# Hiding in Temporal Networks

Marcin Waniek, Petter Holme, and Talal Rahwan

***Abstract*—Social network analysis tools can infer various attributes just by scrutinizing one's connections. Several researchers have studied the problem faced by an evader whose goal is to strategically rewire their social connections in order to mislead such tools, thereby concealing their private attributes. However, to date this literature has only considered static networks, while neglecting the more general case of temporal networks, where the structure evolves over time. Driven by this observation, we study how the evader can conceal their importance from an adversary armed with temporal centrality measures. We consider computational and structural aspects of this problem: Is it computationally feasible to calculate optimal ways of hiding? If it is, what network characteristics facilitate hiding? This topic has been studied in static networks, but in this work, we add realism to the problem by considering temporal networks of edges changing in time. We find that it is usually computationally infeasible to find the optimal way of hiding. On the other hand, by manipulating one's contacts, one could add a surprising amount of privacy. Compared to static networks, temporal networks offer more strategies for this type of manipulation and are thus, to some extent, easier to hide in.**

***Index Terms*—Computational complexity, social media, misinformation, network centrality, network theory.**

## I. Introduction

THE increasing sophistication and ubiquity of computer-based invigilation tools is a persistent threat to the privacy of the general public. The increasing number of privacy-related scandals, such as Cambridge Analytica using data of millions of Facebook users for political agendas [1], demonstrates just how vulnerable our private information is in the age of social media. Network data, from social media in particular, can be used to uncover sensitive information, such as sexual orientation [2], relationship status [3], or political views [4].

Due to this vulnerability of network data, a number of privacy-preservation solutions have been proposed, both in terms of legislature [5] and algorithmic solutions [6], [7]. Most of the literature puts the responsibility of protecting the system users' privacy in the hands of a centralized authority [8]–[10], but such an authority might be prone to error and negligence. A different body of literature proposes methods that can be applied by members of the social network to protect their own privacy, without having to rely on any central entities [11], [12].

Nevertheless, so far, the development of privacy-protection schemes for network data has focused on networks that are static, i.e., whose structure do not change over time. At the same time, network science researchers are starting to shift their attention to temporal networks—the more general case where the structure is allowed to change [13]. In many domains that involve dynamic changes, momentary contacts, and processes unfolding over time, the added complexity of a temporal-network approach can be justified by an added predictive and explanatory power [13]. Temporal networks already found application in such varied areas as communication [14], [15], microbiology [16], [17], and face-to-face interactions [18]. Nevertheless, so far no privacy-protection solutions have been proposed for temporal networks.

Motivated by the relevance of the growing privacy threats and the increasing popularity of temporal networks, in this work we examine how a member of a temporal network can avoid being detected by temporal centrality measures. We set out to investigate the issue from both theoretical and empirical standpoints. As for the theoretical analysis, we evaluate the computational complexity of the problem faced by an evader who wishes to rewire the network in order to obscure their central position in it. We consider both the decision version of the problem—if it is possible to find an optimal solution in polynomial time—as well as the optimization version—if it is possible to identify a solution that is guaranteed to be within a certain bound from optimum. As for the empirical analysis, we consider several real-life temporal network datasets and investigate how effectively the top nodes in the temporal centrality rankings can conceal their importance. We not only observe the results of the hiding process but also identify, using regression analysis, the properties of the nodes that allow them to conceal their importance effectively. Altogether, our study is the first analysis of strategic evasion of social network analysis tools in temporal networks.

### A. Our Contributions

We now briefly summarize the contributions of our work. First, we focus on the theoretical analysis of the problem of hiding from temporal centrality measures. For the decision version of the problem we prove that finding the optimal way of hiding is NP-complete for degree (Theorem 1), closeness (Theorem 2) and betweenness (Theorem 3) temporal centralities. For the optimization version of the problem we show a 2-

Manuscript received July 28, 2021; revised December 9, 2021; accepted January 27, 2022. Date of publication February 7, 2022; date of current version May 23, 2022. The work of Petter Holme was supported by JSPS KAKENHI under Grant JP 21H04595. Recommended for acceptance by Dr. Wei Chen. (*Corresponding author: Talal Rahwan.*)

Marcin Waniek and Talal Rahwan are with New York University Abu Dhabi, Abu Dhabi 129188, UAE (e-mail: mjwaniek@gmail.com; talal.rahwan@nyu.edu).

Petter Holme is with the Institute of Innovative Research, Tokyo Institute of Technology, Tokyo 152-8550, Japan (e-mail: pttrhlm@gmail.com).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSE.2022.3148752, provided by the authors.

Digital Object Identifier 10.1109/TNSE.2022.3148752

approximation algorithm for hiding from degree temporal centrality (Algorithm 1), while proving that the optimal solution cannot be approximated better than logarithmically given closeness (Theorem 5) and betweenness (Theorem 6) temporal centralities. As part of the empirical analysis we investigate a number of heuristic hiding algorithms on a variety of real-life temporal network datasets. Our results indicate the existence of a centrality-influence trade-off in temporal networks, where the removal of network contacts facilitates hiding, but at the cost of influence, while the addition of new contacts increases one's influence, but makes them more exposed. Finally, using regression analysis we find that nodes whose contacts are spread over time have a greater chance of successfully hiding their importance from temporal centrality measures.

## II. Related Work

In recent years there has been a growing interest in temporal networks, where the network structure is not static, but instead changes with the passage of time [13]. Temporal networks found particularly relevant applications in epidemics, where they have been used to predict the infection's reproduction number [19] and to construct static graphs based on temporal contact data [20]. In the context of our work, an essential class of tools for the analysis of temporal networks are centrality measures [21]. The approach to their design varies greatly, ranging from the analysis of network flows [22] and shortest temporal paths [23], to the applications of eigenvector-like techniques [24]–[26].

The topic of avoiding detection by social network analysis tools recently received some attention in the literature devoted to static networks. The greatest amount of attention is focused on hiding from centrality measures, either by lowering the node's centrality, either in absolute terms [27] or in relative terms [28], [29]. Other works provide an axiomatic characterization of centrality measures that are resilient to being fooled [30], analyze the possible strategies of an adversary who is aware of the existence of nodes that want to hide themselves [31], or consider the problem of hiding from centrality measures in multilayer networks [32]. Other hiding problems considered in the literature include preventing the identification of closely-cooperating groups of nodes by community detection algorithms [27], avoiding the detection of private relationships by link prediction algorithms [33], [34], manipulating the node similarity measures [35], and investigating the possibility of concealing the source of network diffusion from source detection algorithms [36]. All of these works consider only static networks.

## III. Preliminaries

### A. Temporal Networks

Throughout the article, we will let $\langle T \rangle$ denote a time interval of $T$ discrete time steps, i.e., $\langle T \rangle = \{0, \ldots, T-1\}$. We will sometimes refer to a particular $t \in \langle T \rangle$ as the *moment* $t$. Let us denote by $G = (V, K, T) \in \mathbb{G}$ a temporal network, where $V$ is the set of $n$ nodes, $K \subseteq V \times V \times \langle T \rangle$ is the set of contacts, and $T$ is the duration of the time interval during which the contacts in $K$ take place. We denote a *contact* (sometimes also called a *temporal edge*) between nodes $v$ and $w$ at time $t$ by $(v, w, t)$. In this work we only consider *undirected* temporal networks, i.e., we do not discern between contacts $(v, w, t)$ and $(w, v, t)$. Moreover, we assume that networks do not contain self-contacts, i.e., $\forall_{v \in V} \forall_{t \in \langle T \rangle} (v, v, t) \notin K$. We denote all contacts of a given node $v$ by $K_G(v)$. Finally, for $K' \subseteq V \times V \times \langle T \rangle$ we denote by $G \cup K'$ the effect of adding the set of contacts $K'$ to $G$, i.e., $G \cup K' = (V, K \cup K,' T)$.

A *time-respecting path* (or a *temporal path*) in a temporal network $G = (V, K, T)$ is an ordered sequence of distinct contacts from $K$, $\langle c_1, \ldots, c_k \rangle$, in which for every two consecutive contacts $(v, w, t)$ and $(v,' w,' t')$ we have that $w = v'$ and $t' > t$. In other words, contacts on a time-respecting path occur chronologically, allowing for the diffusion of information along the path. For example, a message or a piece of news can be transmitted via a path only if it is time-respecting, as any given node on the path has to receive the message before propagating it further. The *duration* of the path is the time difference between the first and the last contact in the path, i.e., the duration of a path $\langle (v, w, t), \ldots, (v,' w,' t') \rangle$ is $t' - t$. Let $\Pi_G(v, w)$ denote the set of temporal paths from $v$ to $w$ with the minimal duration. We say that we can *reach* node $w$ from node $v$ at time $t$ if there exists a time-respecting path from $v$ to $w$ that starts at time greater than or equal to $t$. The *latency* between $v$ and $w$ at time $t$ is the shortest time it takes to reach $w$ from $v$ starting at time $t$ along time-respecting paths, we denote it by $\lambda_G(v, w, t)$. Formally:

$$\lambda_G(v, w, t) = \min_{t \leq t' < t'' : \widehat{\Pi}_{t'}^{t''}(v,w) \neq \emptyset} t'' - t.$$

where $\widehat{\Pi}_{t'}^{t''}(v, w)$ is the set of time-respecting paths from $v$ to $w$ where the first contact takes place at time $t'$ and the last contact takes place at time $t''$. If no such time-respecting path exists, i.e., $\forall_{t \leq t' < t''} \widehat{\Pi}_{t'}^{t''}(v, w) = \emptyset$, then $\lambda_G(v, w, t) = \infty$. Intuitively, latency specifies how quickly one node can be reached from another, e.g., if node $v$ at time $t$ wants to send a message that will reach node $w$ as quickly as possible, it will take the message $\lambda_G(v, w, t)$ time steps to reach $w$. We denote *average latency* between $v$ and $w$ by $\lambda^*(v, w)$. Notice that average latency is the area under the latency plot, divided by the length of the time interval $T - 1$. Following Pan and Saramäki [23] we assume a pair-specific temporal boundary condition, where for every pair of nodes, the first time-respecting path between them is repeated after the end of the observed time interval when computing the average latency for that pair. Hence, if there exists at least one time-respecting path between the pair of nodes, then the average latency between them is finite. An example of a latency plot used to compute the average latency is presented in Fig. 1.

