This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TETC.2018.2840433, IEEE Transactions on Emerging Topics in Computing

Quantifying Privacy Vulnerability to Socialbot Attacks: An Adaptive Non-Submodular Model

Xiang Li, *Student Member, IEEE,* J. David Smith, *Student Member, IEEE,* Tianyi Pan, Thang N. Dinh, *Member, IEEE,* and My Thai, *Member, IEEE*

Abstract—Privacy breaches are one of the biggest concerns on Online Social Networks (OSNs), especially with an introduction of automated attacks by *socialbots*, which can automatically extract victims' private content by exploiting social behavior to befriend them. The key insight of this attack is that by intelligently sending friend requests to a small subset of users, called the Critical Friending Set (CFS), such a bot can evade current defense mechanisms. We study the vulnerability of OSNs to socialbot attacks. Specifically, we introduce a new optimization problem, Min-Friending, which identifies a minimum CFS to friend in order to obtain at least Q *benefit*, which quantifies the amount of private information the bot obtains. The two main challenges of this problem are how to cope with incomplete knowledge of network topology and how to model users' responses to friend requests. In this paper, we show that Min-Friending is inapproximable within a factor of $(1 - o(1)) \ln Q$ and present an adaptive approximation algorithm using adaptive stochastic optimization. The key feature of our solution lies in the adaptive method, where partial network topology is revealed after each successful friend request. Thus the decision of whom to send a friend request to next is made with the outcomes of past decisions taken into account. Traditional tools break down when attempting to place a bound on the performance of this technique with realistic user models. Therefore, we additionally introduce a novel curvature-based technique to construct an approximation ratio of $\ln Q$ for a model of user behavior learned from empirical measurements on Facebook.

Index Terms—SocialBots Attacks; Social Networks Analysis; Content-Centric Privacy; Adaptive Algorithms; Non-Submodularity.

1 Introduction

With a huge amount of personal information ripe for the taking in modern Online Social Networks (OSNs), privacy breaches have become a central concern. During the last two decades, we have witnessed the blossoming of a variety of attacks aimed at collecting users' private content, many of which apply socialbots to automatically infiltrate users' social circles and exfiltrate sensitive data [1], [2]. This information may then be exploited for a number of purposes, including spearphishing and account compromise via security questions [2]. Therefore, studying such threats to online privacy is of great importance both as a way to protect users and a means to improve awareness of these dangers.

Motivated by the above discussion, we present a new paradigm to measure the OSN privacy vulnerability in light of socialbot attacks. Although quantitative analysis of network vulnerability can be addressed from a variety of perspectives, an intuitive measure is the minimum number of users that attackers need to befriend in order to maximally collect private information from a target network. Obviously, if this number is small compared to the benefit gained from collecting private content, one can conclude that the network is vulnerable to attack, whereas if this number is large the network is robust against socialbot attacks. Identifying the set of users to friend, called the *critical friending set (CFS)*, is useful from both attack and defense perspectives. In the former, the attacker identifies an optimal set of users they need to befriend, whereas in the latter, the defender has an opportunity to protect the CFS and interfere with attack in order to protect users' privacy.

Accordingly, we introduce a new optimization problem, called the *Adaptive Minimum Critical Friending Set (Min-Friending)* Problem. It asks us to find the minimum number of users to befriend in order to obtain at least Q benefit, which quantifies valuable private content. The key challenges of this problem are multifold. First of all, the topological information of social networks is partially unavailable. Since only two-hop topology is available by default in closed OSNs such as Facebook, the users' connections are gradually revealed to the attacker when acquiring new friends, thereby requiring us to investigate adaptive strategies. Additionally, the huge number of OSN users and amount of data available on OSNs poses a substantial challenge to mining the critical friending set with incomplete topology. Finally, the variety of potential social responses to friend requests makes it difficult to design an efficient, general friending strategy, and thus challenging to identify the CFS.

A further notable challenge of Min-Friending is that whether the objective function is submodular (i.e. possessing the property of diminishing returns) depends on the way users behave. Although it is well-known that the greedy algorithm obtains an approximation ratio of $(1 + \ln Q)$ for minimization of submodular objective functions [3], [4], there is no tool to theoretically bound the performance of adaptive greedy minimization with non-

- *Xiang Li, J. David Smith, Tianyi Pan and My Thai are with the Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, 32611.*
E-mail: {xixiang, jdsmith, tianyi, mythai}@cise.ufl.edu
- *Thang N. Dinh is with the Computer Science Department, Virginia Commonwealth University, Richmond, VA 23284.*
Email: tndinh@vcu.edu
- *Xiang Li and J. David Smith contributed equally to this work*

submodularity.

Our contributions are summarized as follows:

- Prove Min-Friending cannot be approximated within the ratio of $(1 - o(1)) \ln Q$ unless $NP \subset DTIME(n^{O(loglogn)})$. That being said, no one can design an algorithm which can guarantee its solution is within $(1 - o(1)) \ln Q$ factor from the optimal solution under all instances.
- Design an adaptive approximation algorithm, AReST, to Min-Friending. Using adaptive stochastic optimization, we show that AReST has a performance ratio of $(1 + \ln Q)$ in some cases, which is a tight bound.
- Provide a novel curvature-based technique to theoretically bound the approximation ratio of adaptive and non-submodular greedy minimization. This in turn shows AReST has a bound of $(\delta \ln Q)$ in the general case, where δ denotes the maximum change in acceptance rate.
- Conduct extensive experimental evaluations showing that AReST outperforms several alternate methods. Further, we find that the effectiveness of infiltration has strong dependencies on the network topology and user behavior in addition to the attacker's choice of target. Among target settings, we find that attacking individuals and communities are the most difficult, while attacking the network as a whole or a tightly co-located group is notably easier.

Related Work. Socialbot attacks on OSNs may be simple attempts to collect personal information, but have shown to be critically important [5]. Ryan & Mauch showed that fake profiles can be effectively used to befriend members of the NSA, military intelligence agencies, and Global 500 corporations [6]. These fake profiles can be automated to form socialbots [2], which can be used to automatically infiltrate organizations [7]. Furthermore, these attacks preclude the use of existing defenses to Sybil attacks, which have seen significant study [8], [9], [10] (and references therein). The reasons are twofold. On the one hand, a socialbot attack may consist of only a single bot, and even in cases where multiple bots are involved they do not need to be related by network topology (which is a key assumption of common Sybil defenses). On the other hand, they are carefully designed to behave similarly to the real users. Thus, neither the graph-based nor the feature-based Sybil detection mechanisms are likely to be effective against socialbot attacks. Therefore, there is an urgent need for radically new models and analytical techniques to assess OSNs' vulnerability against these attacks.

When considering friending strategies for socialbots, the number of works on this new direction is rather limited. Two works that are the most relevant to our paper are [11], [12]. However, these papers aimed to find the top k users to friend in order to maximize the gained benefit, which is a dual to our problem. In addition, they only consider a simple constant friend acceptance rate, which may be impractical. Along the line of defense, the only relevant work is in [13], by monitoring a subset of users within an organization, and evaluating the cost to the organization by simulating an attacker on topologies taken from social networks. However, this work is based on heuristics built on known sociological properties and further assumes having complete knowledge of network topology. Instead, we provide a better defense strategy with a theoretical performance guarantee with support for incomplete topological knowledge.

Further, assessing network vulnerability to socialbot attacks has not been addressed. The vulnerability of networks due to malicious actors or external disasters has been characterized in a number of ways [14], [15], [16] (and references therein). However, the differences between these destructive attacks and socialbot attacks precludes applying these vulnerability measurements directly. Where prior work is concerned with the ability of an attacker to disrupt a network in term of network connectivity for example, we instead look at the ability of an attacker to extract information from it, thereby forming a new research direction.

With respect to theoretical tools to analyze the performance of greedy algorithms for optimization problems, Golovin and Krause showed that if the objective function is adaptive submodular and monotone non-decreasing, then the well-known approximation ratio of greedy in a non-adaptive setting can be extended to an adaptive one [4]. When the objective function is non-submodular, Wang *et al.* [17] derived a ratio for the maximization problems. However, the ratio quickly approaches to 0 for non-trivial problem sizes. In addition, this work only applied in the case of non-adaptive settings. Instead, our work introduced a new primal curvature concept to bound the *adaptive non-submodular* greedy minimization.

Organization. The rest of the paper is organized as follows. Section 2 presents the problem models and preliminaries, including the inapproximability. The AReST algorithm and its efficient implementation are introduced in Section 3, followed by the theoretical analysis in Section 4. Section 5 presents our experimental evaluation and Section 6 concludes the paper.

2 Problem Models and Preliminaries

2.1 Friending by Socialbots

In order to solve Min-Friending, we first need to understand how socialbots befriend targets. To generalize our problem, we consider a target set of users T that an attacker wants to collect information from. T could be a set of employees in a targeted organization, one individual with a high profile ($|T| = 1$), or the set of all users in a network. For simplicity, we assume that there is a single attacker s, i.e. a socialbot, who is an online user in the same networking environment[1].

