Attacking Similarity-Based Sign Prediction

Michał Tomasz Godziszewski
and Tomasz P. Michalak
University of Warsaw
m.godziszewski@uw.edu.pl
tpm@mimuw.edu.pl

Marcin Waniek
and Talal Rahwan
NYU Abu Dhabi
mjwaniek@nyu.edu
talal.rahwan@nyu.edu

Kai Zhou
and Yulin Zhu
Hong Kong Polytechnic University
kaizhou@polyu.edu.hk
yulin.zhu@polyu.edu.hk

***Abstract*—In this paper, we present a computational analysis of the problem of attacking sign prediction, whereby the aim of the attacker (a network member) is to hide from the defender (an analyst) the signs of a target set of links by removing the signs of some other, non-target, links. The problem turns out to be NP-hard if either local or global similarity measures are used for sign prediction. We propose a heuristic algorithm and test its effectiveness on several real-life and synthetic datasets.**

I. Introduction

A number of works have recently studied how the graphs can be manipulated to make unwarranted information gathering harder. The literature focused, in particular, on the adversarial analysis of centrality measures [1]–[4], link prediction algorithms [5]–[7] and community detection algorithms [2], [8], mostly in the context of simple networks.Thus far, however, problems of this kind have not been studied in the context of signed networks. In these networks, links are labelled with plus and minus signs representing positive and negative relations between the nodes. Signed networks are often used to model social networks [9], where positive links indicate friendship/support and the negative ones—antagonism/opposition.

A particularly interesting challenge in the context of signed networks is the problem of sign prediction. It involves determining whether a particular link (the label of which is so far unknown or missing) should be assigned a positive or negative label [10]–[13]. Sign prediction can be viewed as analogous to the problem mentioned above of link prediction in unsigned networks, where the aim is to anticipate the existence of links that are missing from the data or that are yet to be created [14]. A large class of methods for link prediction builds upon local and global measures of node similarity in networks. The former class of similarity measures focuses on the *common neighborhood* of two seemingly disconnected nodes. In turn, global similarity measures take into account the entire network. Measuring similarity between nodes becomes more challenging in signed networks [15], as not only the size of the common neighborhood plays the role, but also the composition of "friends" and "foes" within it.

Against this background, we present the first computational analysis of attacking sign prediction, where we focus on the case of undirected signed networks. We formalize five computational problems in which the attacker aims to hide or obfuscate from the network analyst the signs of links in a fixed target set by removing the signs of some non-target links. We prove that all these problems are NP-hard for both local and global similarity measures in their unrestricted versions.

Given the computational intractability of the problems, we propose a variant of a heuristic algorithm for evading local similarity-based link prediction in signed networks, the Tally heuristic. We evaluate the heuristic on real-life and synthetic network datasets and show their effectiveness.

II. Background

Let graph $G = (V, E)$ represent an (unsigned) social network, where V is the set of nodes, E is the set of links. A *similarity measure* is a function that assigns to any pair of nodes a real value that reflects the similarity between them. Two main types of similarity measures can be distinguished: *local* and *global*. The former ones focus on the direct neighborhood of the nodes in question. In contrast, global similarity measures take into account the entire graph. An example is the Katz measure [16].

In this paper we focus on similarity measures defined for signed social networks that are represented by graphs in which links additionally have labels, either positive or negative. Formally, a *signed social network* is a graph $G = (V, E, \sigma)$, where $\sigma : E \to \{+, -\}$ is a *sign function* on the links. Furthermore, let us denote by $N_+(v)$ and $N_-(v)$ the positive and negative neighborhood of a node $v \in V$, i.e.:

$$N_+(v) = \{w \in V : \{v, w\} \in E \,\&\, \sigma(\{v, w\}) = +\}, \quad \text{and}$$
$$N_-(v) = \{w \in V : \{v, w\} \in E \,\&\, \sigma(\{v, w\}) = -\}.$$

Given these definitions, we can introduce a notion of **similar common neighborhood** of $u, v \in V$, denoted $c_s(u, v)$, i.e. the set of nodes adjacent to both u and v connected to them with the links of the same signs. Formally:

$$c_s(u, v) = (N_+(u) \cap N_+(v)) \cup (N_-(u) \cap N_-(v)).$$

Analogously, let $c_d(u, v)$ denote the **dissimilar common neighborhood** of $u, v \in V$, i.e. the set of nodes connected to both u and v with the links of the opposite signs. Formally:

$$c_d(u, v) = (N_+(u) \cap N_-(v)) \cup (N_-(u) \cap N_+(v)).$$

Let d_v denote the degree of the node v, i.e., the number of v's neighbors. We will also use $d_+(v)$ $(d_-(v))$ to denote the number of the neighbors of v with which v has a positive (negative) connection, i.e. $d_+(v) = |N_+(v)|$ $(d_-(v) = |N_-(v)|)$. Let the positive (negative) preferential