To make the notation more readable, we will often omit the temporal network itself from the notation whenever it is clear from the context, e.g., by writing $\lambda(v, w, t)$ instead of $\lambda_G(v, w, t)$. This applies not only to the notation presented thus far, but rather to all notation in this article.

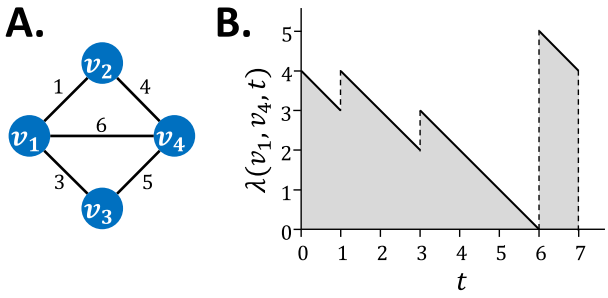


Fig. 1. **An example of temporal network and a latency plot. A.** A temporal network with times of each contact denoted as number next to edges. **B.** A plot of latency between $v_1$ and $v_4$ for under assumption that the duration of time interval is $T = 7$. The gray area under plot is proportional to average latency between $v_1$ and $v_4$. Notice that the latency in time interval (6,7] is finite due to the assumption about a pair-specific temporal boundary condition [23].

### B. Temporal Centrality Measures

Centrality measures quantify the importance of a given node in a network. The concept has also been extended to temporal networks [21]. In this work we consider the following temporal centrality measures:

1) Degree temporal centrality [21]—importance of a node $v$ corresponds to its number of contacts.

$$c_D((V,K,T),v) = \frac{\sum_{t\in\langle T\rangle} |\{w \in V : (v,w,t) \in K\}|}{(n-1)T}.$$

2) Closeness temporal centrality [23]—importance of a node $v$ corresponds to its average latency to other nodes.

$$c_C((V,K,T),v) = \frac{1}{n-1} \sum_{w\in V: v\neq w} \frac{1}{\lambda^*(v,w)}.$$

3) Betweenness temporal centrality [22]—importance of a node $v$ corresponds to the percentage of shortest temporal paths between pairs of other nodes passing through $v$.

$$c_B((V,K,T),v) = \frac{1}{(n-1)(n-2)T} \sum_{\substack{u,w\in V\setminus\{v\}:\\ u\neq w}} \sum_{t=0}^{T-1} \frac{|\widetilde{\Pi}_t^v(u,w)|}{|\Pi(u,w)|}$$

where $\widetilde{\Pi}_t^v(u,w)$ is the set of shortest temporal paths from $u$ to $w$ such that $v$ belongs to the path and either contacts between $v$ and its predecessor and successor take place at moment $t$, or contact between $v$ and its predecessor takes place at or before moment $t$, while the contact between $v$ and its successor takes place after moment $t$:

$$\widetilde{\Pi}_t^v(u,w) = \{\pi \in \Pi(u,w) : (v',v,t') \in \pi \wedge (v,v'',t'') \in \pi \wedge (t' = t = t'' \vee t' \leq t < t'')\}.$$

4) Eigenvector temporal centrality [25]—importance of a node $v$ corresponds to the importance of its neighbors. The temporal version of the eigenvector centrality was defined in a number of ways [24], [37]–[39]. In our work we use algorithm by Lv *et al.* [25], as it allows to efficiently process even relatively large networks.

## IV. Theoretical Analysis

We now present the formal definitions of the computational problems faced by the evader. In both versions of the problem, the evader has to decide which of the contacts from $\widehat{A}$ to add, and which of the contacts from $\widehat{R}$ to remove in order to hide from a temporal centrality measure $c$. In the decision version of the problem (Definition 1), the evader adds or removes at most $b$ contacts from $\widehat{A}$ and $\widehat{R}$ in order to avoid occupying one of the top $\omega$ positions in the ranking of $c$. In the optimization version of the problem (Definition 2), the evader adds or removes as few contacts from $\widehat{A}$ and $\widehat{R}$ as possible in order to avoid occupying one of the top $\omega$ positions in the ranking of $c$.

*Definition 1 (Temporal Hiding):* An instance of the Temporal Hiding problem is defined by a tuple $(G, v^\dagger, \widehat{A}, \widehat{R}, b, c, \omega)$, where $G = (V, K, T)$ is a temporal network, $v^\dagger$ is the evader, $\widehat{A} \subseteq V \times V \times \langle T\rangle$ is the set of contacts allowed to be added, $\widehat{R} \subseteq K$ is the set of contacts allowed to be removed, $b \in \mathbb{N}$ is a budget specifying the maximum number of contacts that can be added or removed, $c : \mathbb{G} \times V \to \mathbb{R}$ is a temporal centrality measure, and $\omega \in \mathbb{N}$ is a chosen safety threshold. The goal is then to identify two sets, $A^* \subseteq \widehat{A}$ and $R^* \subseteq \widehat{R}$, such that $|A^*| + |R^*| \leq b$ and:

$$\left|\left\{w \in V : c(G^*, w) > c(G^*, v^\dagger)\right\}\right| \geq \omega$$

where $G^* = (V, (K \cup A^*) \setminus R^*, T)$.

We also present an optimization version of the problem, where the goal of the evader is to satisfy a certain safety threshold using as few network modifications as possible.

*Definition 2 (Minimum Temporal Hiding):* An instance of the Minimum Temporal Hiding problem is defined by a tuple $(G, v^\dagger, \widehat{A}, \widehat{R}, c, \omega)$, where $G = (V, K, T)$ is a temporal network, $v^\dagger$ is the evader, $\widehat{A} \subseteq V \times V \times \langle T\rangle$ is the set of contacts allowed to be added, $\widehat{R} \subseteq K$ is the set of contacts allowed to be removed, $c : \mathbb{G} \times V \to \mathbb{R}$ is a temporal centrality measure, and $\omega \in \mathbb{N}$ is a chosen safety threshold. The goal is then to identify two sets, $A^* \subseteq \widehat{A}$ and $R^* \subseteq \widehat{R}$, such that the sum of their sizes $|A^*| + |R^*|$ is as small as possible and and:

$$\left|\left\{w \in V : c(G^*, w) > c(G^*, v^\dagger)\right\}\right| \geq \omega$$

where $G^* = (V, (K \cup A^*) \setminus R^*, T)$.

As can be seen, the difference from the decision version of the problem, i.e., from Definition 1, is that now we are not looking for $A^*$ and $R^*$ satisfying a certain budget $b$, but rather $A^*$ and $R^*$ the sum of which is as small as possible. In our analysis we will attempt to identify approximation algorithms that get as close to this optimal, smallest size as possible. Notice that this kind of analysis would not be possible with the decision version of the problem, where there a solution for a given $b$ is either identified or not.

TABLE I
SUMMARY OF OUR COMPUTATIONAL COMPLEXITY RESULTS

| Centrality | Temporal Hiding | Minimum Temporal Hiding |
|---|---|---|
| Degree | NP-complete (Theorem 1) | We show a 2-approximation algorithm (Theorem 4) |
| Closeness | NP-complete (Theorem 2) | Inapproximable within $(1-\epsilon)\ln\lvert\widehat{A}\rvert$ for any $\epsilon > 0$ (Theorem 5) |
| Betweenness | NP-complete (Theorem 3) | Inapproximable within $(1-\epsilon)\ln\lvert\widehat{A}\rvert$ for any $\epsilon > 0$ (Theorem 6) |

As can be seen, the difference between the decision version of the problem (Definition 1) and the optimization version (Definition 2) is that the former involves looking for $A^*$ and $R^*$ that satisfy a certain budget $b$, whereas the latter involves looking for $A^*$ and $R^*$ the sum of which is as small as possible. In our analysis of the optimization version, we will attempt to identify approximation algorithms that get as close as possible to the optimal (i.e., smallest) size. Notice that this kind of analysis would not be possible with the decision version, because a solution for any given b may or may not exist.

Table I presents the summary of our computational complexity results. Due to space constraints we present the proofs of some of our theorem in the Supplementary Materials. It should be noted that we investigate the complexity of both problems given the degree, closeness, and betweenness temporal centralities, since their simple closed-form formulas make them very amenable to theoretical analysis. On the other hand, the eigenvector temporal centrality lacks such a simple formula, which explains why we only consider it in our empirical analysis in Section V.

*Theorem 1:* Temporal Hiding problem is NP-complete given the degree temporal centrality.

*Theorem 2:* Temporal Hiding problem is NP-complete given the closeness temporal centrality.

*Theorem 3:* Temporal Hiding problem is NP-complete given the betweenness temporal centrality.

*Proof:* The problem is trivially in NP, since after adding and removing a given set of contacts we can computed the ranking of all nodes according to the betweenness temporal centrality in polynomial time.