Conceptually, the socialbot attack works as follows. The attacker s first obtains a master-list of target users T through some public channels, e.g. organization's website or the OSNs themselves which expose a certain amount of personal information. This has been made even easier by the fact that popular social networks such as Facebook enforce a real-name policy, which facilitates identifying users. Because

[1] The solution proposed in this paper can be easily extended to handle multiple attackers.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TETC.2018.2840433, IEEE Transactions on Emerging Topics in Computing

of the privacy settings available to users[2], the only way s can reliably gather the information about $t \in T$, assuming t has privacy setting to "Friends", is befriending t. To successfully become friends with t, s needs to achieve the following: 1) s needs to mimic a normal user, and thus needs to have a few friends initially. This can be easily done by sending friend requests to users who have a high number of friends as they tend to accept all friend requests [2], [7]. 2) Since s and t usually have no mutual friends, the probability of t accepting s's request is likely low. Thus s should attempt to befriend t's friends first, which in turn means s should send friend requests to friends of friends of t, and so on.

However, s cannot send too many friend requests or it will easily get detected by any network monitoring service/manager (e.g. based on anomaly detection methods). The best strategy for s is to mimic normal behavior by sending friend requests to a small number of users, observing the response and then sending to another set of users. Once a user v accepts the friend request of s, s can collect all v's information, and all the friends of v become visible to s. This strategy of repeating the process of making decisions subject to previous decisions and observing new results is called an adaptive strategy.

Based on the above discussion, the central concern is who s should select in each decision making step to minimize the number of friend requests while successfully gathering information about T. Addressing this concern is equivalent to solving our Min-Friending problem.

2.2 Network Models

From the aforementioned insights, we abstract an OSN as a directed graph $G = (V, E)$ where $V = \{v_1, v_2, \ldots, v_n, s\}$ is the set of n users and attacker s, who initially has no connections to other users. E is the set of m directed edges where each edge $(u, v) \in E$ represents the friendship between u and v. Note that due to the privacy settings, the friendship information (network topology) available to s is incomplete. However, s can estimate these friendship probabilities based on link prediction methods [18], [19], [20] which may combine both the publicly observable connections and users' public profiles. Therefore, we model this by letting each edge $e \in E$ exist with some probability $p_e \in [0, 1]$. Once u accepts the friend request from s and all of u's friends are visible to s, then $p_{uv} = 1$ iff v is u's friend. Else $p_{uv} = 0$.

Friend Request Acceptance Model. Let accept(u) denote the probability that u accepts a friend request from s. This function accept(u) is complex due to the social behaviors of users. For example, when u has a very high number of friends in his circle, accept(u) tends towards 1 [7]. Boshmaf et al. found that increasing the number of mutual friends dramatically increased the friend acceptance rate on Facebook, which they explained as a result of the Triadic Closure principle [2].

To derive an acceptance model, we fit the data generated in [2] to a degree-1 polynomial with a natural log term.

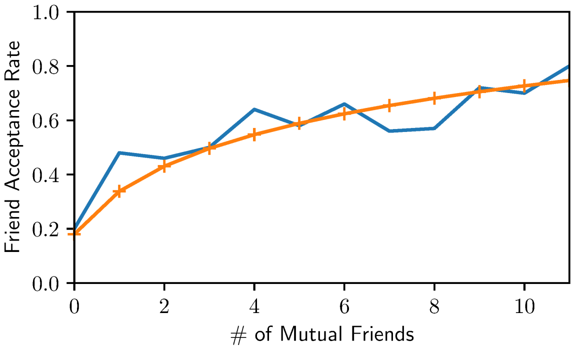


Fig. 1: The friend acceptance rate from the experiments of Boshmaf et al. [2] as a function of the number of mutual friends, with a logarithmic function fit to the data.

Figure 1 shows the original data and the estimated function. We use the following fitted function as the main friend request acceptance model in our paper:

$$\text{accept}(u) = \rho_1 \log(\mathbb{E}[|N(u) \cap N(s)|] + 1) + \rho_0 \quad (1)$$

with $\rho_1 = 0.22805837$ and $\rho_0 = 0.18014571$. $N(\cdot)$ denotes the set of neighbors. In a more general sense, this formula incorporates the willingness of a user to accept a new, unknown friend (ρ_0) and how much sharing mutual friends improves that willingness (ρ_1). Given the limited amount of data available, learning the distribution of per-user weights is currently infeasible, though we conduct experiments in the special case of each user u having independent $\rho_0(u)$.

Information Benefit Model. In order to quantify the benefit that s achieves by gathering the information in OSNs, each node $u \in V$ is associated with a benefit $B_{fof}(u) \in \mathbb{R}_0^+$ when u becomes a friend of friend of s, i.e., 2 hops away from s[3]. Each node u is also associated with a benefit $B_f(u) \geq B_{fof}(u), B_f(u) \in \mathbb{R}_0^+$ when u becomes a friend of s. Note that when u is both a friend and a friend of friend of s, only the friend benefit $B_f(u)$ is in effect. Moreover, when each edge $(u, v) \in E$ is revealed, (i.e. the attacker learns about the existence of (u, v)), the attacker gains an information benefit $B_i(u, v) \in \mathbb{R}^+$. The existence of edge (u, v) is revealed only when node u becomes a friend of s. At this point, $p_{uv} = 1$.

We note that a target set T can be encoded in the benefit functions $B = (B_f, B_{fof}, B_i)$. For example, we can define $B_f(u) = \mathbf{1}_{u \in T}$ where $\mathbf{1}_p$ is the indicator function taking value 1 when predicate p is true. Therefore, rather than take a target set T as a parameter, we take the more general benefit functions B.

2.3 Problem Definitions and Formulations

Based on the above model, the goal of attacker s is to gain the greatest total benefit in the minimum number of friend requests. Accordingly, we study the following problem:

Definition 1 (Adaptive Minimum Critical Friending Set (Min-Friending))**.** *Given a social network* $G = (V, E, p, B, accept)$ *where* V *is the set of user accounts,* E *is*

[2] Privacy settings allow account owners to specify who can see what in their accounts. For example, in Facebook, users can specify who are able to see their friends list based on three categories: Public (everybody), Friends (only whom you are connected with) and Only Me (nobody else but me).

[3] $\mathbb{R}_0^+$ is the set of non-negative reals.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TETC.2018.2840433, IEEE Transactions on Emerging Topics in Computing

a set of possible friendships between users, each edge $e \in E$ exists with a probability $p_e \in [0,1]$, and a threshold $Q \in \mathbb{Z}^+$. The benefit function B and acceptance probability function accept$(\cdot)$ are defined earlier. The problem asks us to find a set of nodes to befriend $F \subset V$ with minimum size so as when s sends friend requests to F, the total expected benefit gain is at least Q.

Note that finding F is equivalent to finding an adaptive attack strategy π, in which s will befriend $u \in F$ iteratively. Each time s becomes a friend of u, the network topology G will be updated to reveal all edges incident with u. As $|F|$ is minimized, the number of friend request steps is also minimized.

Since G is partially unknown to s and friend requests sent from s to u may fail, we use adaptive stochastic optimization to tackle our problem. We begin by introducing our notation. For each node $u \in V$, let $X_u \in \{0,1,?\}$ denote the state of u where 1 indicates that u accepts the friend request from s, 0 indicates that u rejects the friend request, and ? represents an unknown, i.e., s has not sent a request to u yet. Initially, the states of all us should be ?. Likewise, for each edge $(u,v) \in E$, define $Y_{uv} \in \{0,1,?\}$. 1 means the edge (u,v) exists (revealed when s befriends u and v is u's friend), 0 indicates edge (u,v) is not present (revealed when s befriends u and we learn for certain that v is not a friend of u), and ? means unknown, i.e. u rejects the friend request from s, or s has not sent a friend request to u yet, or u has the privacy setting to himself only (not friend of friend). Let Ω be the collection of all possible states of G and $\phi = \{X_v\}_{v \in V} \cup \{Y_{uv}\}_{(u,v) \in E} \to \Omega$ be a possible state, called a *realization*. Thus we call $\phi(u)$ the state of node u and $\phi(e)$ the state of edge e under realization ϕ. We require each realization to be consistent. That is, each node and edge must be in only one of the states $\{0,1,?\}$. Clearly there are many possible realizations which follow a probability distribution $\Pr[\phi]$. We denote Φ as a random realization and $\Pr[\phi] = \Pr[\Phi = \phi]$ over all realizations.

We will consider the problem where s sequentially sends a friend request to u, sees the state $\Phi(u)$, and sees the states $\Phi(e)$ for all e incident to u iff $\Phi(u) = 1$. s then picks the next user to befriend, see its state, and so on. We use the notation $F(\pi, \phi)$ be the set of nodes selected by strategy π under realization ϕ. After each friend request, our observations thus far can be represented as a *partial realization* ω. We use dom(ω) to refer to the domain of ω, ie., the set of nodes and edges observed in ω. A partial realization ω is consistent with a realization ϕ if they are equal everywhere in the domain of ω, written as $\phi \sim \omega$. If ω and ω' are both consistent with some ϕ and dom$(\omega) \subseteq$ dom(ω'), we say ω is a subrealization of ω'.