2374-8486/21/$31.00 ©2021 IEEE
DOI 10.1109/ICDM51629.2021.00173

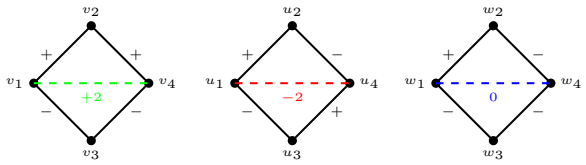

Fig. 1: Three sample signed networks. The weights of the colored dashed links are the SCN scores of the pairs of nodes $\{u_1, u_4\}$, $\{v_1, v_4\}$, and $\{w_1, w_4\}$. In contrast, the unsigned counterpart of SCN, i.e. the Common Neighborhood measure, outputs a score equal to 2 in all the three cases.

attachments between u and v be denoted as $\mathrm{PA}_+(u,v) = d_+(u) \cdot d_+(v)$ ($\mathrm{PA}_+(u,v) = d_-(u) \cdot d_-(v)$).

While there are many local similarity measures for unsigned networks [17], only recently some of their counterparts for signed networks have been analyzed in the literature [15], [18]:

- **Signed Common Neighborhood** (SCN, see [19], [15]):

$$\begin{aligned}\mathrm{SCN}(u,v) &= |c_s(v,u)| - |c_d(v,u)| \\ &= (|N_+(u) \cap N_+(v)| + |N_-(u) \cap N_-(v)|) + \\ &- (|N_+(u) \cap N_-(v)| + |N_-(u) \cap N_+(v)|)\,.\end{aligned}$$

- **Signed Jaccard** (SJ, see [15]):

$$\mathrm{SJ}(u,v) = \frac{|c_s(u,v)| - |c_d(u,v)|}{|N(u)_+ \cup N(u)_- \cup N_+(v) \cup N_-(u)|}.$$

To illustrate the difference between the above two measures for signed networks and their counterparts for unsigned networks let us consider the graphs in Figure 1. Let us first disregard signs of links as standard similarity measures do. Then, in all the three graphs, pairs $\{v_1, v_4\}$, $\{u_1, u_4\}$, and $\{w_1, w_4\}$ have 2 common neighbors and this would be the value of the common neighborhood similarity measure for unsigned graphs. However, when we take into account the signs of link, the values of SCN for these pairs are 2, -2, and 0, respectively. The value of $\mathrm{SJ}(\{u,v\})$ is equal to the quotient of $\mathrm{SCN}(\{u,v\})$ by the cardinality of the sum of all signed neighborhoods of u and v.

We now move to discussing the global similarity measures for signed networks. Let $A = (a_{ij})_{v_i,v_j \in V}$ denote the adjacency matrix of a given signed graph, where $a_{ij} = 1$ if $\sigma(\{v_i, v_j) = +$, where $a_{ij} = 0$ if $\{v_i, v_j\} \notin E$, and where $a_{ij} = -1$ if $\sigma(\{v_i, v_j\}) = -$. Furthermore, let A^+ (A^-) denote the adjacency matrix of positive (negative) links in the graph, i.e. $a_{ij} = 1$ if $\sigma(\{v_i, v_j\}) = +$ ($\sigma(\{v_i, v_j\}) = -$) and 0 otherwise. According to the balance theory [20], a path between two nodes in the graph is balanced if it contains an even number of negative links and it is unbalanced otherwise. For a given signed graph G, we can define the matrices B_l and U_l, the entries of which are the numbers of balanced and unbalanced paths of length l, respectively, i.e., $B_l = (b^l_{ij})_{v_i,v_j \in V}$, and $U_l = (u^l_{ij})_{v_i,v_j \in V}$, where b^l_{ij} (u^l_{ij}) is the number of paths of length l between v_i and v_j where the number of negative links is even (odd). These matrices are inductively defined as:

$$B_1 := A^+; \quad U_1 := A^-;$$
$$B_{l+1} := B_l \cdot A^+ + U_l \cdot A^-; \quad U_{l+1} := U_l \cdot A^+ + B_l \cdot A^-.$$

Having defined B_l and U_l, we can now define a signed version of the Katz global similarity measure:

- **Signed Katz** (SK, see [15]):

$$\mathrm{SK}(v_i, v_j) = \sum_{l=1}^{\infty} \beta^l (b^l_{ij} - u^l_{ij}),$$

where $\beta \in (0,1)$ is a parameter that gives an exponential decay on the count of paths with their length increasing.