We first give a high-level overview of the proof of NP-hardness of the problem. We will show that any algorithm that can solve an instance of the Temporal Hiding problem for the temporal betweenness centrality in polynomial time, can be used to solve a well-known NP-complete problem of 3-Set Cover. To this end, we take an arbitrary instance of the 3-Set Cover problem, construct a corresponding instance of the Temporal Hiding problem, and prove that based on a solution to the constructed instance of Temporal Hiding, we can obtain a solution to the arbitrary instance of 3-Set Cover. Since 3-Set Cover is NP-complete, this implies that any problem in NP can be reduced to 3-Set Cover, which then in turn can be reduced to Temporal Hiding. Thus, the Temporal Hiding problem is as hard as any problem in NP.

We will now prove that the problem is NP-hard. To this end, we will show a reduction from the NP-complete *3-Set Cover* problem. The decision version of this problem is defined by a universe, $U = \{u_1, \ldots, u_{|U|}\}$, a collection of sets $S = \{S_1, \ldots, S_{|S|}\}$ such that $\forall_i S_i \subset U \wedge |S_i| = 3$, and an integer $k \in \mathbb{N}$ where the goal is to determine whether there exist $k$ elements of $S$ the union of which equals $U$.

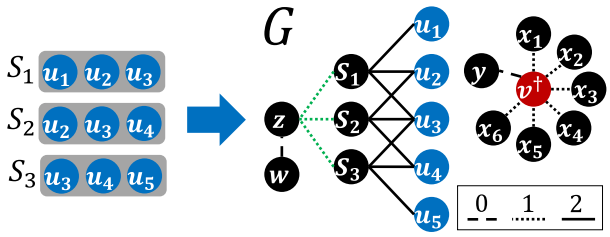


Fig. 2. Construction used in the proof of Theorem 3. On the left there is an instance of the 3-Set Cover problem, while on the right there is a network $G$ constructed as part of the Temporal Hiding problem instance. The blue nodes in $G$ correspond to the elements of the universe $U$, while the evader is marked red. Dashed lines correspond to contacts at moment 0, dotted lines to contacts at moment 1, and solid lines to contact at moment 2. Green lines depict the contacts allowed to be added.

Let $(U, S, k)$ be a given instance of the 3-Set Cover problem. Let us assume that $|U| \geq 6$, note that all instances where $|U| < 6$ can be solved in polynomial time. We will now construct an instance of the Temporal Hiding problem. First, let us construct a temporal network $G = (V, K, T)$ where:

1) $V = \{v^\dagger, w, y, z\} \cup U \cup S \cup \bigcup_{i=1}^{|U|+k-1}\{x_i\}$,
2) $K = \{(z, w, 0), (v^\dagger, y, 0)\} \cup \bigcup_{x_i \in V}\{(v^\dagger, x_i, 1)\} \cup \bigcup_{u_i \in S_j}\{(u_i, S_j, 2)\}$,
3) $T = 3$.

An example of the construction of the network $G$ is presented in Fig. 2.

Now, consider the instance $(G, v^\dagger, \widehat{A}, \widehat{R}, b, c, \omega)$ of the Temporal Hiding problem, where:

1) $G$ is the temporal network we just constructed,
2) $v^\dagger$ is the evader,
3) $\widehat{A} = \{(z, S_i, 1) : S_i \in V\}$,
4) $\widehat{R} = \emptyset$, i.e., none of the edges can be removed,
5) $b = k$,
6) $c$ is the temporal betweenness centrality,
7) $\omega = 1$ is the safety threshold.

Since $\widehat{R} = \emptyset$, for any solution to the constructed instance of the Temporal Hiding problem we must have $R^* = \emptyset$. Hence, we will omit mentions of $R^*$ in the remainder of the proof, and we will assume that a solution consists just of $A^*$.

First, let us analyze the temporal betweenness centrality values of the nodes in $G$ after the addition of any $A \subseteq \widehat{A}$. Let $q_A$ denote the number of nodes $u_j \in U$ such that $\lambda^*(z, u_j) < \infty$, i.e., the number of nodes $u_j \in U$ such that $z$ is connected (via the contacts from $A$) with at least one node $S_i$ connected with $u_j$ (notice that there are no other possibilities for $z$ to have finite average latency to a node in $U$). The temporal betweenness centrality values can be computed as follows:

1) $c_B(G \cup A, v^\dagger) = \frac{|U|+k-1}{(n-1)(n-2)T}$, as it belongs to the shortest temporal paths from $y$ to all $|U| + k - 1$ nodes $x_i$,
2) $c_B(G \cup A, z) = \frac{|A|+q_A}{(n-1)(n-2)T}$, as it belongs to the shortest temporal paths from $w$ to all $|A|$ nodes $S_i$ it is connected with and to all $q_A$ nodes $u_i$ that can be reached from it,
3) $c_B(G \cup A, S_i) \leq \frac{6}{(n-1)(n-2)T} \leq c_B(G \cup A, v^\dagger)$, for any $i$, as $S_i$ can belong to the shortest temporal paths from nodes in $\{z, w\}$ to the three nodes in $U$ that it is connected with,
4) $c_B(G \cup A, w) = c_B(G \cup A, y) = c_B(G \cup A, u_i) = c_B(G \cup A, x_i) = 0$, for any $i$, as none of these nodes belong to the shortest temporal paths between any pairs of other nodes.

Since the safety threshold is $\omega = 1$, only one node has to have greater temporal betweenness centrality than $v^\dagger$ after he addition of a given $A \subseteq \widehat{A}$ in order for said $A$ to be solution to the constructed instance of the Temporal Hiding problem. However, notice that the only node that can have greater temporal betweenness centrality than $v^\dagger$ (and therefore higher position in the ranking) is $z$. In other words, a given $A \subseteq \widehat{A}$ is a solution to the constructed instance of the Temporal Hiding problem if and only if we have $c_B(G \cup A, z) > c_B(G \cup A, v^\dagger)$.

We will now show that the constructed instance of the Temporal Hiding problem has a solution if and only if the given instance of the 3-Set Cover problem has a solution.

Assume that there exists a solution to the given instance of the 3-Set Cover problem, i.e., a subset $S^* \subseteq S$ of size $k$ the union of which is the universe $U$. We will show that $A^* = \{z\} \times S^* \times \{1\}$ (i.e., connecting $z$ with nodes corresponding to all sets in $S^*$) is a solution to the constructed instance of the Temporal Hiding problem. First, notice that after the addition of $A^*$ there exists a time-respecting path from $z$ to every node $u_j \in U$, leading through the node corresponding to an element $S_i \in S^*$ containing $u_j$, with which $z$ is now connected. Hence, we have that $q_{A^*} = |U|$ and $|A^*| = k$, which gives us:

$$c_B(G \cup A^*, z) = \frac{|U| + k}{(n-1)(n-2)T} > \frac{|U| + k - 1}{(n-1)(n-2)T} = c_B(G \cup A^*, v^\dagger).$$

Therefore, after the addition of $A^*$ node $z$ has greater temporal betweenness centrality than the evader. We showed that if there exists a solution to the given instance of the 3-Set Cover problem, then there also exists a solution to the constructed instance of the Temporal Hiding problem.

Assume that there exists a solution $A^*$ to the constructed instance of the Temporal Hiding problem. We will show that $S^* = \{S_i \in S : (z, S_i, 1) \in A^*\}$ covers the universe $U$. Let us compute the difference between $c_B(G \cup A^*, v^\dagger)$ and $c_B(G \cup A^*, z)$:

$$c_B(G \cup A^*, z) - c_B(G \cup A^*, v^\dagger) = \frac{q_{A^*} + |A^*| - |U| - k + 1}{(n-1)(n-2)T}.$$

Since $A^*$ is a solution, $z$ must have greater temporal betweenness centrality than $v^\dagger$, implying $c_B(G \cup A^*, z) - c_B(G \cup A^*, v^\dagger) > 0$, which gives us:

$$q_{A^*} + |A^*| + 1 > |U| + k.$$

Notice however that $|A^*| \leq k$ (since the evader's budget is $b = k$) and that $q_{A^*} \leq |U|$ (since $q_{A^*}$ is the number of nodes in $U$ at finite average latency from $z$). Therefore, we must have $|A^*| = k$ and $q_{A^*} = |U|$, which means that after the addition of $A^*$ for every $u_j \in U$ we have a node $S_i \in S$ connected to both $z$ and $u_j$ (there is no other way for $z$ to have a finite average latency to $u_j$). Consequently, every element of the universe $u_j \in U$ is covered by at least one set $S_i \in S^*$, as $z$ is connected with a node $S_i$ only if it belongs to $S^*$, and $S_i$ is connected with $u_j$ only if it contains $u_j$ in the 3-Set Cover problem instance. We showed that if there exists a solution to the constructed instance of the Temporal Hiding problem, then there also exists a solution to the given instance of the 3-Set Cover problem. This concludes the proof. ■

*Theorem 4:* Algorithm 1 is a 2-approximation for the Minimum Temporal Hiding problem given the degree temporal centrality. The bound is tight, i.e., there exists a network for which the approximation ratio of Algorithm 1 is exactly 2.

*Proof:* We first give a high-level overview of Algorithm 1. The algorithm begins its operation by using the dynamic programming technique to compute the minimal possible size of the solution under an assumptions that any two nodes appearing as part of the contacts in $\widehat{A}$ can be connected. The solution of this size might be impossible to obtain (either because it does not exist, or because its computation would involve solving an NP-complete problem), but it serves as the lower bound for a given instance of the problem. The algorithm then proceeds to compute a solution that is at most twice as costly as the computed theoretical optimum. We conclude by showing an instance of the problem where the identified solution is exactly twice as big as the optimal solution, thus proving the tightness of the algorithm bound.