Let π be an adaptive attack strategy of s. The total benefit gain from this strategy π under realization ϕ can be formulated as follows.

$$f(\pi,\phi) = \sum_{u \in N_f(\pi,\phi)} B_f(u) + \sum_{v \in N_{fof}(\pi,\phi)} B_{fof}(v) + \sum_{(u,v) \in N_i(\pi,\phi)} B_i(u,v) \qquad (2)$$

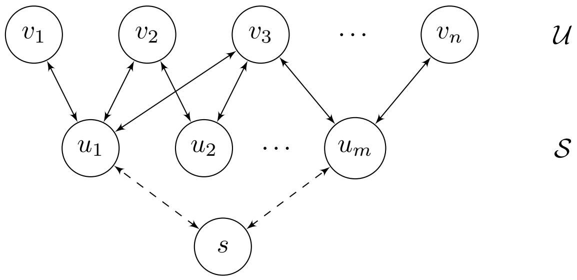


Fig. 2: An instance Π' of Min-Friending after the attacker s friends u_1 and u_m (corresponding to S_1 and S_m).

where

$$\begin{aligned} N_f(\pi,\phi) &= \{u | u \in F(\pi,\phi), \phi(u) = 1\}, \\ N_{fof}(\pi,\phi) &= \{v | \exists u \in N_f(\pi,\phi) : \phi(u,v) = 1\} \backslash N_f(\pi,\phi), \\ \text{and } N_i(\pi,\phi) &= \{(u,v) | u \in N_f(\pi,\phi), \phi(u,v) = 1\} \end{aligned}$$

Therefore, the Min-Friending problem can be stated formally as:

$$\begin{aligned} &\min \mathbb{E}[|F(\pi,\Phi)|] \\ &s.t.\ \mathbb{E}[f(F(\pi,\Phi),\Phi)] \geq Q \end{aligned} \qquad (3)$$

That is: find a policy π that minimizes the expected number of friend requests to obtain at least Q benefit in expectation (where benefit is defined by Eqn. (2)). Each expectation is taken w.r.t. the set of potential realizations.

2.4 Inapproximability

Instead of proving that Min-Friending is NP-hard, we prove a stronger theorem, showing the inapproximability of Min-Friending.

Theorem 1. *The Min-Friending problem cannot be approximated within a factor $(1 - o(1)) \ln Q$ unless $NP \subset DTIME(n^{O(loglogn)})$*

Proof. Let $\Pi = (\mathcal{S}, \mathcal{U}, K)$ be an instance of the set-cover problem in which $\mathcal{U} = \{e_1, e_2, \ldots, e_n\}$ is the set of n elements and $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$ is the collection of m subsets of $\mathcal{U}$. The set cover problem asks if there are k subsets which cover at least $K \leq n$ items in $\mathcal{U}$. We construct an instance $\Pi' = G(V, E, p, B, accept)$ with target benefit Q of the Min-Friending problem as follows.

- For each element $e_i \in \mathcal{U}$, we include into V a vertex v_i. Similarly, we put into V a vertex u_j for each subset $S_j \in \mathcal{S}$.
- For each pair of element e_i and subset S_j, we connect u_j and v_i if $e_i \in S_j$. For all edges $(u_j, v_i) \in E$, we set $p_{u_j v_i} = 1$.
- For benefit B, we set $B_{fof}(v_i) = 1$ and $B_f(v_i) = 0$ for each $v_i \in V$ that corresponds to $e_i \in \mathcal{U}$. And we set $B_f(u_j) = B_{fof}(u_j) = 0$ for all u_j, associated with subset S_j, and $B_i(u_j, v_i) = 0$ for all edges in the graph.
- For accept(.), set accept$(u_j) = 1\ \forall u_j$, set accept$(v_i) = \rho_1 \log(\mathbb{E}[|N(v) \cap N(s)|] + 1) + \rho_0\ \forall v_i$.
- Finally, set $Q = K$.

So the Min-Friending problem asks us if there exists a CFS of size q such that the total benefit is at least Q. The construction is illustrated in Fig. 2.

Since $B_f(v_i) = 0$, there is no incentive to friend v_i. Thus attacker s will befriend u_j. This friend request to u_j is always successful as accept$(u_j) = 1$. In order to have $Q = K$ benefit, s needs to have at least K v_i in his two-hop neighbors. Then the users u_j that s chooses to befriend are corresponding to the k sets S_j of Π. Clearly if we have an approximation algorithm with an approximation factor $\alpha(Q)$ for Min-Friending, then we also have an $\alpha(K)$ approximation algorithm for the set cover problem. Thus, due to the inapproximability of set cover [21], the Min-Friending problem cannot be approximated within a factor $(1 - o(1)) \ln Q$ unless NP has $n^{O(\log \log n)}$ deterministic time algorithms. □

3 Adaptive Reconnaissance Strategies

In this section, we present our solution to Min-Friending, namely Adaptive Reconnaissance Strategy algorithm (AReST), followed by the discussion of its efficient implementation.

3.1 Algorithm Description

At an abstract level, the Adaptive Reconnaissance Strategy algorithm (AReST) has two main phases: *Selection* and *Feedback*. At the *Selection* phase, AReST will select a node u for s to friend so as to increase the potential function (which will be discussed later) the most. After selecting u and sending a friend request to u, if u accepts the friend request, AReST executes the *Feedback* phase, which will (1) update the network topology with more exact information on p_e ; and (2) update the accept(v) for all $v \in N(u)$. If u rejects the friend request, AReST will continue the first phase to select another node. These two phases will be iteratively executed until the total expected benefit exceed Q.

Algorithm 1: Adaptive Reconnaissance Strategies (AReST)

```
   Input: Graph $G = (V, E, p, B, accept)$, and $Q \in \mathbb{Z}^+$
   Output: An ordered set of nodes $F \subset V$ for $s$ to friend
           with.
1  $F \leftarrow \emptyset; \omega \leftarrow \emptyset$
2  while $\mathbb{E}[f(F)] < Q$ do
3      foreach $u \in V \backslash F$ do
4          $\Delta(u|\omega) = \text{accept}(u)(P_1 + P_2)$
5      Select $u^* \in \arg\max_u \Delta(u|\omega)$
6      Set $F \leftarrow F \cup \{u^*\}$
7      Send a friend request to $u^*$
8      if $u^*$ accepts the friend request then
9          Feedback: Update $\omega$ with new observed
           information of $p_{u^*v}$ and accept$(v)$ for all
           $v \in N(u^*)$
10 Return $F$
```

The main challenge of the first phase is to define an efficient potential function. As the friend request acceptance probability may change after each successful friend request, the potential function must account for the likelihood of increasing the acceptance probability in a later iteration, in addition to the gain defined by the benefit function B. Let F denote the set of s' friends at the current step, with corresponding partial realization ω. In order to select a node u at the next step, we define the potential function as follows:

$$\Delta(u|\omega) = \text{accept}(u)(P_1 + P_2)$$

where P_1 and P_2 represent the gain in increasing the acceptance probability for later stages and the gain in increasing the benefit function B, respectively. Mathematically, we have:

$$P_1 = \frac{1}{|N(u) \setminus N(s)|} \sum_{v \in N(u) \setminus N(s)} p_{uv} \times \Delta P_u(s, v) \times \Delta_{uv} B$$

where $\Delta P_u(s, v)$ denotes the gain in the acceptance probability when u becomes a friend of s. $\Delta P_u(s, v)$ can be calculated based on the definition of accept(.) function: $\Delta P_u(s, v) = \rho_1 \log(1 + p_{uv}/(\mathbb{E}[|N(u) \cap N(s)|] + 1))$. In the special cases of u placing low value on mutual friends or u having many friends, this tends to 0. $\Delta_{uv} B$ represents the benefit gain assuming s adds u as a friend, and then add v as a friend. Thus $\Delta_{uv} B = f(\omega \cup \{u, v\}) - f(\omega \cup \{u\})$ and

$$\begin{aligned} P_2 = {} & B_f(u) - I_{fof}(u) B_{fof}(u) \\ & + \Bigg(\sum_{v \in N(u) \setminus N(s)} p_{uv} (1 - I_{fof}(v)) B_{fof}(v) \\ & \quad + \sum_{(u,v) \in E} p_{uv} B_i(u, v) \Bigg) \end{aligned}$$

where $I_{fof}(u) = 1$ if u is a friend of a friend of s, and 0 otherwise. Algorithm AReST is depicted in Algorithm 1.

3.2 Improving the Efficiency of Potential Updates

Lines 3-4 of Alg. 1 update the potential of every element that could be added to the solution in the next step. We note that the efficiency of this operation can be dramatically improved by maintaining a cache of the potentials and updating only the entries whose potential might have changed as a result of adding a node w to the solution F.