III. Attack Model

Let $G' = (V, E, \sigma(E))$ be a signed graph and let $H \subseteq (V \times V) \setminus E$ be a set of pairs of nodes not in E. Assume that the pairs of nodes $\{u, w\} \in H$ are actually linked, i.e., we consider the graph $G = (V, E \cup H, \sigma(E))$. Now, the links in H are the attacker's target. Specifically, the aim of the attacker is to make the sign in H as much difficult to predict as possible by the defender. To do so, the attacker is allowed to remove the signs of no more than k links in E.

We formalize our computational problems as follows:

Problem 1 (NEUTRALIZING SIGN PREDICTION (NSP))**.** *Given a signed graph $G = (V, E \cup H, \sigma(E))$, where $H \subseteq (V \times V) \setminus E$ is the attacker's target set of links, a subset $D \subseteq E$ of links the signs of which can be deleted, an integer $k \leq |D|$ denoting the budget of the attacker, i.e., the maximum number of signs which can be deleted, and a non-negative real number r, decide if there exists a subset $C \subseteq D$ such that $|C| \leq k$ and such that for all $\{u, v\} \in H$ it holds that*

$$|sim'(u,v)| \leq r,$$

for a similarity measure $sim : V \times V \to \mathbb{R}$, where $sim'(u,v)$ denotes the value of similarity $sim(u,v)$ in $G' = (V, E \cup H, \sigma(E \setminus C))$. An instance of this problem is a tuple (G, H, D, k, r).

Problem 2 (NEUTRALIZING SIGNED COMMON NEIGHBORHOOD (NSCN))**.** NSCN *is a variant of* NSP *in which we ask if there exists a subset $C \subseteq D$ such that $|C| \leq k$ and such that for all $\{u, v\} \in H$ it holds that for some $r \in \mathbb{R}$:*

$$|c_s(u,v)| - |c_d(u,v)| \leq r.$$

Secondly, we ask if it possible to completely eliminate the signed common neighborhood of all the pairs in the target set.

Problem 3 (ELIMINATING SIGNED COMMON NEIGHBORHOOD (ESCN))**.** ESCN *is a variant of* NSP *in which we ask if there exists a subset $C \subseteq D$ such that $|C| \leq k$ and such that for all $\{u, v\} \in H$ it holds that:*

$$c_s(u,v) \cup c_d(u,v) = \emptyset.$$

An instance of this problem is a tuple (G, H, D, k).

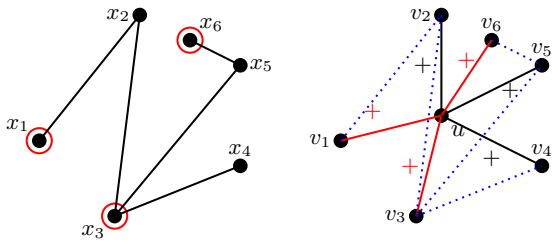

Fig. 2: The illustration of the reduction from VC to NSP in the proof of Theorem 1. The set of nodes $\{x_1, x_3, x_6\}$ forms a vertex cover of the graph $G = (X, E_X)$ on the left. Let us consider now the graph (V, E_V, σ) on the right and let us define the pairs in the target set H as corresponding to the links in E_X, i.e. $\{v_i, v_j\} \in H$ iff $\{x_i, x_j\} \in E_X$. Removing the signs of the links connecting v_1, v_3 and v_6 to u makes the SCN of all the pairs in H equal to 0.

Note that for the SCN or SJ similarity measures, a positive solution to ESCN is also a positive solution both to the NSCN and NSP problems with $r = 0$.

We also study an interesting version of NSP that takes into consideration the sum of moduli of the similarity scores of all the pairs in the target set:

Problem 4 (NEUTRALIZING TOTAL SIGN PREDICTION (NTSP))**.** *Given G, H, D, k, and r as in Problem 1, decide if there exists $C \subseteq D$ such that $|C| \leq k$ and such that:*

$$\sum_{\{u,v\} \in H} |sim'(u, v)| \leq r.$$

An instance of this problem is a tuple (G, H, D, k, r).

Finally, we also analyze a variant of the problem, where the goal of the attacker is to reverse the sign of the similarity score of the links from the target set.