We now move to the description of the algorithm. First, notice that we can satisfy the safety threshold by lowering the degree of $v^\dagger$ or by increasing the degrees of other nodes (as a given node $v$ contributes to the safety thresholds if and only if the difference between the degree of $v^\dagger$ and the degree of $v$ is negative). Thus, removing contacts that are not incident with $v^\dagger$ and adding edges that are incident with $v^\dagger$ is always suboptimal, as in case of both these changes the difference between the degree of the evader and all other nodes either increases or stays the same. Hence, no solution of the optimal size will include these types of modification and we exclude them in lines 2-3.

Let the cost of a given solution be expressed by a number of stubs (half-contacts), i.e., twice the number of contacts. In lines 4-27 the algorithm computes the minimal cost of a solution that satisfies the safety threshold, under assumption that we are allowed to connect any two stubs from contacts appearing in $\widehat{A}$ (thus we are guaranteed that the actual optimum is greater or equal). It does so using the dynamic programming technique. The loop in line 4 iterates over the number of contacts removed from the network. The algorithm fills the arrays $X$ and $Y$ in order to identify $C_z$, the optimal cost of solution that satisfies the safety thresholds while removing $z$ contacts from the

**Algorithm 1:** Finding a 2-approximate solution for the Minimum Temporal Hiding problem given the degree temporal centrality.

**Input:** Temporal networks $G = (V, K, T)$, evader $v^\dagger$, set of edges allowed to be added $\widehat{A}$, set of edges allowed to be removed $\widehat{R}$, safety threshold $\omega$.
**Output:** Solution $(A^*, R^*)$ to the instance $(G, v^\dagger, \widehat{A}, \widehat{R}, c_C, \omega)$ of the Temporal Hiding problem or $\perp$ if there is no solution.

1: Let $\langle v\rangle_{i=1}^{n-1}$ be a sequence of all nodes in $V$ other than $v^\dagger$
2: $\widehat{A} \leftarrow \widehat{A} \setminus (\{v^\dagger\}\times V\times\langle T\rangle)$
3: $\widehat{R} \leftarrow \widehat{R} \cap (\{v^\dagger\}\times V\times\langle T\rangle)$
4: **for** $z \leftarrow 0,\ldots,|\widehat{R}|$ **do**
5: $\quad X_i^z[\theta,q] \leftarrow \infty$ for every $i,\theta,q$
6: $\quad X_0^z[0,0] \leftarrow 0$
7: $\quad$ **for** $i \leftarrow 1,\ldots,n-1$ **do**
8: $\quad\quad \phi_i \leftarrow |\widehat{A} \cap (\{v_i\}\times V\times\langle T\rangle)|$
9: $\quad\quad$ **for** $r \leftarrow 0,\ldots,|\widehat{R} \cap (\{v^\dagger\}\times\{v_i\}\times\langle T\rangle)|$ **do**
10: $\quad\quad\quad a \leftarrow \max(0, |K_G(v^\dagger)| - z + 1 - |K_G(v_i)| + r)$
11: $\quad\quad\quad$ **for** $\theta \leftarrow 0,\ldots,\min(i-1,\omega)$ **do**
12: $\quad\quad\quad\quad$ **for** $q \leftarrow 0,\ldots,z$ **do**
13: $\quad\quad\quad\quad\quad$ **if** $a \le \phi_i \wedge X_{i-1}^z[\theta,q] + a < X_i^z[\theta+1, q+r]$ **then**
14: $\quad\quad\quad\quad\quad\quad X_i^z[\theta+1, q+r] \leftarrow X_{i-1}^z[\theta,q] + a$
15: $\quad\quad\quad\quad\quad\quad Y_i^z[\theta+1, q+r] \leftarrow (a, r)$
16: $\quad\quad\quad\quad\quad$ **if** $a > 0 \wedge X_{i-1}^z[\theta,q] < X_i^z[\theta, q+r]$ **then**
17: $\quad\quad\quad\quad\quad\quad X_i^z[\theta, q+r] \leftarrow X_{i-1}^z[\theta,q]$
18: $\quad\quad\quad\quad\quad\quad Y_i^z[\theta, q+r] \leftarrow (0, r)$
19: $\quad C_z \leftarrow X_{n-1}^z[\omega, z] + 2z$
20: $z^* \leftarrow \operatorname{argmin}_z C_z$
21: **if** $C_{z^*} = \infty$ **then**
22: $\quad$ **return** $\perp$
23: $(A^*, R^*) \leftarrow (\emptyset, \emptyset)$
24: $x^* \leftarrow C_{z^*} - 2z^*$
25: $\theta^* \leftarrow \omega$
26: $q^* \leftarrow z^*$
27: **for** $i \leftarrow n-1,\ldots,1$ **do**
28: $\quad (a, r) \leftarrow Y_i^{z^*}[\theta^*, q^*]$
29: $\quad A^* \leftarrow A^* \cup \text{select}\left(a,\ \widehat{A} \cap (\{v_i\}\times V\times\langle T\rangle)\right)$
30: $\quad R^* \leftarrow R^* \cup \text{select}\left(r,\ \widehat{R} \cap (\{v^\dagger\}\times v_i\times\langle T\rangle)\right)$
31: $\quad x^* \leftarrow x^* - a$
32: $\quad q^* \leftarrow q^* - r$
33: $\quad$ **if** $|K_G(v_i)| + a - r > |K_G(v^\dagger)| - z^*$ **then**
34: $\quad\quad \theta^* \leftarrow \theta^* - 1$
35: **return** $(A^*, R^*)$

network. The value of the element $X_i^z[\theta, q]$ is the minimal number of stubs that have to added to nodes $v_1,\ldots,v_i$ so that $\theta$ of them have greater degrees than the evader, assuming that we removed $q$ stubs incident with them and that the degree of the evader is $|K_G(v^\dagger)| - z$. The element $Y_i^z[\theta, q]$ stores the number of stubs that are added to and removed from node $v_i$ to achieve the solution with the cost $X_i^z[\theta, q]$. The loop in line 7 iterates over all nodes in the network other than the evader, while the loop in line 9 iterates over the number of stubs removed from the node $v_i$. The value $a$ computed in line 10 is the number of stubs that have to added to node $v_i$ in order for it to contribute towards the safety threshold, i.e., in order for it to have greater

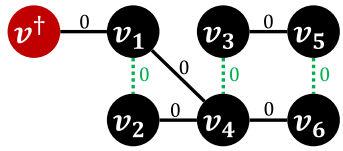


Fig. 3. Example of a network for which the approximation ratio of Algorithm 1 is exactly 2. The red node corresponds to the evader. Numbers next to edges are the moments in which the contacts occur. Green dotted lines depict the contacts allowed to be added.

degree than the evader. The loops in lines 11 and 12 iterate over all values of $\theta$ and $q$ valid for arrays $X_{i-1}^z$ and $Y_{i-1}^z$, and based on them fill arrays $X_i^z$ and $Y_i^z$. If the addition of the required number of $a$ stubs is possible (test in line 13) then we record a solution where the node $v_i$ contributes towards the safety threshold (value $\theta + 1$ in lines 14-15). In lines 18-19) we record a solution where the node $v_i$ does not contribute towards the safety threshold, but is only used to decrease the degree of the evader. We do it only if counting towards the threshold actually required adding any stubs (test in line 17). Finally, in line 25 we compute the cost of the optimal solution that removes $z$ contacts from the network, while satisfying the safety threshold $\omega$.

We showed that $C_{z^*}$ is the minimal cost of a solution that satisfies the safety threshold, under assumption that we are allowed to connect any two stubs from contacts appearing in $\widehat{A}$. If no solution was found (which is tested in line 28), then the algorithm returns $\perp$ in line 29. Otherwise, the algorithm constructs the solution in lines 31-43. Since we removed all contact including $v^\dagger$ from $\widehat{A}$, any solution that removes all contacts from the optimal solution and adds all the stubs from the optimal solution satisfies the safety threshold. What is more, such a solution contains at most twice as many contacts as the optimum (as it is possible that each stub in $\widehat{A}$ from the optimal solution gets connected to a stub not appearing in the optimal solution). Hence, the algorithm is a 2-approximation. The complexity of the algorithm is $\mathcal{O}(|\widehat{R}|^3 n\omega)$.

Notice that the approximation ratio of the algorithm cannot be easily improved, as the main inefficiency is caused by the way in which contacts from $\widehat{A}$ are added to the network. Adding these contacts in an optimal way would involve solving the equivalent of an NP-complete problem, namely Finding $k$-Clique. More formally, assuming that we would have to add $\frac{k(k-1)}{2}$ edges to increase a degree of $k$ nodes by $k-1$ each, one can create an instance of this problem following steps that are very similar to those used in the proof of Theorem 1).

Finally, we prove the tightness of the bound of Algorithm 1. To this end, Fig. 3 presents an example of a network for which the approximation ratio of the algorithm is exactly 2. Given a safety threshold $\omega = 4$, Algorithm 1 will add to the network contacts $(v_1, v_2, 0)$ and $(v_3, v_4, 0)$ (because of the order of nodes in $\langle v_i\rangle_{i=1}^{n-1}$), whereas the optimal solution is to add the contact $(v_5, v_6, 0)$. This concludes the proof. ■

*Theorem 5:* The Minimum Temporal Hiding problem given the closeness temporal centrality cannot be approximated within a ratio of $(1-\epsilon)\ln n$ for any $\epsilon > 0$, unless $P = NP$.