Lemma 1. *$\Delta(u \mid \omega)$ is a function only of the nodes and edges within an unweighted distance of 2 of u (the "2-hop" neighborhood of u), the bot s, and the current solution F.*

We only sketch the proof of Lemma 1 here. Fundamentally, the proof comes down to examining each term of $\Delta(u \mid \omega)$ to see which nodes and edges are needed for computation. In large part, it is straightforward to verify under the stated assumptions that each term is a function only of s, F, $N(N(u)) = N(u) \cup \bigcup_{v \in N(u)} N(v)$, and edges with both endpoints in $N(N(u))$. While there are a few terms which appear to require other inputs for computation, in each case after expanding the definition it becomes clear that the apparent requirement for other input is only a product of notational shortcuts.

As an example, consider the term $N(u) \setminus N(s)$. This term appears to require that one examine $N(s)$ so that one knows the correct elements to remove from $N(u)$. However, this can be flipped around to only examine nodes and edges in the 2-hop neighborhood of u as follows: for each node $v \in N(u)$ check the set $N(v)$. If it contains s, omit it.

Corollary 1. *For any $u \in V \setminus F$, if $w \notin N(N(u))$, $w \neq u$ is added to the solution, then $\Delta(u \mid \omega \cup \{w\}) = \Delta(u \mid \omega)$.*

4 Theoretical Performance Analysis

In this section, we provide a comprehensive analysis of AResT with respect to various friend request acceptance models. We first focus on a special case of the model where our objective function $f(\cdot)$ is adaptive submodular. We next cover a wide range of friend request acceptance models which makes $f(\cdot)$ no longer submodular.

4.1 Adaptive Submodular: A Special Case

We are going to analyze the performance of AReST where $\Delta P_u(s, v) = \rho_1 \log(1 + 1/\mathbb{E}[|N(u) \cap N(s)|]) = 0$. This relates to u placing low value on mutual friends or u having many friends, and thus their friend acceptance probility depends on $\rho_0(u)$. We show that AReST has an approximation ratio of $(1+\ln Q)$. As shown in Theorem 1, this ratio is tight.

Note that AReST calculates $\Delta(u|\omega)$ for all $u \in V\backslash F$ and chooses u^* with the maximal gain over all realization. Thus in this case, the expected marginal gain of u conditioned on having partial realization ω is defined as follows:

$$\Delta(u|\omega) = \mathbb{E}[f(\mathrm{dom}(\omega) \cup \{u\}, \Phi) - f(\mathrm{dom}(\omega), \Phi)|\Phi \sim \omega]$$

Before continuing with our analysis, we state the following definitions, which are defined in [4].

Definition 2 (Strongly Adaptive Monotone). *A function $f(\cdot)$ is strongly adaptive monotone with respect to the distribution $\Pr[\phi]$ if the following condition holds. For all ω, all $v \notin$ dom(ω), and all possible states o of node v such that $\Pr[\Phi(v) = o|\Phi \sim \omega] > 0$, we have:*

$$\begin{aligned}&\mathbb{E}[f(dom(\omega), \Phi)|\Phi \sim \omega]\\ &\quad \leq \mathbb{E}[f(dom(\omega) \cup \{v\}, \Phi)|\Phi \sim \omega, \Phi(v) = o]\end{aligned} \tag{4}$$

Definition 3 (Adaptive Submodularity). *A function $f(.)$ is adaptive submodular w.r.t the distribution $\Pr[\phi]$ of all realizations if for all ω and ω' such that $\omega \subseteq \omega'$ and for all $v \in V\backslash dom(\omega')$, we have:*

$$\Delta(v|\omega) \geq \Delta(v|\omega') \tag{5}$$

Definition 4 (Self-certifying). *An instance $(f, \Pr[\phi])$ is self-certifying if for all ϕ, ϕ', and ω such that $\phi \sim \omega$ and $\phi' \sim \omega$, we have $f(dom(\omega), \phi) = f(V, \phi)$ iff $f(dom(\omega), \phi') = f(E, \omega')$.*

Lemma 2. *The objective function f of the Min-Friending problem is strongly adaptive monotone.*

Proof. Consider a fixed ω, $v \notin$ dom(ω), and status o. Let $A(\omega)$ be a set of nodes and edges that can be reached from s after selecting dom(ω) and observing ω. Clearly, for all paths from s to $u \in A(\omega)$ consisting of $\omega(e) = 1$. Therefore, every path from any $u \in A(\omega)$ to any $v \in V\backslash A(\omega)$ must consist at least one $\omega(e) \neq 1$ or $\omega(w) \neq 1$ for some w on the path. Thus we have $f(A(\omega)) = \mathbb{E}[f(\mathrm{dom}(\omega), \Phi)|\Phi \sim \omega]$.

With a similar argument, we have $f(A(\omega \cup \{v\})) = \mathbb{E}[f(\mathrm{dom}(\omega) \cup \{u\}, \Phi)|\Phi \sim \omega, \Phi(u) = o]$. Note that $\omega \subseteq \omega'$ implies $A(\omega) \subseteq A(\omega')$. Since f is a monotone function by definition, we have $f(A(\omega)) \leq f(A(\omega'))$. Thus we obtain $\mathbb{E}[f(\mathrm{dom}(\omega), \Phi)|\Phi \sim \omega] \leq \mathbb{E}[f(\mathrm{dom}(\omega) \cup \{v\}, \Phi)|\Phi \sim \omega, \Phi(v) = o]$. This completes the proof. □

Lemma 3. *The objective function f of the Min-Friending problem is adaptive submodular.*

Proof. Consider two fixed partial realizations ω and ω' where $\omega \subseteq \omega'$ and a node $v \in V\backslash$dom(ω'), we need to prove that $\Delta(v|\omega) \geq \Delta(v|\omega')$. We first prove the following claim:

Given $\omega \subseteq \omega'$ and define a coupled distribution μ over pairs of realization $\phi \sim \omega$, $\phi' \sim \omega'$ such that $\phi(v) = \phi'(v)$ for all $v \notin$ dom(ω'). For all (ϕ, ϕ') in support of μ, we have:

$$\Delta(v|\omega, \phi \sim \omega) \geq \Delta(v|\omega', \phi' \sim \omega')$$

where $\Delta(v|\omega, \phi) = f(\mathrm{dom}(\omega) \cup \{v\}, \phi) - f(\mathrm{dom}(\omega), \phi)$.

Define $A(\omega)$ and $A(\omega')$ as in the proof of Lemma 2. We have:

$$\begin{aligned}\Delta(v|\omega, \phi \sim \omega) &= f(\mathrm{dom}(\omega) \cup \{v\}, \phi) - f(\mathrm{dom}(\omega), \phi)\\ &= f(A(\omega) \cup \{(v, \phi(v))\}) - f(A(\omega))\\ &\geq f(A(\omega') \cup \{(v, \phi'(v))\}) - f(A(\omega'))\\ &= f(\mathrm{dom}(\omega') \cup \{v\}, \phi') - f(\mathrm{dom}(\omega'), \phi')\\ &= \Delta(v|\omega', \phi' \sim \omega')\end{aligned}$$

Having proven the above claim, we can straightforwardly finish our proof. Since $\omega \subseteq \omega'$, we have:

$$\begin{aligned}\Delta(v|\omega) &= \mathbb{E}[f(\mathrm{dom}(\omega) \cup \{v\}, \Phi) - f(\mathrm{dom}(\omega), \Phi)|\Phi \sim \omega]\\ &= \sum_{(\phi,\phi')} \mu(\phi, \phi')\Delta(v|\omega, \phi \sim \omega)\\ &\geq \sum_{(\phi,\phi')} \mu(\phi, \phi')\Delta(v|\omega', \phi' \sim \omega')\\ &= \Delta(v|\omega') \qquad \square\end{aligned}$$

Theorem 2. *The AReST algorithm has an approximation ratio of $(1 + \ln Q)$.*

Proof. Let OPT represent an optimal solution to Min-Friending. According to [4], if f is strongly adaptive monotone, adaptive submodular, and self-certifying, then we have:

$$|F| \leq (1 + \ln(\frac{Q}{\eta}))OPT$$

where η is any value such that $f(F, \phi) > Q - \eta$ implies $f(F, \phi) = Q$ for all F and ϕ.

If range$(f) \in \mathbb{Z}$, we set $\eta = 1$ and thus have $|F| \leq (1 + \ln Q)OPT$.

Therefore, the only thing left we need to prove is that f is self-certifying. Clearly, we have $f(V, \phi) = \min\{Q, f(V)\} = Q$. We have shown that $f(\mathrm{dom}(\omega), \phi) = f(A(\omega))$ for every $\phi \sim \omega$. It follows that $f(\mathrm{dom}(\omega), \phi) = f(\mathrm{dom}(\omega), \phi')$ for all ω and ϕ, $\phi' \sim \omega$. Thus, we obtain $f(\mathrm{dom}(\omega), \phi) = Q$ iff $f(\mathrm{dom}(\omega), \phi') = Q$, which completes the proof. □

4.2 Adaptive Non-Submodular - A Generalization

Outside of the special case analyzed in the previous section, the objective function for AReST may not be submodular, and therefore the $(1 + \ln Q)$ ratio does not hold in general. In this section, we introduce the notion of *adaptive primal curvature* (APC) to assist in the derivation of an adaptive approximation ratio in the general case without adaptive submodularity. The notion is motivated by the *elemental curvature* defined in [17], with the capability of coping with adaptive functions and greater precision. Intuitively, the APC depicts the rate at which the marginal gain $\Delta(u \mid \omega)$ of the objective function changes would change if ω were extended with an additional element v prior to adding u. The elemental curvature, when extended to the adaptive case, is the maximum of all such rates of change. Formally, we have:

Definition 5 (Adaptive Primal Curvature)**.** *The APC of an adaptive monotone non-decreasing function f is*

$$\nabla_f(i, j \mid \omega) = \mathbb{E}\left[\frac{\Delta(i \mid \omega \cup s)}{\Delta(i \mid \omega)} \,\middle|\, s \in S(j)\right]$$

where $S(j)$ is the set of possible states of j and Δ is the conditional expected marginal gain [4].