Problem 5 (REVERSING SIGN PREDICTION (RSP))**.** *Given $G = (V, E \cup H, \sigma)$, $H \subseteq (V \times V) \setminus E$, $D \subseteq E$, k, and r as in Problem 1, decide if there exists $C \subseteq D$ such that $|C| \leq k$ and such that for all $\{u, v\} \in H$ it holds that*

$$sgn(sim'(u, v)) = -sgn(sim(u, v)).$$

An instance of this problem is a tuple (G, H, D, k).

IV. Complexity Analysis

In this section we first show that all the problems are NP-hard. We then analyze parameterized complexity.

A. NP-Hardness Results

We begin by demonstrating that NSP is NP-hard for local metrics, even for the most restricted case, and that NSCN and ESCN are NP-hard as well:

Theorem 1. *Let sim be a local similarity measure such that, for any $G = (V, E, \sigma)$ and any $u, v \in V$, it holds that:*

$$|c_s(u, v)| - |c_d(u, v)| = 0 \implies sim(u, v) = 0, \qquad (1)$$

Then, NSP *is* NP*-hard even if $r = 0$.*

Proof. We will prove a stronger result which says that even attacking similarity-based sign prediction by eliminating common signed neighborhood (i.e. making $c_s(u, v) \cup c_d(u, v) = \emptyset$ for all $u, v \in H$) is NP-hard. The theorem trivially follows. We reduce from the VERTEX COVER problem (VC) which is to decide, for a given graph $G = (X, E_X)$ and an integer $k \in \mathbb{N}$ whether there exists a vertex cover of G of size at most k, i.e. a subset $U \subseteq X$ with $|U| \leq k$ such that each link in E_X is incident to some node from U. Let $I = (G, k)$ be an instance of VC. Assume $|X| = n$ and fix any numbering of X, that is let $X = \{x_1, \ldots, x_n\}$. First, we construct a signed graph $(V, E_V \cup H, \sigma(E_V))$. The set of nodes V is made of the **original nodes**: for each node $x_i \in X$ construct its *copy* $v_i \in V$, and a **root node**: a fresh single element $u \in V$. The set of links E_V and their signs σ is defined as the set of **root links**: for each $i \leq n$ we construct a positive link between u and v_i. Let the target set H be the copy of the set E_X, and set $D := E_V$. Finally, let the budget of links, in the constructed instance of NSP, be equal to k.

We now show that the reduction is correct. Suppose (G, k) is a "yes" instance of VC. We first prove that then $(V, E_V \cup H, \sigma(E_V))$ is a "yes" instance of NSP. Let $U \subseteq X$ with $U = \{x_{i_1}, \ldots, x_{i_k}\}$ be a vertex cover of G of size k. In such the case, the attacker removes the positive signs of the links (v_{i_j}, u) for all $j \leq k$. We now claim that for all pairs in the target set, i.e. for each $\{v_i, v_j\} \in H$, the value of $sim(v_i, v_j)$ is equal to 0, since the similar common neighborhood of v_i and v_j is empty, i.e. $c_s(v_i, v_j) = \emptyset$. Indeed if $\{v_i, v_j\} \in H$, then by construction, $\{x_i, x_j\} \in E_X$. By the definition of vertex cover, at least one of the nodes x_i, x_j is in U. This means that either the sign of the link $\{v_i, u\}$ or $\{v_j, u\}$ is removed. Therefore, since before the sign removal, $c_s(v_i, v_j) = \{u\}$, now we have that $c_s(v_i, v_j) = \emptyset$, and thus $sim'(v_i, v_j) = 0$.

For the other direction, suppose that for each $v_i, v_j \in H$ we have $sim(v_i, v_j) = 0$ after deleting at most k signs of links in the graph (V, E_V, σ). We will now demonstrate that then there is a vertex cover of size at most k in the graph $G = (X, E_X)$. Indeed, since all the links had the positive sign, from the fact that $sim(v_i, v_j) = 0$ holds for all $v_i, v_j \in H$ it follows that $c_s(v_i, v_j) = \emptyset$ after the removal of some signs. But this implies that the links signs of which are removed are of the form $\{v, u\}$ for $v \in dom(H)$, since for each pair $v_i, v_j \in H$, their similar common neighborhood before the removal was $c_s(v_i, v_j) = \{u\}$. Let the links with their signs removed be $\{v_{i_1}, u\}$, $\{v_{i_2}, u\}$, …, $\{v_{i_k}, u\}$. We claim that $U = \{x_{i_1}, \ldots, x_{i_k}\}$ is a vertex cover of G. Indeed, let $\{x_l, x_m\} \in E_X$ be any link from G. Then, by assumption$\{v_l, v_m\} \in H$. Since after the sign removal $c_s(v_l, v_m) = \emptyset$, one of the links $\{v_l, u\}$ or $\{v_m, u\}$ had its sign removed, but by construction this means exactly that either x_l or x_m belongs to U. Since the choice of the link was arbitrary, this end the proof.

