea and Taiwan. *Social Indic. Res.*, 1–22.

van de Weijer, S. G.; Leukfeldt, R.; and Bernasco, W. 2018. Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. *European Journal of Criminology*.

Venkatanathan, J.; Karapanos, E.; Kostakos, V.; and Gonçalves, J. 2012. Network, personality and social capital. In *Proc. 4th Annual ACM WebSci'12*, 326–329.

Wang, Z.; Liao, J.; Cao, Q.; Qi, H.; and Wang, Z. 2015. Friendbook: A semantic-based friend recommendation system for social networks. *IEEE Trans. Mobile Computing*, 14(3): 538–551.

Warburton, D. 2020. 2020 Phishing and Fraud Report. https://www.f5.com/labs/articles/threat-intelligence/2020-phishing-and-fraud-report. Accessed: 2021-05-01.

Westerman, D.; Spence, P. R.; and Van Der Heide, B. 2012. A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior*, 28(1): 199–206.

Xu, Y.; Zhou, D.; and Ma, J. 2019. Scholar-friend recommendation in online academic communities: An approach based on heterogeneous network. *Decision Support Systems*, 119: 1–13.

Yang, C.; Harkreader, R.; Zhang, J.; Shin, S.; and Gu, G. 2012. Analyzing spammers' social networks for fun and profit: A case study of cyber criminal ecosystem on twitter. In *Proc. the 21st Int'l Conf. on World Wide Web*, 71–80.

Yang, C.; Harkreader, R. C.; and Gu, G. 2011. Die free or live hard? Empirical evaluation and new design for fighting evolving twitter spammers. In *Int'l Workshop on Recent Advances in Intrusion Detection*, 318–337. Springer.

Yang, K. 2007. Individual social capital and its measurement in social surveys. *Survey Research Methods*, 1(1): 19–27.

Ye, S.; Ho, K. K.; and Zerbe, A. 2021. The effects of social media usage on loneliness and well-being: Analysing friendship connections of Facebook, Twitter and Instagram. *Information Discovery and Delivery*.

Zheng, N.; Song, S.; and Bao, H. 2015. A temporal-topic model for friend recommendations in Chinese Microblogging systems. *IEEE Trans. Systems, Man, and Cybernetics: Systems*, 45(9): 1245–1253.