For notational clarity, we also define the fixed adaptive primal curvature in terms of a single state $s \in S(j)$:

$$\nabla'(i, s \mid \omega) = \frac{\Delta(i \mid \omega \cup s)}{\Delta(i \mid \omega)}$$

We also introduce the adaptive total primal curvature (ATPC), which intuitively captures the total change in the marginal value from a partial realization ω to another (ω') such that $\text{dom}(\omega) \subset \text{dom}(\omega')$.

Definition 6 (Adaptive Total Primal Curvature)**.** *Let $dom(\omega) \subset dom(\omega')$ and let $\omega \to \omega'$ represent the set of possible state sequences leading from ω to ω'. Then the ATPC is*

$$\Gamma(i \mid \omega', \omega) = \mathbb{E}\left[\prod_{s_j \in R} \nabla'(i, s_j \mid \omega \cup \{s_1, \ldots, s_{j-1}\}) \,\middle|\, R \in \omega \to \omega'\right]$$

In the following, we derive the approximation ratio of AReST in two steps. First, we provide a relation between the optimal policy and greedy policy of any size. Then, we prove the approximation ratio based on the bound.

Denote π_k^* as the optimal policy of size k, and π_l^g as the greedy policy of size l. We introduce the following auxiliary Lemmas before the main result.

Lemma 4.

$$\Gamma(i \mid \omega', \omega) = \frac{\Delta(i \mid \omega')}{\Delta(i \mid \omega)}$$

Proof. Fix a sequence $R \in \omega \to \omega'$ of length r and $R = \{s_1, \ldots, s_r\}$. Then, expanding the product we obtain

$$\frac{\Delta(i \mid \omega \cup \{s_1\})}{\Delta(i \mid \omega)} \cdot \frac{\Delta(i \mid \omega \cup \{s_1, s_2\})}{\Delta(i \mid \omega \cup \{s_1\})} \cdots \frac{\Delta(i \mid \omega')}{\Delta(i \mid \omega' \setminus \{s_{r-1}\})}$$

Notice that for any sequence in $\omega \to \omega'$, the product reduces to the same ratio. From this, we directly obtain the statement of the lemma. □

From Lemma 4, we can also see that $\Gamma(i \mid \omega', \omega)$ is independent of the order in which elements of R are considered. Therefore, we can treat R as simply a set rather than a sequence. Further, we can derive an upper bound on all Γ in terms of $\text{accept}(u \mid \omega)$.

Lemma 5. *$\Gamma(i \mid \omega', \omega) \leq \delta, \forall i, \omega', \omega$ where*

$$\delta = \max_{u,\omega,\omega'} \frac{accept(u \mid \omega)}{accept(u \mid \omega')}$$

Proof. For any i, ω', ω, we have

$$\Gamma(i \mid \omega', \omega) = \frac{\text{accept}(i \mid \omega')\Delta B(i \mid \omega')}{\text{accept}(i \mid \omega)\Delta B(i \mid \omega)} \leq \delta \frac{\Delta B(i \mid \omega')}{\Delta B(i \mid \omega)}$$

where $\Delta B(u \mid \omega)$ is the marginal gain of successfully befriending u in partial realization ω. $\Delta B(u \mid \omega') \leq \Delta B(u \mid \omega)$ by definition, and therefore $\Gamma(i \mid \omega', \omega) \leq \delta$. □

Corollary 2. *For the ETC acceptance function, $\delta = O(1)$.*

Proof. Recall that the ETC acceptance function is defined as:

$$\text{accept}(u) = \rho_1 \log(\mathbb{E}[|N(u) \cap N(s)|] + 1) + \rho_0$$

Thus, for any u, $\min_\omega \text{accept}(u \mid \omega)$ is achieved in all partial realizations that guarantee $|N(u) \cap N(s)| = 0$ and $\max_\omega \text{accept}(u \mid \omega) \leq 1$. Thus, $\rho_0 \leq \text{accept}(u \mid \omega) \leq 1, \forall \omega$. So we have:

$$\begin{aligned}\delta &\leq \frac{\max_\omega \text{accept}(u \mid \omega)}{\min_\omega \text{accept}(u \mid \omega)} \\ &\leq \frac{1}{\rho_0}\end{aligned}$$

As ρ_0 is a constant, $\delta = O(1)$ for the ETC acceptance function.

In the following, we prove the approximation ratio of AReST. We first bound the arbitrary marginal gain by the marginal gain in the greedy solution utilizing ATPC, then we relate the optimal solution with the greedy solution and eventually derive the approximation ratio.

Lemma 6.

$$\Delta(i|\omega') \leq \delta \Delta(g_{l+1}|\omega)$$

where ω denotes an arbitrary partial realization that can be the result of the greedy policy π_l^g that sent l requests, ω' is any realization that ω is a subrealization of, i is an element not in $dom(\omega')$ and g_{l+1} denotes the $(l+1)th$ step of the greedy policy.

Proof. By Lemma 4, we have $\Delta(i|\omega') = \Gamma(i|\omega', \omega)\Delta(i|\omega)$. By Lemma 5, $\Gamma(i|\omega', \omega) \leq \delta$. By property of the greedy policy, $\Delta(i|\omega) \leq \Delta(g_{l+1}|\omega)$. Combining all the facts together guarantees the desired result. □

Now we are ready to relate the greedy policy to the optimal policy.

Theorem 3.

$$f_{avg}(\pi_k^*) - f_{avg}(\pi_l^g) \leq k\delta(f_{avg}(\pi_{l+1}^g) - f_{avg}(\pi_l^g))$$

where $f_{avg}(\pi) = \mathbb{E}[f(E(\pi, \Phi), \Phi)]$ and $E(\pi, \Phi)$ denotes the observed states of poilcy π in a random realization Φ.

Proof. As the optimal policy selects k elements, the difference between $f_{\text{avg}}(\pi_k^*)$ and $f_{\text{avg}}(\pi_l^g)$ is upper bounded by the marginal gain of sending the k extra requests of the optimal policy on top of the l requests in the greedy policy. By Lemma 6, the marginal gain of adding each element i is bounded for each partial realization that can be a result of the greedy policy π^g_l. Thus:

$$\begin{aligned}
&f_{\text{avg}}(\pi^*_k) - f_{\text{avg}}(\pi^g_l) \leq \mathbb{E}\left[k\delta\Delta(g_{l+1} \mid \omega) \mid \omega\right] \\
&= k\delta\mathbb{E}\left[\Delta(g_{l+1} \mid \omega) \mid \omega\right] \\
&= k\delta\mathbb{E}\left[\mathbb{E}\left[f(\text{dom}(\omega) \cup G', \Phi) - f(\text{dom}(\omega), \Phi)) \mid \Phi \sim \omega\right] \mid \omega\right] \\
&= k\delta\mathbb{E}\left[f(E(\pi^g_{l+1}), \Phi), \Phi) - f(E(\pi^g_l, \Phi), \Phi))\right] \\
&= k\delta(f_{\text{avg}}(\pi^g_{l+1}) - f_{\text{avg}}(\pi^g_l))
\end{aligned}$$

where $G' = \{g_{l+1}\}$.

□

Denote OPT as the optimal solution to the Min-Friending problem and $\Lambda_k = k\delta$, we have the following theorem on the performance of AReST.

Theorem 4. *AReST can reach the benefit of at least $Q - 1$ with approximation ratio of $(\delta \ln Q)$.*

Proof. Based on Theorem 3, for any j, we have:

$$f_{\text{avg}}(OPT) \leq \sum_{i=1}^{j} (f_{\text{avg}}(\pi^g_{i+1}) - f_{\text{avg}}(\pi^g_i)) + \delta f^{j+1}_{\text{avg}} \Lambda_{|OPT|}$$

$$f_{\text{avg}}(OPT) \leq \sum_{i=1}^{j} \delta f^i_{\text{avg}} + \delta f^{j+1}_{\text{avg}} \Lambda_{|OPT|} \tag{6}$$

where $\delta f^{i+1}_{\text{avg}} = f_{\text{avg}}(\pi^g_{i+1}) - f_{\text{avg}}(\pi^g_i)$. (Note that $f_{\text{avg}}(\pi^g_0) = 0$) By similar technique as in [17], we can multiply both sides of (6) by $(1 - \frac{1}{\Lambda_{|OPT|}})^{l-j}$ for any l and sum both sides from $j = 1$ to $j = l$, which gives the lhs as:

$$\Lambda_{|OPT|} \left(1 - (1 - \frac{1}{\Lambda_{|OPT|}})^l\right) f_{\text{avg}}(OPT)$$

For rhs, the coefficient of δf^j_{avg} is in the form of:

$$\left(\Lambda_{|OPT|}(1 - \Lambda^{-1}_{|OPT|})^{l-j} + \sum_{i=j+1}^{l} (1 - \Lambda^{-1}_{|OPT|})^{l-i}\right)$$

and can be simplified to $\Lambda_{|OPT|}$.