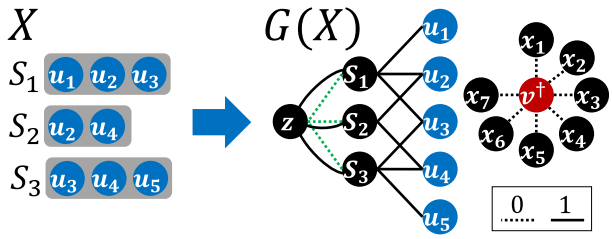


Fig. 4. Construction used in the proof of Theorem 5. On the left there is an instance $X$ of the Minimum Set Cover problem, while on the right there is a network $G(X)$ constructed as part of the Minimum Temporal Hiding problem instance. The blue nodes in $G(X)$ correspond to the elements of the universe $U$, while the evader is marked red. Dotted lines correspond to contacts at moment 0, and solid lines to contact at moment 1. Green lines depict the contacts allowed to be added.

*Proof:* In order to prove the theorem, we will use the result by Dinur and Steurer [40] that the Minimum Set Cover problem cannot be approximated within a ratio of $(1-\epsilon)\ln n$ for any $\epsilon > 0$, unless $P = NP$.

We first give a high-level overview of the proof of inapproximability of the problem. We will show that any algorithm that can approximate the Temporal Hiding problem for the temporal closeness centrality with ratio $r$, can be used to approximate the Minimum Set Cover problem with ratio $r$. To this end, we take an arbitrary instance of the Set Cover problem, construct a corresponding instance of the Minimum Temporal Hiding problem, and prove that their optimal solutions are of the same size. Hence, if there would exist a sub-logarithmic approximation algorithm for the Temporal Hiding problem, this would imply an existence of a sub-logarithmic algorithm for the Minimum Set Cover problem. However, from the aforementioned result by Dinur and Steurer [40] we know that such an algorithm cannot exist. Therefore, there can be no sub-logarithmic approximation algorithm for the Temporal Hiding problem.

Let $X = (U, S)$ be an instance of the Minimum Set Cover problem, where $U$ is the universe $\{u_1, \ldots, u_{|U|}\}$, and $S$ is a collection $\{S_1, \ldots, S_{|S|}\}$ of subsets of $U$. The goal here is to find subset $S^* \subseteq S$ such that the union of $S^*$ equals $U$ and the size of $S^*$ is minimal.

First, we will show a function $f(X)$ that based on an instance of the Minimum Set Cover problem $X = (U, S)$ constructs an instance of the Minimum Temporal Hiding problem. Let us assume that $|S| \geq 3$, note that all instances where $|S| < 3$ can be solved in polynomial time.

Let the temporal network $G(X) = (V, K, T)$ be defined as:

1) $V = \{v^\dagger, z\} \cup U \cup S \cup \bigcup_{i=1}^{|U|+|S|-1} \{x_i\}$,
2) $K = \bigcup_{x_i \in V} \{(v^\dagger, x_i, 0)\} \cup \bigcup_{S_i \in V} \{(z, S_i, 1)\} \cup \bigcup_{u_i \in S_j} \{(u_i, S_j, 1)\}$,
3) $T = 2$.

An example of the construction of the network $G(X)$ is presented in Fig. 4.

The formula of function $f$ is then $f(X) = (G(X), v^\dagger, \widehat{A}, \widehat{R}, c, \omega)$, where:

1) $G(X)$ is the temporal network we just constructed,
2) $v^\dagger$ is the evader,
3) $\widehat{A} = \{(z, S_i, 0) : S_i \in V\}$,
4) $\widehat{R} = \emptyset$, i.e., none of the edges can be removed,
5) $c$ is the temporal closeness centrality,
6) $\omega = 1$ is the safety threshold.

Let $A^*$ be the solution to the constructed instance of the Minimum Temporal Hiding problem $f(X)$ (notice that since $\widehat{R} = \emptyset$, we must have $R^* = \emptyset$). The function $g$ computing corresponding solution to the instance $X$ of the Minimum Set Cover problem is now $g(X, A^*) = \{S_i \in S : (z, S_i, 0) \in A^*\}$, i.e., $S^*$ is the set of all sets $S_i$ such that their corresponding nodes $S_i$ are connected with $z$ via the contacts in $A^*$.

Now, we will show that $g(X, A^*)$ is indeed a correct solution to $X$, i.e., that it covers the entire universe. Let $q_A$ denote the number of nodes $u_j \in U$ such that $\lambda^*_{G(X)\cup A}(z, u_j) < \infty$ for a given set of contacts $A \subseteq \widehat{A}$. Centrality values after adding $A$ to the network $G(X)$ are as follows:

1) $c_C(G(X) \cup A, v^\dagger) = \frac{2(|U|+|S|-1)}{n-1}$,
2) $c_C(G(X) \cup A, z) = \frac{2(|S|+q_A)}{n-1}$,
3) $c_C(G(X) \cup A, S_i) \leq \frac{2(|U|+1)}{n-1} < c_C(G(X) \cup A, v^\dagger)$ for every $S_i \in V$,
4) $c_C(G(X) \cup A, u_i) = \frac{2(|\{S_j \in S : u_i \in S_j\}|)}{n-1} \leq \frac{2|S|}{n-1} < c_C(G(X) \cup A, v^\dagger)$ for every $u_i \in V$,
5) $c_C(G(X) \cup A, x_i) = \frac{2}{n-1} < c_C(G(X) \cup A, v^\dagger)$ for every $x_i \in V$.

Therefore, the safety threshold $\omega$ is satisfied if and only if $c_C(G(X) \cup A, z) > c_C(G(X) \cup A, v^\dagger)$, which is the case when $q_A = |U|$. Hence, if $A^*$ is a solution to $f(X)$, then for every node $u_j$ there exists a node $S_i$ such that $(z, S_i, 0) \in A^*$ and $u_j$ is connected with $S_i$. However, because of the way we constructed the network $G(X)$, this can only be the case when $u_j \in S_i$ in the given instance $X$ of the Minimum Set Cover problem. Therefore, for every element of the universe $u_j$ there exists a set $S_i$ such that $u_j \in S_i$ and $S_i$ in $g(X, A^*)$, which implies that $g(X, A^*)$ is a solution to the given instance $X$ of the Minimum Set Cover problem, i.e., it covers the universe. What is more, since $|g(X, A^*)| = |A^*|$, we also have that the optimal solutions to both instances are of the same size.

Now, assume that there exists an $r$-approximation algorithm for the Minimum Temporal Hiding problem where $r = (1-\epsilon)\ln|\widehat{A}|$ for some $\epsilon > 0$. Let us use this algorithm to solve the constructed instance $f(X)$ and consider solution $g(X, A^*)$ to the given instance $X$ of the Minimum Set Cover problem. Since the size of the optimal solution is the same for both instances, we obtained an approximation algorithm that solves Minimum Set Cover problem to within $(1-\epsilon)\ln n$ for $\epsilon > 0$. However, Dinur and Steurer [40] showed that the Minimum Set Cover problem cannot be approximated within a ratio of $(1-\epsilon)\ln n$ for any $\epsilon > 0$, unless $P = NP$. Therefore, such approximation algorithm for the Minimum Temporal Hiding problem cannot exist, unless $P = NP$. This concludes the proof. ■

*Theorem 6:* The Minimum Temporal Hiding problem given the betweenness temporal centrality cannot be approximated within a ratio of $(1-\epsilon)\ln n$ for any $\epsilon > 0$, unless $P = NP$.

## V. Empirical Analysis

In this section, we present the experimental analysis of the problem using simulations.

TABLE II
DATASETS CONSIDERED IN THE SIMULATIONS

| Network | $\|V\|$ | $\|K\|$ |
|---|---|---|
| Bluetooth [41] | 74 | 87491 |
| Call center [42] | 52 | 1182 |
| Cambridge [43] | 187 | 8769 |
| Conference 1 [44] | 113 | 8218 |
| Conference 2 [43] | 199 | 27165 |
| Conference 3 [45] | 402 | 28954 |
| Conference 4 [45] | 361 | 19304 |
| Copenhagen call [46] | 484 | 1667 |
| Copenhagen SMS [46] | 535 | 6165 |
| Diary [47] | 49 | 1427 |
| High school 1 [48] | 312 | 28780 |
| High school 2 [48] | 310 | 35592 |
| High school 3 [48] | 303 | 30383 |
| High school 4 [48] | 295 | 28112 |
| High school 5 [48] | 299 | 26391 |
| Hospital [49] | 75 | 9313 |
| Hospital colocation [49] | 73 | 35200 |
| Intel [43] | 113 | 7408 |
| Kenya [50] | 52 | 2070 |
| Office [51] | 92 | 2679 |
| Office colocation [51] | 95 | 45357 |
| Polish manufacturer email [52] | 167 | 43360 |
| Primary school 1 [53] | 236 | 52339 |
| Primary school 2 [53] | 238 | 56313 |
| Reality [54] | 64 | 4251 |
| Romania [55] | 42 | 19045 |
| St Andrews [56] | 25 | 1483 |
| Undergrads call [41] | 75 | 3574 |
| Undergrads SMS [41] | 41 | 758 |
| University couples call [57] | 126 | 31155 |
| University couples SMS [57] | 110 | 11962 |

### A. Datasets

In our simulations we consider a number of real-life temporal network datasets. All datasets are presented in Table II.

Given that the time resolutions of different datasets are vastly different, we perform the following normalization procedure. We divide the time interval between the first and the last timestamp in the original dataset into 1000 equal subintervals. For each pair of nodes we consider them to have contact at time step $t$ if and only if the $t$-th subinterval contains at least one contact in the original dataset. The length of the time interval of such normalized dataset is then $T = 1000$.

The third column of Table II contains the number of contacts after performing the normalization procedure. Such normalized datasets are then used in the simulations.