□

Algorithm 1 Tally heuristic

Input: a signed network G, the target set of pairs to be hidden H, the set links the signs of which can be removed $D \subseteq E$, the number of signs that can be removed k, and the positive contribution condition $\phi(u, w, v) : V \times V \times V \to \{0, 1\}$.

```
for i = 1, ..., k do
    for {v, w} ∈ E do
        g({v, w}) := 0
    for {u, w} ∈ H do
        for v ∈ c_s(u, v) ∪ c_d(u, v) do
            if φ(u, w, v) then
                g({v, u}) := g({v, u}) + 1
                g({v, w}) := g({v, w}) + 1
            else
                g({v, u}) := g({v, u}) − 1
                g({v, w}) := g({v, w}) − 1
    {v*, w*} := arg max_{{v,w}∈D} g({v, w})
    if g({v*, w*}) > 0 then
        σ := σ \ {σ({v*, w*})}
```

Corollary 1. ESCN *is* NP*-hard.* NSCN *is* NP*-hard, even when* $r = 0$ *and if* sim *is a local measure satisfying the condition 1. Then,* NTSP *is* NP*-hard, even when* $r = 0$*.*

A modification of the above reduction allows us to prove the following result (that applies e.g. to both SCN and SJ):

Theorem 2. *Let* sim *be a local similarity measure such that, for any signed graph* $G = (V, E, \sigma)$ *and any* $u, v \in V$*, it holds that:*

$$sgn(sim(u, v)) = sgn(|c_s(u, v)| - |c_d(u, v)|) \qquad (2)$$

Then, RSP *is* NP*-hard.*

As it turns out, computing the solution to NSP for signed global measures is hard as well—the following negative result holds for Signed Katz, by a reduction from the HAMILTONIAN CYCLE problem:

Theorem 3. *The NSP problem for the Signed Katz measure is* NP*-hard, even if the target set* H *contains only one link.*

The computational results have been so far negative. We can show now that when the target set of links H is induced by a set of *important* nodes U in the sense that H is simply the set of pairs between all nodes in U, i.e. $H = U \times U$, then some of the problems become tractable. Specifically:

Theorem 4. *Assume that in an instance of* ESCN *(or* NSCN*),* $D = E_G$*, and that the target set* H *consists solely of all the links between any two nodes in some set* $U \subseteq V$*, i.e.,* $H = \{\{u, v\} : u, v \in U\}$*. Then, there exists a polynomial-time algorithm for solving* ESCN *(or* NSCN*).*

V. The Tally Heuristic

Given the intractability of the NSP and RSP problems shown in the previous section, we now propose a heuristic solution that we call the Tally heuristic.

One of the arguments of the Tally heuristic is the positive contribution condition $\phi(u, w, v)$ that decides whether the removal of a sign from link $\{u, v\}$ or $\{w, v\}$ has a desired effect on the link $\{u, w\} \in H$. We use the following formula ϕ for the NSP problem:

$$\begin{aligned}\phi(u, w, v) = &(\text{SCN}(u, w) > 0 \wedge v \in c_s(u, v)) \\ &\vee (\text{SCN}(u, w) < 0 \wedge v \in c_d(u, v))\end{aligned}$$

and the following ϕ for the RSP problem:

$$\begin{aligned}\phi(u, w, v) = &(\sigma(u, w) = + \wedge v \in c_s(u, v)) \\ &\vee (\sigma(u, w) = - \wedge v \in c_d(u, v)) .\end{aligned}$$

Name	Nodes	Links	Negative signs percentage
Bitcoin Alpha	3,775	14,120	9.59%
Bitcoin OTC	5,875	21,489	14.94%
Wikipedia RFA	11,379	181,041	26.75%
Slashdot	79,116	467,731	25.38%
Epinions	119,130	704,267	17.10%
Wikipedia trust	137,740	715,334	12.25%

TABLE I: Basic characteristics of real-life datasets used.

VI. Experimental Evaluation

In this section, we evaluate the effectiveness of the heuristic presented in the previous section using simulations on both randomly-generated and real-life networks. Table I contains information about preprocessed datasets.