Thus, the rhs is exactly $\Lambda_{|OPT|} f_{\text{avg}}(\pi^g_l)$ and we have

$$\left[1 - \left(1 - \frac{1}{\Lambda_{|OPT|}}\right)^l\right] f_{\text{avg}}(OPT) \leq f_{\text{avg}}(\pi^g_l) \tag{7}$$

Now, we consider the problem of finding the minimum l such that $f_{\text{avg}}(\pi^g_l)$ is at least $f_{\text{avg}}(OPT) - c$ where c is a constant. Thus, we need to solve:

$$\left[1 - \left(1 - \frac{1}{\Lambda_{|OPT|}}\right)^{l^*}\right] f_{\text{avg}}(OPT) = f_{\text{avg}}(OPT) - c$$

Notice that the minimum of l, l^*, is the size of the solution outputted by AReST. Rearranging terms and taking log on both side gives:

$$-l^* \ln(1 - \frac{1}{\Lambda_{|OPT|}}) = \ln f_{\text{avg}}(OPT) - \ln c$$

Using the fact that $\ln(1 + x) < x$ (when $x \neq 0$), we have

$$\frac{l^*}{\Lambda_{|OPT|}} < \ln f_{\text{avg}}(OPT) - \ln c$$

Selecting $c = 1$ gives

$$\frac{l^*}{|OPT|} < \left(\frac{\Lambda_{|OPT|}}{|OPT|} \ln f_{\text{avg}}(OPT)\right) \leq O(\delta \ln Q) \quad □$$

Notice that the result is bi-criteria as $f_{\text{avg}}(\pi^g_{l^*})$ is one less than $f_{\text{avg}}(OPT)$ by the selection of c. The gap is hard to remove as f is non-submodular and the value $f_{\text{avg}}(OPT)$ may be reached after arbitrary number of greedy selections starting from l^*.

Corollary 3. *For the ETC acceptance model, AReST achieves an approximation ratio of $\rho_0^{-1} \ln Q = O(\ln Q)$ with benefit at least $Q - 1$.*

Proof. The corollary immediately follows when combining Corollary 2 and Theorem 4. □

5 Experimental Evaluations

Having established our model for a near-optimal attacker, we now apply it to understand the structure of vulnerability on online social networks. In addition to the broad, overarching question of *how vulnerable are OSNs to socialbot attacks?* we also wish to understand the impact of user behavior and the attacker's priorities and knowledge level. To this end, we structure our experiments as follows. First, we introduce the user and attacker models we use in our experiments (Sec. 5.1). We next apply these to the networks listed in Table 1 and examine how vulnerability changes as a function of the model used (Sec. 5.2). In particular, we find that user behavior is a dominant factor in determining the attacker's success across all attack models. Beyond this, it generally appears to be easiest to conduct *untargeted* or *structured* attacks as within each user model these kinds of attacks achieve the greatest success.

5.1 Models Used

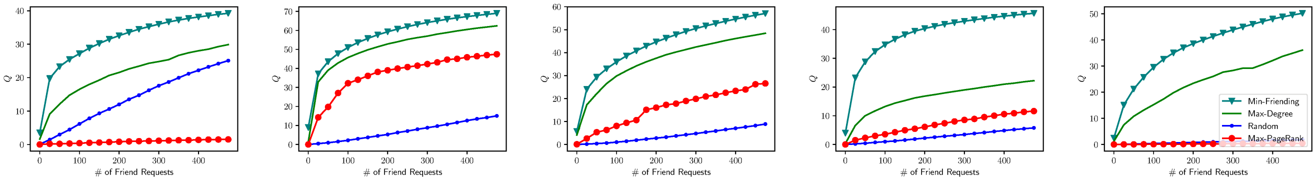
We begin by presenting our model of user behavior. We consider three models in total: the fixed-probability and Expected Triadic Closure (ETC, eqn. 1) models described in previous sections, and the Expected Shared Neighbors (ESN) model:

$$\text{accept}(u) = \mathbb{E}\left[\frac{\mid N(u) \cap N(s) \mid}{\mid N(u) \cup N(s) \mid}\right] \tag{8}$$

This model describes users as increasingly more likely to accept a friend request based on the number of shared friends, normalized by the sum of friends of u and s. Intuitively, this matches the expectation that users want to see mutual friends before accepting a friend request, but also penalize simply befriending users in bulk. We additionally model a commonly-used bootstrapping strategy in the ESN and ETC settings. This strategy prioritizes high-degree users early in the attack because they have been observed to have significantly higher acceptance rates. This is useful for evading automated detection by increasing the proportion of friend requests accepted. We model this with the function

$$\text{di}(u) = \left(\frac{\mathbb{E}\left[\deg(u)\right]}{M}\right)^5$$

where $M = \max_{u \in V} \mathbb{E}\left[\deg(u)\right]$ is the maximum expected degree of any node in V. This term is added to



(a) Facebook (b) Enron Email (c) Slashdot (d) Twitter (e) DBLP

Fig. 3: Mean number of friend requests sent before reaching benefit threshold Q on each dataset using the ETC user model and Structured Set attack model. Note that although the parameter we vary is Q, we flip the axes to simplify interpretation.

Network	Nodes	Edges	Kind
Facebook	4k	88k	Social
Enron Email	37k	184k	Communication
Slashdot	77k	905k	Social
Twitter	81k	1.77M	Social
DBLP	317k	1.05M	Collaboration

TABLE 1: The networks used in our simulations. All networks are from SNAP [22].

the acceptance probability in the ESN and ETC models and modifies the acceptance probability such that nodes with high degree begin with near-guaranteed acceptance (as observed in prior work [2], [7]) while having virtually no impact on the bulk of nodes with relatively small degree.

We now turn our attention to the attacker models we consider. Five models are used to explore the variation in vulnerability due to the goals of the attacker.

Untargeted. In this model, the attacker is indiscriminate; interested only in extracting as much private content as possible from the network. This simple model is used as a baseline.

Individual. The attacker is interested only in information about a specific individual. This corresponds to an attacker using socialbots for stalking or attempting to mine private information to use for blackmail.

Unstructured Set. The attacker is interested in information pertaining to a target set of users T, where the selection of T is not based on network topology. This kind of attack would be seen when the attacker selects targets based on criteria orthogonal to social ties. We model this by selecting T uniformly at random from V, with $|T|$ fixed as 100.

Structured Set. The attacker seeks information related to users that are socially interconnected, and uses noisy data from the targeted network to collect this information. We model this kind of set by selecting a user at random, then performing a stochastic breadth-first search to collect up to 100 users. The breadth-first search is stochastic in the sense that whether the BFS traverses an edge or not is determined by the edge probability p_{uv}. This traversal happens *independently* of the state of the edge in the simulation (e.g. the BFS may traverse an edge (u, v) that is not present when the socialbot reveals the edges around u in the simulation).

Community. The attacker seeks information as in the Structured Set setting, and bases their target list on an ground-truth selection of socially related users. For example, an attacker may target an organization by scraping LinkedIn profiles or from illicitly obtained employment rolls. We model this by using the ground-truth communities on the DBLP network as target sets.

Lastly, we consider the possibility of the attacker prioritizing certain users within the target set (note that for the untargeted setting we can equivalently take $T = V$). If the attacker assigns a user u priority w_u, then we write the benefit functions as $B_f(u) = \mathbf{1}_{u \in T} w_u$, $B_{fof}(u) = \mathbf{1}_{u \in T} w_u/2$, and $B_e(u, v) = 2^{\mathbf{1}_{u \in T} + \mathbf{1}_{v \in T}} w_u w_v / M$, where $\mathbf{1}_p$ is the indicator function taking value 1 when predicate p is true and 0 otherwise. We examine three *Benefit Models*: **Unweighted** ($w_u = 1$), weighted by a sense of **Naïve Importance** ($w_u = \mathbb{E}[\deg(u)]$), and weighted according to **External Priorities** (modeled as weights being distributed according to a uniform distribution on $[1, 10]$).

5.2 Vulnerability Analysis

We now turn to the results of our simulations. We compare against several simple heuristics as a baseline: randomly selecting users to friend (Random), greedily selecting users with maximum expected degree (Max-Degree), and greedily selecting users with maximum PageRank (Max-PageRank). Each result presented is the mean of 500 independent runs of the associated algorithm. Due to space constraints, we cannot give a complete representation of the vulnerability of each combination of settings.