### B. Heuristics

As discussed in Section IV, the task of finding an optimal way to hide from temporal centralities is computationally intractable. Hence, we now consider a number of heuristic solutions that add or remove contacts from the network in the hope of improving the concealment of the evader. In these heuristics all random choices are made uniformly:

**Remove Random** Randomly selects a neighbor of the evader, then removes one of the contacts with it (chosen randomly).

**Remove Min** Removes one of the contacts (chosen randomly) with a neighbor who has the smallest number of contacts with the evader.

**Remove Max** Removes one of the contacts (chosen randomly) with a neighbor who has the greatest number of contacts with the evader.

**Add Random** Randomly selects a neighbor of the evader, then adds another contact with it (the moment of which is chosen randomly).

**Add Min** Adds a contact (the moment of which is chosen randomly) with a neighbor who has the smallest number of contacts with the evader.

**Add Max** Adds a contact (the moment of which is chosen randomly) with a neighbor who has the greatest number of contacts with the evader.

**Add New** Adds a contact (the moment of which is chosen randomly) between the evader and its second degree neighbor (chosen randomly), i.e., a node that has at least one common neighbor with the evader, but is not the evader's neighbor.

For both adding and removing contacts we consider one heuristic selecting contacts to add or remove uniformly at random, and a few with a more strategic approach, focused either on finding new neighbors or on minimizing/maximizing the degree of the affected neighbor.

### C. Basic Simulations

Given a temporal network $G$ and a centrality measure $c$, we select 10 top nodes in the centrality ranking as the potential evaders. We then attempt to hide each such evader $v^\dagger$ using each of the considered heuristics described above. For each heuristic, we add or remove 5% of the evader's contacts (at least ten for nodes with exceptionally few contacts).

We measure the following values before and after the hiding process:

1) the position of $v^\dagger$ in the ranking of $c$; this value serves as a measure of how well $v^\dagger$ is able to hide from $c$,
2) the influence of $v^\dagger$ over the network, measured as the average probability that $v^\dagger$ gets infected if a Susceptible-Infected process starts in a different node times the expected number of infected nodes if a Susceptible-Infected process starts in $v^\dagger$ (we consider the process in which the probability of infection is 10%); this value serves as a measure of the influence of $v^\dagger$ over the network.

We repeated the simulation 100 times for each network and presented the results as averages.

Fig. 5 presents the results of our simulations regarding the trade-off between hiding (measured using the centrality ranking) and influence (measured using the Susceptible-Infected process). As shown in the figure, in the vast majority of cases, the removal of contacts results in the evader being more hidden, i.e., dropping in centrality ranking and becoming less influential (see how the data points corresponding to the removal heuristics are mostly grouped in quadrant II of the figure). At the same time, adding contacts to the network has

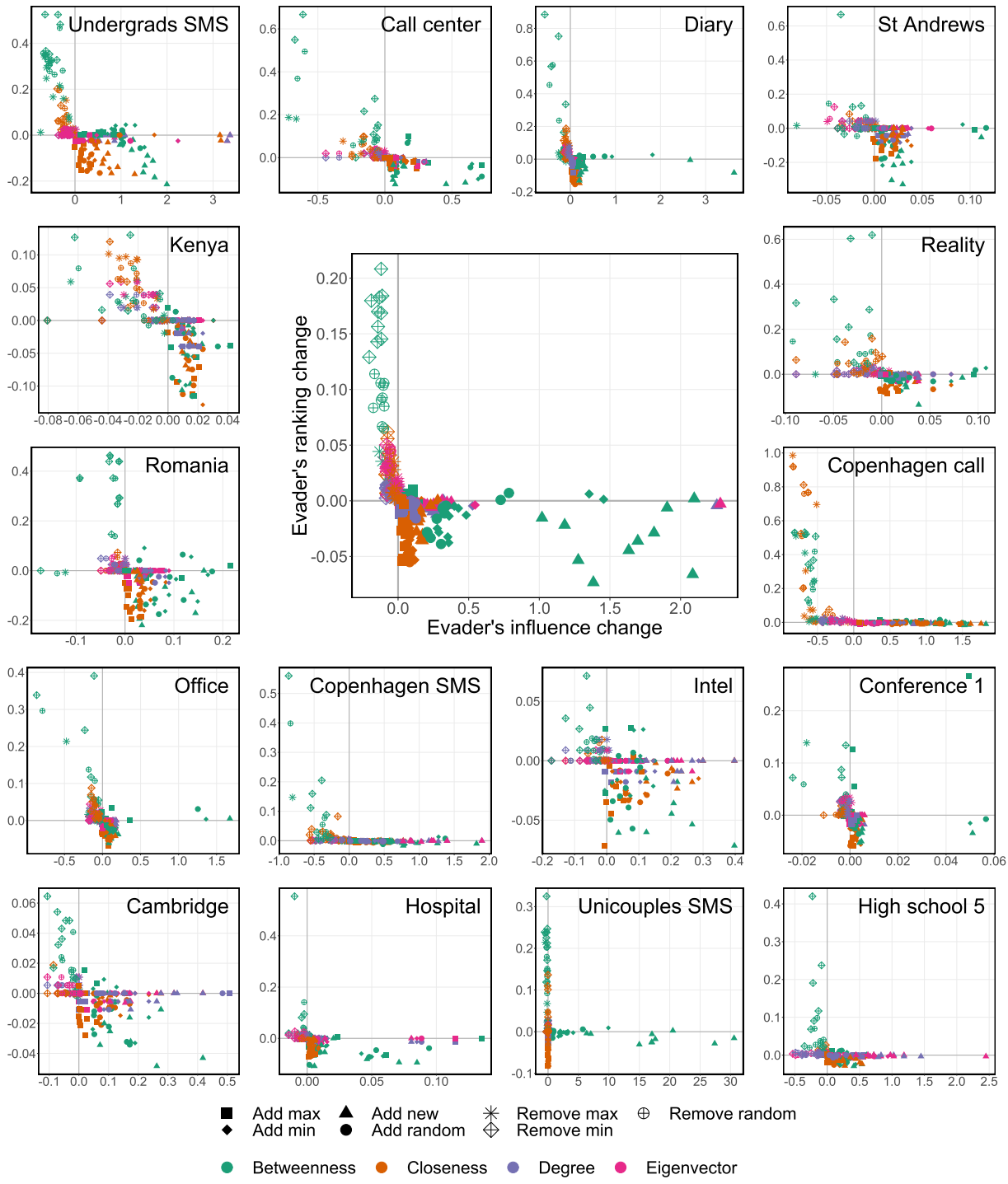


Fig. 5. Effects of hiding the top 10 nodes in each centrality ranking. Smaller plots present results for different real-life networks. The large central plot presents data averaged over all networks (with each point corresponding to an average over nodes at a particular position in the ranking of a given centrality). The y-axis corresponds to the change in the evader's centrality ranking relative to the number of all nodes (with positive values indicating that the evader becomes better hidden due to the heuristic). The x-axis corresponds to the evader's influence relative to their initial influence. Different colors correspond to centrality measures, while various shapes correspond to heuristics.

the opposite effects, i.e., the evader becomes more influential and more exposed to temporal centrality analysis (see how the data points corresponding to the addition heuristics are mostly grouped in quadrant IV of the figure). It suggests a centrality-influence trade-off in temporal networks. When comparing the effects of heuristics performing random changes, be it adding or removing contacts, with those performing strategic manipulation, i.e., Add/Remove Min/Max and Add New heuristics, we can see that using strategic heuristics result in a more pronounced effect, i.e., the magnitude of change in centrality ranking or influence value is greater than in the case of their random counterparts. If the goal is to hide from temporal

TABLE III
TEMPORAL NODE DESCRIPTORS

| Type | Descriptor | Definition |
|---|---|---|
| Time evolution | $\phi_C$ | Fraction of node's contacts when half of all contacts happened |
| | $\phi_T$ | Fraction of node's contacts at half the time $T$ |
| | $\rho_C$ | Fraction of node's edges present when half of all contacts happened |
| | $\rho_T$ | Fraction of node's edges present at half the time $T$ |
| | $\rho_c$ | Fraction of node's edges present at both the first and last $5\%$ of all contacts |
| | $\rho_t$ | Fraction of node's edges present at both the first and last $5\%$ of the time $T$ |
| Edge activity | $\varepsilon_m$ | Mean of intercontact times over node's edges |
| | $\varepsilon_s$ | Standard deviation of intercontact times over node's edges |
| | $\varepsilon_v$ | Coefficient of variation of intercontact times over node's edges |
| | $\varepsilon_k$ | Skewness of intercontact times over node's edges |
| | $\lambda_m$ | Mean of the number of other contacts between two consecutive contacts of a node's edge |
| | $\lambda_s$ | Standard deviation of the number of other contacts between two consecutive contacts of a node's edge |
| | $\lambda_v$ | Coefficient of variation of the number of other contacts between two consecutive contacts of a node's edge |
| | $\lambda_k$ | Skewness of the number of other contacts between two consecutive contacts of a node's edge |
| Node activity | $\nu_m$ | Like $\varepsilon_m$ but for node's contacts |
| | $\nu_s$ | Like $\varepsilon_s$ but for node's contacts |
| | $\nu_v$ | Like $\varepsilon_v$ but for node's contacts |
| | $\nu_k$ | Like $\varepsilon_k$ but for node's contacts |
| | $\eta_m$ | Like $\lambda_m$ but for node's contacts |
| | $\eta_s$ | Like $\lambda_s^c$ but for node's contacts |
| | $\eta_v$ | Like $\lambda_v^c$ but for node's contacts |
| | $\eta_k^c$ | Like $\lambda_k^c$ but for node's contacts |
| Network structure | $\delta$ | Degree of the node |
| | $\delta_R$ | Degree of the node divided by the average degree |
| | $\kappa$ | Number of contacts of the node |
| | $\kappa_R$ | Number of contacts of the node divided by the average number of contacts |
| | $\zeta$ | Local clustering coefficient of the node |

centralities, the Remove Min heuristic offers the best performance on average, probably as it can easily lead to disconnecting the evader from one of their neighbors completely. On the other, executing the Add New heuristic does not help the evader to hide, but is most effective in increasing the evader's influence over the network. Finally, when comparing different centrality measures, the betweenness centrality is, on average, the most sensitive to manipulation, i.e., executing the hiding process results in the greatest change in the betweenness centrality ranking. The next most sensitive centrality measure is the closeness centrality, with the degree and the eigenvector centrality measures being the most resilient on average.