A. Datasets

In our simulations we consider the following real-life signed networks datasets: **Bitcoin Alpha** [21], **Bitcoin OTC** [21], **Wikipedia RFA** [22], **Slashdot** [23], **Epinions** [24], and **Wikipedia trust** [25].We also consider the three typical types of randomly-generated networks: **Barabási-Albert**, **Erdős-Rényi**, and **Watts-Strogatz** networks.

Unless stated otherwise, we consider randomly generated networks with 2000 nodes and the average degree of 30. We set the sign of 10% of the links (chosen uniformly at random) to minus, with all remaining links sign set to plus. We chose this value as it reflects the percentage of negative links in the real-life signed networks datasets.

B. Experimental Procedure

We now describe the experimental procedure of our simulations. Given an undirected signed network we first select the set of links H the sign of which we will attempt to hide. We consider two different ways of selecting the set H: either uniformly at random or as the links with the greatest values of the SCN score. In the simulations we select 50 links to be the elements of H.

We then use the Tally heuristic presented in the previous section in an attempt to hide the signs of the links in H. We compare the results of the Tally heuristic to a random baseline, that removes a sign from a random link incident with at least one link in H. We set the hiding budget to be the same as the

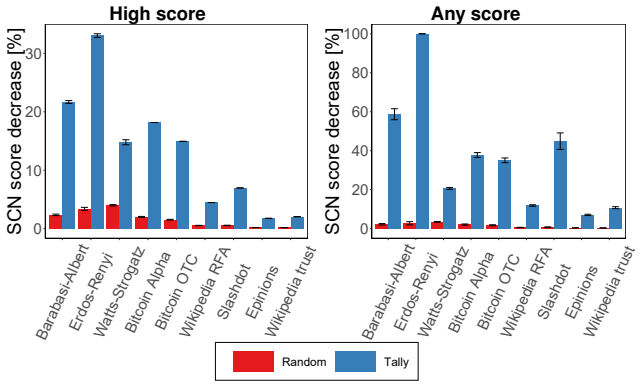

Fig. 3: Decrease of the sum of the absolute SCN scores of the links in H after the entire hiding process for the NSP problem instances (large values indicate better effectiveness of the heuristic). The left plot corresponds to H selected as the links with the greatest SCN scores, while the right plot to H selected uniformly at random. Error bars represent the 95%-confidence intervals.

size of H. During the hiding process we record the absolute value of the SCN score for the NSP problem instances, and the difference between the initial SCN score and the current SCN score for the RSP problem instances.

For the real-life network datasets we run the process for 1000 different sets H per selection criterion (i.e., either H selected uniformly at random or as the links with greatest SCN scores). For each random network generation model we generate 100 different networks, and for each of them we select 10 different sets H per selection criterion.

C. Simulation Results

When considering the NSP problem instances, Figure 3 compares the drop in the SCN scores at the end of the hiding process. As can be seen, the Tally heuristic is significantly more effective than the random approach. Moreover, hiding links that are chosen uniformly at random is on average much more effective than the attempt to hide links with very high SCN scores from the beginning.

Figure 4 allows us to investigate the effectiveness of the heuristics in random network with varying size and density. As can be seen, hiding signs in the NSP setting is generally more effective in sparse networks. As for the size of the network, trends observed in our simulations are less uniform. In Watts-Strogatz networks the hiding process is more effective in smaller networks, while in their Erdős-Rényi counterparts the same is true in larger structures. Interestingly, in Barabási-Albert networks hiding is more effective in larger networks when the links to be hidden are selected uniformly at random, but more effective is smaller networks when said links are those with greatest SCN scores.

Figures 5 presents the comparison of the effectiveness of the heuristics in the RSP setting. As in the NSP simulations, the Tally heuristic is significantly more effective than the