5.2.1 Efficiency of the AReST Model

Our analysis begins with a comparison of AReST to the heuristic baselines in Fig. 3. It is easy to see that AReST outperforms each, and often does so by a large margin. As one might expected, the performance of the Random heuristic decreases as network size increases. However, we can see that the performance of both the Max-PageRank and Max-Degree heuristics relative to AReST varies independently of network size or density. In particular, the performance of AReST displays complex behavior with respect to topology: the most difficult (Facebook; 500 requests needed to reach 40 benefit) and second most difficult (Twitter; 200 requests needed to reach similar benefit) networks are dramatically different sizes – but share the common factor that both are popular OSNs. The behavior we observe in Fig. 3 indicates that topology may be an important factor in determining the difficulty of an attack.

Difficulty, for the purposes of our analysis, is defined to be the time required to obtain some level of benefit. If the bot can obtain large amounts of benefit very quickly, we say that the attack it performs is easy. Similarly, if it requires a significant length of time (corresponding to many

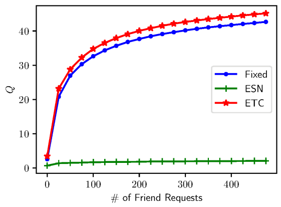


Fig. 4: Benefit gained by the bot as we vary the assumed acceptance model. The true acceptance model is fixed to ETC.

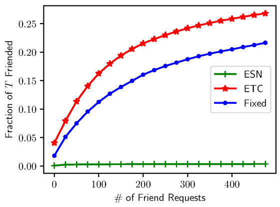


Fig. 5: Fraction of T friended under each assumed acceptance model. The true acceptance model is fixed to ETC.

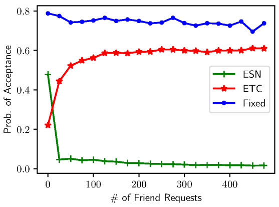


Fig. 8: Mean acceptance probability of the node targeted at each step of the AReST algorithm.

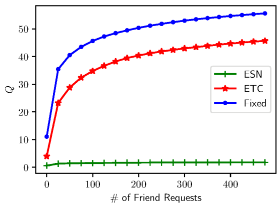


Fig. 9: Benefit obtained under each acceptance model, assuming the bot knows the true model.

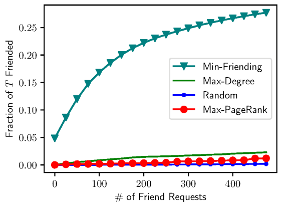


Fig. 6: Fraction of T friended under ETC acceptance and Structured Set settings on Twitter.

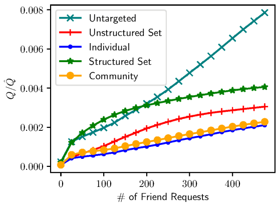


Fig. 7: Performance of AReST as the target structure is varied on DBLP.

friend requests) to obtain benefit, we say the attack is hard. Visually, an easy attack should have a steep slope upwards initially, while a difficult attack will have a lower slope.

Next, we examine the importance having an accurate model of user behavior. To model this scenario, we set the underlying acceptance model to the ETC model described previously. This matches the literature on real-world acceptance probabilities [2]. Then, the bot makes decisions using one of our three models but the true acceptance probability is determined using the underlying model. From Fig. 4 we can see that incorrectly assuming that user acceptance follows the fixed-probability model has little impact on overall performance. However, this is tempered by the results shown in Fig. 5: assuming the fixed model dramatically reduces the fraction of the target set befriended. This is caused by the bot aiming for valuable, untargeted users in the fixed-probability case merely due to their high probability of accepting requests (often coupled with a large value from revealing edges). On the other hand, when the bot knows the true model, it is able to exploit the increasing probabilities to more quickly infiltrate the target set. This experiment implicitly compares to prior simulation work, which assume a constant acceptance model [11], [12].

Continuing along this line of analysis, we compare the source of benefit gained by each heuristic. Fig. 6 illustrates that the heuristics, which do not take into account the target set T, perform poorly at the task of infiltrating T. This is an unsurprising result. This further supports the above interpretation of the benefit gained in the mis-applied fixed-probability model: we see from Fig. 3 that the Max-Degree heuristic in general performs relatively well as measured by our objective. However, Fig. 6 likewise indicates that little of this benefit comes from befriending targeted users. By the process of elimination, we conclude that these heuristics are primarily obtaining benefit from revealing edges on the network from requests to high-degree users and not by befriending the target set.

Finally, we conclude this subsection by examining the mean acceptance probability of each request under each acceptance model. Figures 8 shows even though the mean acceptance probability on the Twitter network is 0.5 under the Fixed-Probability model, the average request remains near 0.8 for the course of the run. We see a similar bias towards highly-likely requests in the ETC model. However, this model begins with a uniform probability of $\rho_0 \approx 0.18$ for each user — thus, it is clear that a socialbot run by AReST crawls along from an initial position, exploiting the friendships it has already made to improve the likelihood of further accepted requests. Before moving on, we note that the ESN model performs atrociously: the likelihood of a user accepting a request rapidly approaches 0. This appears to be due to the slow growth of acceptance probabilities under this model, leading to the bot rapidly exhausting the reasonably likely requests given by the degree incentive and then floundering with many low-probability requests being rejected as it fails to gain a foothold. This view is supported by the performance shown in Figures 4 & 9, where the bot gains nearly no benefit under the ESN model.

5.2.2 *Factors Influencing Attack Difficulty*

In the previous sub-section, we noted several factors that may influence the difficulty of the reconnaissance attack from the perspective of the attacker. In this section, we explore those factors in greater detail. For this purpose, we use the *vulnerability* of each setting, measured as the expected number of friend requests to obtain 1% of network benefit, as a means to understand the difficulty. This normalization allows more direct comparison across target settings, since each has a different amount of benefit available. Table 2 shows overall scores.

The intractability of attacking the ESN model is again immediately obvious. However, this observation is tempered by the fact that it is not a model based on observed user behavior. When examining small networks, it seems that when users follow the ETC acceptance model, their private content is more vulnerable. However, on larger networks this property vanishes. However, these numbers are promising from a defense perspective. In every case, an attacker seeking to attack an individual or community requires more time than for attacking either topologically-close users or

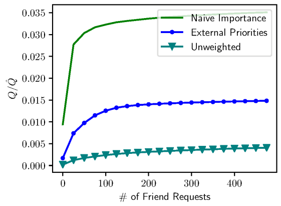


Fig. 10: Vulnerability of DBLP under each benefit model under ETC targeting a Structured Set.

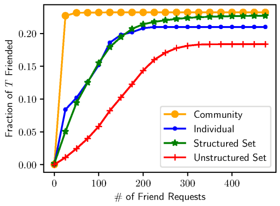


Fig. 11: Fraction of T friended under the ETC acceptance and varying target settings.

Enron Email	Fixed	ESN	ETC
Untargeted	35.08	139.61	31.55
Individual	169.47	677.49	153.45
Unstructured Set	46.20	387.21	67.55
Structured Set	194.19	38.03	53.63
Slashdot	Fixed	ESN	ETC
Untargeted	43.32	468.40	47.17
Individual	243.83	2767.71	271.69
Unstructured Set	63.60	1305.19	126.75
Structured Set	73.98	471.72	76.96
Facebook	Fixed	ESN	ETC
Untargeted	25.63	831.05	43.74
Individual	236.74	7055.66	355.80
Unstructured Set	54.68	3481.76	109.08
Structured Set	84.21	1766.10	109.108
Twitter	Fixed	ESN	ETC
Untargeted	41.53	2897.87	84.31
Individual	446.39	33208.39	935.10
Unstructured Set	43.63	12142.53	128.43
Structured Set	54.85	2077.276	58.62
DBLP	Fixed	ESN	ETC
Untargeted	264.68	4899.80	621.74
Individual	1027.81	15494.18	2370.87
Unstructured Set	678.17	14994.75	1646.89
Structured Set	728.41	12933.14	1277.30
Community	999.43	15490.64	2217.69

TABLE 2: *Normalized Resistance Score* (mean number of friend requests required to earn 1% of the maximum benefit on the network) for each combination of user and attacker models with the unweighted benefit model on each network. Larger numbers indicate higher resistance.

the network as a whole. This is supported further by Figure 7. Figure 11 indicates that successfully infiltrating a user under the ETC model only happens roughly 1 in 5 times, and require hundreds of friend requests to reach that point. On the other hand, infiltrating a community can be done relatively rapidly and quickly reaches saturation due to failed requests blocking off the remaining portions of the community.

Tables 3 and 4 give insight into the vulnerability as the attackers' priorities vary. From Table 3 we see that when an adversary aims for "important" users, very little changes from the unweighted case. However, when the attacker has external priorities towards individual users within the target set, the users are notably more vulnerable. We note that under the External-Priorities model, a greater portion of the benefit is tied to members of the target set than either of the other two models, which indicates that the more focused an attacker is on their targets, the more vulnerable those targets are. This does not appear to include the case of an *Individual* target, which would seem to indicate that regardless of how much an individual is prioritized (beyond a certain point, which is reached by the Unweighted priority model) there is a limit on the ability of an attacker to successfully automate the theft of private content. However, this must be taken with a grain of salt as Fig. 10 indicates that this pattern does not hold on the Structured Set setting on DBLP.