### D. Regression Analysis

The results of the simulations presented so far give us some insight into how effective different hiding heuristics can be. Still, they do not explain what qualities of the evader determine how successfully it can hide from temporal centralities. To this end, we now perform the regression analysis of the hiding process. Since the analysis presented in the previous section showed that only removal heuristics successfully hide the evader, we will focus our attention on them.

For each node considered an evader in our experiments, we compute several descriptors, i.e., its characteristics relating to its contacts and the network structure in which it is embedded. We consider four different categories of descriptors: time evolution, edge activity, node activity, and network structure. All descriptors are presented in Table III.

Fig. 6 presents the regression coefficients computed using the Lasso regression methods [58]. Since the Lasso regression performs variable selection, not all descriptors appear in the figure (the coefficients for the omitted descriptors are deemed zero). As seen from the figure, the exact values of the coefficients depend on the temporal centrality under consideration. Nevertheless, we can see some trends.

There are two node descriptors that show strong positive correlation with the evader's ability to hide for the majority of the temporal centralities. The first of these descriptors is the average intercontact time of the evader's contacts $\nu_m$, while the other is the average number of other contacts between two consecutive contacts over the evader's contacts $\eta_m$. Interestingly, both of them describe the distribution of the entirety of the evader's contacts over time (see how they belong to the "node activity" type in Table III), rather than focusing on the contacts with each neighbor separately (as for the "edge activity" type in Table III). This result suggests that nodes who wish to hide from temporal centrality measures should space out their contacts evenly in time.

When it comes to the descriptors that show negative correlation with hiding abilities, they vary greatly between the centrality measures. Interestingly, the local clustering coefficient

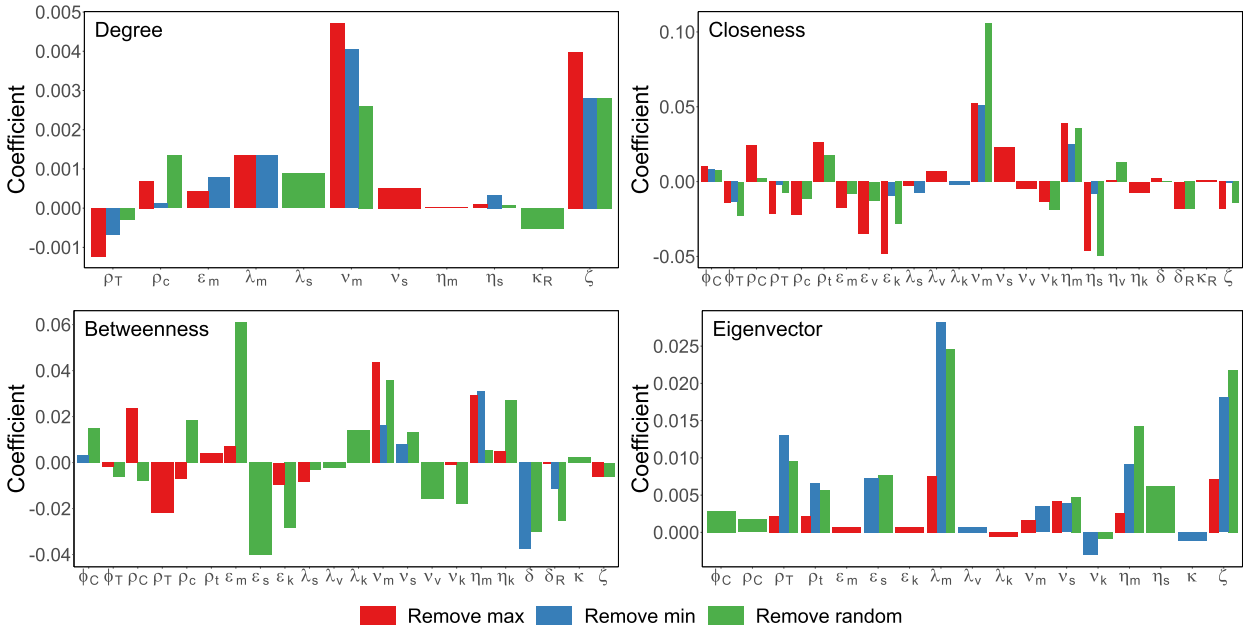


Fig. 6. Lasso regression coefficients of the node descriptors. In each plot the x-axis corresponds to descriptors (defined in Table III) with non-zero coefficients in the Lasso regression, while the y-axis corresponds to the value of said coefficients. Different colors corresponds to different hiding heuristics.

of the evader $\zeta$ has positive regression coefficient values for more locally-oriented centralities, i.e., degree and eigenvector, but negative values for more global centralities, i.e., closeness and betweenness. This finding suggests that the knowledge about which centrality measure will be used to analyze the network may be very valuable for the hiding node.

## VI. Conclusion

In this article, we analyzed both theoretically and empirically the problem faced by an evader—a member of a temporal social network who wishes to avoid being detected by temporal centrality measures. As part of the theoretical analysis, we defined the decision and the optimization versions of the problem faced by the evader. We proved that the decision version is NP-complete even for very basic temporal centrality measures. As for the optimization version, we presented a 2-approximation algorithm for the degree temporal centrality while showing that the problem is inapproximable within logarithmic bounds for the closeness and betweenness centrality measures. As part of the empirical analysis, we compared the effectiveness of several hiding heuristics in real-life temporal social networks, finding that removing existing contacts is significantly more effective in avoiding detection by temporal centralities than adding new contacts. Moreover, using regression analysis, we determined that the nodes whose contacts are more distributed throughout the analyzed period are, on average, more successful in obscuring their central position in the network. More broadly, our study contributes to the literature on hiding from social network analysis tools by considering temporal networks, which are not only more general, but also much more challenging to analyze compared to static networks.

Our findings suggest that while it is computationally infeasible to find an optimal way of hiding from temporal centrality measures, skillful manipulation of one's contacts can still help maintain some semblance of safety even when utilizing heuristic solutions. At the same time, considering the temporal aspects of the network activity offers much more diverse ways of hiding one's importance. Whereas in a static network, a member of the network can only decide who to connect or avoid, in a temporal network, they can also determine the distribution of the contacts, thus creating a much richer strategic space that should be further analyzed in the future research.

## References

[1] J. Isaak and M. J. Hanna, "User data privacy: Facebook, Cambridge analytica, and privacy protection," *Computer*, vol. 51, no. 8, pp. 56–59, 2018.
[2] C. Jernigan and B. F. Mistree, "Gaydar: Facebook friendships expose sexual orientation," *First Monday*, vol. 14, no. 10, 2009.
[3] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in *Proc. ACM Workshop Privacy Electron. Soc.*, 2005, pp. 71–80.
[4] A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: Inferring user profiles in online social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 251–260.
[5] EU, "Regulation 2016/679 of the European Parliament and of the Council of 27 Apr. 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC," *Official J. Eur. Union,* vol. L119, pp. 1–88, May 2016.
[6] J. I. Lane, V. Stodden, S. Bender, and H. Nissenbaum, Eds., *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. New York, NY, USA: Cambridge Univ. Press, 2014.
[7] M. Kearns, A. Roth, Z. S. Wu, and G. Yaroslavtsev, "Private algorithms for the protected in social network search," *Proc. Nat. Acad. Sci. USA*, vol. 113, 2016, Art. no. 201510612.
[8] M. Hay, G. Miklau, D. Jensen, P. Weis, and S. Srivastava, "Anonymizing social networks," Comput. Sci. Dept., Univ. Massachusetts Amherst, Tech. Rep. 180, 2007, pp. 7–19.
[9] V. Khatri and C. V. Brown, "Designing data governance," *Comm. ACM*, vol. 53, no. 1, pp. 148–152, 2010.

[10] B. C. Fung, K. Wang, R. Chen, and P. S. Yu, "Privacy-preserving data publishing: A survey of recent developments," *Assoc. Comput. Machinery Comput. Surv.*, vol. 42, no. 4, pp. 1–53, 2010.

[11] S. Yu *et al.*, "Target defense against link-prediction-based attacks via evolutionary perturbations," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 2, pp. 754–767, Feb. 2021.

[12] T. P. Michalak, T. Rahwan, and M. Wooldridge, "Strategic social network analysis," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 727–732.

[13] P. Holme and J. Saramäki, "Temporal networks," *Phys. Rep.*, vol. 519, no. 3, pp. 97–125, 2012.

[14] J.-P. Eckmann, E. Moses, and D. Sergi, "Entropy of dialogues creates coherent structures in e-mail traffic," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 40, pp. 14333–14337, 2004.

[15] J. Candia, M. C. González, P. Wang, T. Schoenharl, G. Madey, and A.-L. Barabási, "Uncovering individual and collective human dynamics from mobile phone records," *J. Phys. A*, vol. 41, no. 22, 2008, Art. no. 224015.