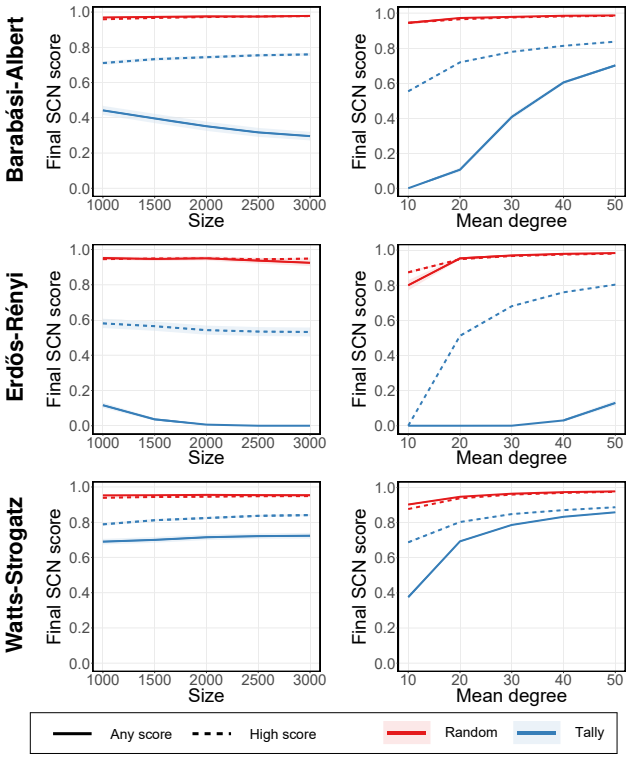

Fig. 4: The performance of the heuristics in random networks with varying size and average degree for NSP problem instances. The y-axis represents the relative sum of the absolute SCN scores of the links in H after the hiding process (small values indicate better effectiveness of the heuristic), while the x-axis represents either the changing size of the network (left column) or the average degree of the nodes (right column). Colored areas represent the 95%-confidence intervals.

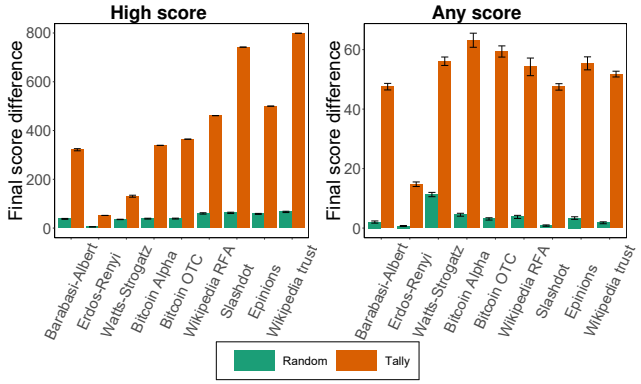

Fig. 5: The summed up difference in the SCN values of the links in H after the hiding process for the RSP problem instances (large values indicate better effectiveness of the heuristic). The left plot shows H selected as the links with the greatest SCN scores, while the right plot H selected uniformly at random. Error bars represent the 95%-confidence intervals.

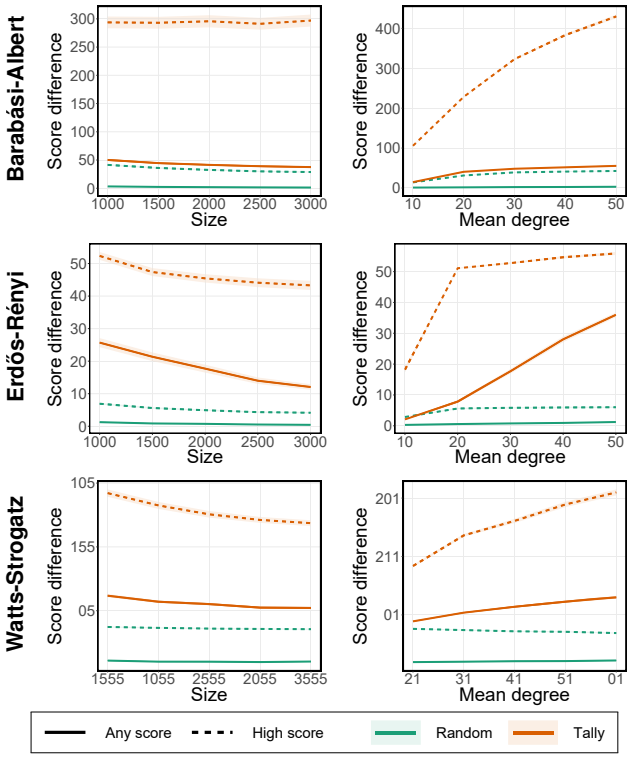

Fig. 6: The performance of the heuristics in random network with varying size and average degree for RSP problem instances. In each plot the y-axis represents as the summed up difference in SCN values of the links in H (large values indicate better effectiveness of the heuristic), while the x-axis represents either changing size of the network (left column) or average degree of the network nodes (right column). Colored areas represent the 95%-confidence intervals.

random alternative. Figure 6 presents the results of the RSP simulations in randomly generated networks with varying size density. The general trends remain consistent for both hiding random links and links with high SCN scores: on average the hiding process is more effective in small and dense networks.