This leads us to a concrete set of properties that influence the vulnerability of users' private content on an OSN. First – and most impactful – is the users' behavior, which indicates that effective user education directed at altering this behavior may be the most effective means of reducing vulnerability. However, users have proven to be notoriously difficult to train away from poor security practices (e.g. Dhamija & Perrig found that "the level of security training did not prevent users from choosing trivial passwords or from storing them insecurely" [23]), so the feasibility of this approach is somewhat suspect. Next, we observe that network topology plays a significant role. Facebook, a topology drawn from a closed network, is less vulnerable than a topology drawn from an open network (Slashdot and Twitter) when an attacker is interested in topologically co-located users. This indicates that perhaps restructuring OSNs to encourage the growth of certain topologies could produce safer environments for OSN users. Finally, the structure of the target set plays a significant role. Both attacking the network as a whole and attacking many co-located users are more efficient than alternatives. Based on this, we conclude that defenses to these kinds of attacks are most pressing.

6 Conclusion

In this paper, we present a new paradigm to quantify the OSN privacy vulnerability with respect to socialbot attacks. Specifically, we introduce a new optimization problem, namely Min-Friending, which identifies a minimum CFS to friend with in order to obtain at least Q benefits. We show that Min-Friending is inapproximable within a factor of $(1 - o(1)) \ln Q$ and present an adaptive approximation algorithm which has a tight performance bound of $(1 + \ln Q)$ using adaptive stochastic optimization, when the friend request acceptance rate is constant. The key feature of our solution lies in the adaptive method, where partial network topology is revealed during each successful friend request. We further generalize our analysis to cope with a more practical friend request acceptance model, which requires us to introduce a novel theoretical tool to analyze the adaptive non-submodular greedy minimization. Extensive experiments not only confirm the performance of our algorithm, but also provide new insights towards privacy protection under socialbot attacks.

Acknowledgments

This work is supported in part by the NSF grant #CCF-1422116 and DTRA HDTRA1-14-1-0055.

Twitter	Fixed	ESN	ETC
Untargeted	41.55	3023.58	83.98
Individual	446.69	32234.25	937.17
Unstructured Set	43.50	13423.26	127.94
Structured Set	54.31	2478.89	56.85

TABLE 3: Normalized Resistance Scores under the Naïve-Importance benefit model.

Twitter	Fixed	ESN	ETC
Untargeted	44.90	3299.76	91.60
Individual	436.29	33540.81	936.96
Unstructured Set	15.88	6612.41	51.61
Structured Set	19.71	838.65	20.42

TABLE 4: Normalized Resistance Scores under the External-Priorities benefit model.

References

[1] Ed Novak and Qun Li. A survey of security and privacy in online social networks. *College of William and Mary Computer Science Technical Report*, 2012.

[2] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov, and Matei Ripeanu. The Socialbot Network: When Bots Socialize for Fame and Money. In *Proceedings of the 27th Annual Computer Security Applications Conference*, ACSAC '11, pages 93–102. ACM, 2011.

[3] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions – I. *Mathematical Programming*, 14(1):265–294, 1978.

[4] Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, pages 427–486, 2011.

[5] Inkyung Jeun, Youngsook Lee, and Dongho Won. A Practical Study on Advanced Persistent Threats. In *Computer Applications for Security, Control and System Engineering*, number 339 in Communications in Computer and Information Science, pages 144–152. Springer Berlin Heidelberg, 2012.

[6] Thomas Ryan and G Mauch. Getting in bed with robin sage. In *Black Hat Conference*, 2010.

[7] Aviad Elyashar, Michael Fire, Dima Kagan, and Yuval Elovici. Homing socialbots: intrusion on a specific organization's employee using socialbots. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1358–1365. ACM, 2013.

[8] Yazan Boshmaf, Konstantin Beznosov, and Matei Ripeanu. Graph-based sybil detection in social and information systems. In *Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on*, pages 466–473. IEEE, 2013.

[9] Gang Wang, Tristan Konolige, Christo Wilson, Xiao Wang, Haitao Zheng, and Ben Y Zhao. You are how you click: Clickstream analysis for sybil detection. In *USENIX Security Symposium*, volume 9, pages 1–008, 2013.

[10] Kuan Zhang, Xiaohui Liang, Rongxing Lu, Kan Yang, and Xuemin Sherman Shen. Exploiting mobile social behaviors for sybil detection. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 271–279. IEEE, 2015.

[11] Xiang Li, J. David Smith, and My T. Thai. Adaptive reconnaissance attacks with near-optimal parallel batching. In *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*. IEEE, 2017.

[12] Hung T. Nguyen and Thang N. Dinh. Targeted Cyber-attacks: Unveiling Target Reconnaissance Strategy via Social Networks. In *Proceedings of the IEEE Int Conf. on Computer Com., Security and Privacy in BigData Workshop*, INFOCOM BigSecurity 2016, 2016.

[13] Abigail Paradise, Asaf Shabtai, and Rami Puzis. Hunting Organization-Targeted Socialbots. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, pages 537–540, New York, NY, USA, 2015. ACM.

[14] T. N. Dinh, Y. Xuan, M. T. Thai, P. M. Pardalos, and T. Znati. On New Approaches of Assessing Network Vulnerability: Hardness and Approximation. *IEEE/ACM Transactions on Networking*, 20(2):609–619, 2012.

[15] Md Abdul Alim, Nam P. Nguyen, Thang N. Dinh, and My T. Thai. Structural Vulnerability Analysis of Overlapping Communities in Complex Networks. In *Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01*, WI-IAT '14, pages 5–12. IEEE Computer Society, 2014.

[16] Sebastian Neumayer, Gil Zussman, Reuven Cohen, and Eytan Modiano. Assessing the vulnerability of the fiber infrastructure to disasters. *Networking, IEEE/ACM Transactions on*, 19(6):1610–1623, 2011.

[17] Zengfu Wang, Bill Moran, Xuezhi Wang, and Quan Pan. Approximation for maximizing monotone non-decreasing set functions with a greedy method. *Journal of Combinatorial Optimization*, 31(1):29–43, 2016.

[18] Michael Fire, Lena Tenenboim, Ofrit Lesser, Rami Puzis, Lior Rokach, and Yuval Elovici. Link prediction in social networks using computationally efficient topological features. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 73–80. IEEE, 2011.

[19] Michael Fire, Rami Puzis, and Yuval Elovici. Link prediction in highly fractional data sets. In *Handbook of computational approaches to counterterrorism*, pages 283–300. Springer, 2013.

[20] Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 635–644. ACM, 2011.

[21] Uriel Feige. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

[22] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data, June 2014.

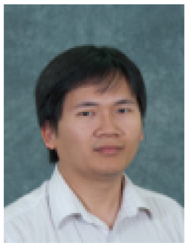
[23] Rachna Dhamija and Adrian Perrig. Deja Vu – A User Study: Using Images for Authentication. In *USENIX Security Symposium*, volume 9.

Xiang Li received the M.Sc. degree from the Academy of Mathematics Systems and Science, Chinese Academy of Sciences, and the M.Sc. degree in industrial and systems engineering from the University of Florida, where she is currently pursuing the Ph.D. degree with the Department of Computer and Information Science and Engineering. Her current research interests include security on online social networks, Network Science, Approximation Algorithms, Optimization.

J. David Smith received his Bachelors of Science degree from the University of Kentucky. He is currently working towards his PhD degree in Computer Science at CISE Department, University of Florida, USA under the supervision of Dr. My T. Thai. His current research interests are in security on online social networks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TETC.2018.2840433, IEEE Transactions on Emerging Topics in Computing

Tianyi Pan is currently working toward the PhD degree in the Department of Computer and Information Science and Engineering, University of Florida, under the supervision of Dr. My T. Thai. His research focuses on optimization problems in online social networks, smart grids and cellular networks.



Thang N. Dinh (S'11-M'14) received the Ph.D. degree in computer engineering from the University of Florida in 2013. He is an Assistant Professor at the Department of Computer Science, Virginia Commonwealth University. His research focuses on security and optimization challenges in complex systems, especially social networks, wireless and cyber-physical systems. He serves as PC chair of COCOON???16 and CSoNet???14 and on TPC of several conferences including IEEE INFOCOM, ICC, GLOBECOM, and SOCIALCOM. He is also an associate editor of Computational Social Networks journal and a guest-editor for Journal of Combinatorial Optimization.

My T. Thai (M'06) received the Ph.D. degree in computer science from the University of Minnesota, in 2005. She is a Professor with the Department of Computer and Information Science and Engineering, University of Florida. Her current research interests include algorithms and optimization on network science and engineering. She was a recipient of several research awards, including a UF Provosts Excellence Award for Assistant Professors, a DoD YIP, and an NSF CAREER Award. She has engaged in many professional activities, such as being the PC Chair of the EEE IWCMC 2012, the IEEE ISSPIT 2012, and COCOON 2010. She is the Founding Editor-in-Chief of *Computational Social Networks*, an Associate Editor of *JOCO* and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, and a Series Editor of *SpringerBriefs* in *Optimization*.