Acknowledgements

Michał Tomasz Godziszewski and Tomasz Michalak were supported by the Polish National Science Centre grant 2016/23/B/ST6/03599. K. Zhou and Y. Zhu were supported by the PolyU Internal Fund (No. BE3U).

References

[1] P. Crescenzi, G. D'angelo, L. Severini, and Y. Velaj, "Greedily improving our own closeness centrality in a network," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, no. 1, p. 9, 2016.

[2] M. Waniek, T. P. Michalak, M. J. Wooldridge, and T. Rahwan, "Hiding individuals and communities in a social network," *Nature Human Behaviour*, vol. 2, no. 2, p. 139, 2018.

[3] T. Was, M. Waniek, T. Rahwan, and T. Michalak, "The manipulability of centrality measures-an axiomatic approach," in *19th International Conference on Autonomous Agents and MultiAgent Systems*. Auckland, New Zealand: AAMAS, 2020, pp. 1467–1475.

[4] M. Waniek, J. Woźnica, K. Zhou, Y. Vorobeychik, T. Rahwan, and T. Michalak, "Strategic evasion of centrality measures," in *Proceedings of the 20th International Conference on Autonomous Agents and Multi-Agent Systems*. UK: IFAAMAS, 2021.

[5] K. Zhou, T. P. Michalak, M. Waniek, T. Rahwan, and Y. Vorobeychik, "Attacking similarity-based link prediction in social networks," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 2019, pp. 305–313.

[6] K. Zhou, T. P. Michalak, and Y. Vorobeychik, "Adversarial robustness of similarity-based link prediction," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2019, pp. 926–935.

[7] P. Dey and S. Medya, "Manipulating node similarity measures in networks," in *AAMAS '20: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, 2020, p. 321–329.

[8] J. Chen, L. Chen, Y. Chen, M. Zhao, S. Yu, Q. Xuan, and X. Yang, "Ga-based q-attack on community detection," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 3, pp. 491–503, 2019.

[9] J. Tang, Y. Chang, C. Aggarwal, and H. Liu, "A survey of signed network mining in social media," *ACM Computing Surveys (CSUR)*, vol. 49, no. 3, pp. 1–37, 2016.

[10] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of the 19th international conference on World wide web*, 2010, pp. 641–650.

[11] P. Agrawal, V. K. Garg, and R. Narayanam, "Link label prediction in signed social networks," in *Twenty-third international joint conference on artificial intelligence*, 2013.

[12] K.-Y. Chiang, C.-J. Hsieh, N. Natarajan, I. S. Dhillon, and A. Tewari, "Prediction and clustering in signed networks: a local to global perspective," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1177–1213, 2014.

[13] B. Hu, H. Wang, X. Yu, W. Yuan, and T. He, "Sparse network embedding for community detection and sign prediction in signed social networks," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 1, pp. 175–186, 2019.

[14] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: statistical mechanics and its applications*, vol. 390, no. 6, pp. 1150–1170, 2011.

[15] T. Derr, C. Wang, S. Wang, and J. Tang, "Signed node relevance measurements," *arXiv preprint arXiv:1710.07236*, 2017.

[16] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[17] L. Linyuan and Z. Tao, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 6, pp. 1150–1170, 2011.

[18] X. Chen, J.-F. Guo, X. Pan, and C. Zhang, "Link prediction in signed networks based on connection degree," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 5, pp. 1747–1757, 2019.

[19] X. Chen, J. Guo, and X. e. a. Pan, "Link prediction in signed networks based on connection degree," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, p. 1747–1757, 2019.

[20] D. Cartwright and F. Harary, "Structural balance: a generalization of heider's theory." *Psychological review*, vol. 63, no. 5, p. 277, 1956.

[21] S. Kumar, F. Spezzano, V. Subrahmanian, and C. Faloutsos, "Edge weight prediction in weighted signed networks," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 2016, pp. 221–230.

[22] R. West, H. S. Paskov, J. Leskovec, and C. Potts, "Exploiting social network structure for person-to-person sentiment analysis," *Transactions of the Association for Computational Linguistics*, vol. 2, pp. 297–310, 2014.

[23] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The slashdot zoo: mining a social network with negative edges," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 741–750.

[24] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2010, pp. 1361–1370.

[25] S. Maniu, B. Cautis, and T. Abdessalem, "Building a signed network from interactions in wikipedia," in *Databases and Social Networks*. Association for Computing Machinery, 2011, pp. 19–24.