[16] T. M. Przytycka, M. Singh, and D. K. Slonim, "Toward the dynamic interactome: It's about time," *Brief. Bioinf.*, vol. 11, no. 1, pp. 15–29, 2010.

[17] A. Rao, A. O. Hero, D. J. States, and J. D. Engel, "Inferring time-varying network topologies from gene expression data," *EURASIP J. Bioinf. Syst. Biol.*, vol. 2007, pp. 1–12, 2007.

[18] T. Takaguchi, M. Nakamura, N. Sato, K. Yano, and N. Masuda, "Predictability of conversation partners," *Phys. Rev. X*, vol. 1, no. 1, 2011, Art. no. 011008.

[19] P. Holme and N. Masuda, "The basic reproduction number as a predictor for epidemic outbreaks in temporal networks," *PLoS One*, vol. 10, no. 3, 2015, Art. no. e0120567.

[20] P. Holme, "Epidemiologically optimal static networks from temporal network data," *PLoS Comput. Biol.*, vol. 9, no. 7, 2013, Art. no. e1003142.

[21] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Phys. Rev. E*, vol. 85, no. 2, 2012, Art. no. 026107.

[22] J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia, "Analysing information flows and key mediators through temporal centrality metrics," in *Proc. 3rd Workshop Social Netw. Syst.*, 2010, pp. 1–6.

[23] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks," *Phys. Rev. E*, vol. 84, no. 1, 2011, Art. no. 016105.

[24] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, "Eigenvector-based centrality measures for temporal networks," *Multiscale Model. Sim.*, vol. 15, no. 1, pp. 537–574, 2017.

[25] L. Lv, K. Zhang, T. Zhang, X. Li, J. Zhang, and W. Xue, "Eigenvector centrality measure based on node similarity for multilayer and temporal networks," *IEEE Access*, vol. 7, pp. 115725–115733, 2019.

[26] D. Taylor, M. A. Porter, and P. J. Mucha, "Supracentrality analysis of temporal networks with directed interlayer coupling," in *Temporal Network Theory*. Cham, Switzerland: Springer, 2019, pp. 325–344.

[27] M. Waniek, T. P. Michalak, M. J. Wooldridge, and T. Rahwan, "Hiding individuals and communities in a social network," *Nature Hum. Behav.*, vol. 2, no. 2, pp. 139–147, 2018.

[28] M. Waniek, T. P. Michalak, T. Rahwan, and M. Wooldridge, "On the construction of covert networks," in *Proc. 16th Conf. Auton. Agents MultiAgent Syst.*, 2017, pp. 1341–1349.

[29] P. Dey and S. Medya, "Covert networks: How hard is it to hide?," in *Proc. 18th Int. Conf. Auton. Agents MultiAgent Syst.*, Montreal, Canada, 2019, pp. 628–637.

[30] T. Was, M. Waniek, T. Rahwan, and T. Michalak, "The manipulability of centrality measures-An axiomatic approach," in *Proc. 19th Int. Conf. Auton. Agents MultiAgent Syst.*, 2020, pp. 1467–1475.

[31] M. Waniek, J. Woundefinednica, K. Zhou, Y. Vorobeychik, T. Rahwan, and T. P. Michalak, "Strategic evasion of centrality measures," in *Proc. 20th Int. Conf. Auton. Agents MultiAgent Syst.*, 2021, pp. 1389–1397.

[32] M. Waniek, T. Michalak, and T. Rahwan, "Hiding in multilayer networks," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, pp. 1021–1028.

[33] M. Waniek, K. Zhou, Y. Vorobeychik, E. Moro, T. P. Michalak, and T. Rahwan, "How to hide one's relationships from link prediction algorithms," *Sci. Rep.*, vol. 9, no. 1, 2019, Art. no. 12208.

[34] K. Zhou, T. P. Michalak, M. Waniek, T. Rahwan, and Y. Vorobeychik, "Attacking similarity-based link prediction in social networks," in *Proc. 18th Int. Conf. Auton. Agents Multi-Agent Syst.*, 2019, pp. 305–313.

[35] P. Dey and S. Medya, "Manipulating node similarity measures in networks," in *Proc. 19th Int. Conf. Auton. Agents MultiAgent Syst.*, 2020, pp. 321–329.

[36] M. Waniek, M. Cebrian, P. Holme, and T. Rahwan, "Social diffusion sources can escape detection," 2021, *arXiv:2102.10539*.

[37] Q. Huang, C. Zhao, X. Zhang, X. Wang, and D. Yi, "Centrality measures in temporal networks with time series analysis," *Event Process. Lang. Europhys. Lett.*, vol. 118, no. 3, 2017, Art. no. 36001.

[38] R.-R. Yin, Q. Guo, J.-N. Yang, and J.-G. Liu, "Inter-layer similarity-based eigenvector centrality measures for temporal networks," *Physica A*, vol. 512, pp. 165–173, 2018.

[39] D. Taylor, M. A. Porter, and P. J. Mucha, "Tunable eigenvector-based centralities for multiplex and temporal networks," *Multiscale Model. Sim.*, vol. 19, no. 1, pp. 113–147, 2021.

[40] I. Dinur and D. Steurer, "Analytical approach to parallel repetition," in *Proc. 46th Annu. ACM Symp. Theory Comput.*, 2014, pp. 624–633.

[41] A. Madan, M. Cebrian, S. Moturu, K. Farrahi, and A. Pentland, "Sensing the 'health state' of a community," *IEEE Pervasive Comput.*, vol. 11, no. 4, pp. 36–45, Oct.–Dec. 2012.

[42] D. Olguin, B. Waber, T. Kim, A. Mohan, K. Ara, and A. Pentland, "Sensible organizations: Technology and methodology for automatically measuring organizational behavior," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 39, no. 1, pp. 43–55, Feb. 2009.

[43] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. Mob. Comput.*, vol. 6, no. 6, pp. 606–620, Jun. 2007.

[44] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, "What's in a crowd? Analysis of face-to-face behavioral networks," *J. Theor. Biol.*, vol. 271, no. 1, pp. 166–180, 2011.

[45] J. Stehlé *et al.*, "Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees," *Bounded Model Checking Med.*, vol. 9, no. 1, p. 87, 2011.

[46] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, J. E. Larsen, and S. Lehmann, "Measuring large-scale social networks with high resolution," *PLoS One*, vol. 9, no. 4, 2014, Art. no. e95978.

[47] J. M. Read, K. T. Eames, and W. J. Edmunds, "Dynamic social networks and the implications for the spread of infectious disease," *J. Roy. Soc. Interface*, vol. 5, no. 26, pp. 1001–1007, 2008.

[48] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys," *PLoS One*, vol. 10, no. 9, 2015, Art. no. e0136497.

[49] P. Vanhems *et al.*, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PLoS One*, vol. 8, no. 9, 2013, Art. no. e73970.

[50] M. C. Kiti *et al.*, "Quantifying social contacts in a household setting of rural Kenya using wearable proximity sensors," *EPJ Data Sci.*, vol. 5, no. 1, pp. 1–21, 2016.

[51] M. Genois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat, "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers," *Netw. Sci.*, vol. 3, pp. 326–347, 2015. [Online]. Available: http://journals.cambridge.org/article_S2050124215000107

[52] R. Michalski, S. Palus, and P. Kazienko, "Matching organizational structure and social network extracted from email communication," in *Lecture Notes in Business Information Processing*, vol. 87. Berlin, Germany: Springer, 2011, pp. 197–206.

[53] J. Stehlé *et al.*, "High-resolution measurements of face-to-face contact patterns in a primary school," *PLoS One*, vol. 6, no. 8, 2011, Art. no. e23176.

[54] N. Eagle and A. S. Pentland, "Reality mining: Sensing complex social systems," *Pers. Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, 2006.

[55] R.-C. Marin, C. Dobre, and F. Xhafa, "Exploring predictability in mobile interaction," in *Proc. 3rd Int. Conf. Emerg. Intell. Data Web Technol.*, 2012, pp. 133–139.

[56] G. Bigwood, T. Henderson, D. Rehunathan, M. Bateman, and S. Bhatti, "Crawdad dataset St Andrews/Sassy (v. 2011-06-03)" CRAWDAD wireless network data archive, 2011. Accessed: Jun. 2011. [Online]. Available: http://crawdad.org/st_andrews/sassy/20110603/mobile

[57] N. Aharony, W. Pan, C. Ip, I. Khayal, and A. Pentland, "Social fMRI: Investigating and shaping social mechanisms in the real world," *Pervasive Mob. Comput.*, vol. 7, no. 6, pp. 643–659, 2011.

[58] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy Statist. Soc. Ser. B. Statist. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

**Marcin Waniek** received the Ph.D. degree in computer science from the University of Warsaw, Warsaw, Poland. He is currently a Postdoctoral Associate with Computer Science Department, New York University Abu Dhabi, Abu Dhabi, UAE. His research interests include social network analysis, graph theory, computational complexity theory, artificial intelligence, and game theory. It earned him the Polish Artificial Intelligence Society Award for the Best Ph.D. Dissertation in artificial intelligence in 2017.

**Petter Holme** received the M.A. degree in chinese from Stockholm University, Stockholm, Sweden, and the Ph.D. degree in theoretical physics from Umeå University, Umeå, Sweden. He is a Specially Appointed Professor with the World Research Hub Initiative, Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan. Before joining Tokyo Institute of Technology, he was a Professor of energy science with Sungkyunkwan University, Seoul, South Korea. He has coauthored more than 150 research papers. His research interests include networks, ranging from data-driven topics about social, biological and technological systems, and from theoretical questions. His current research focuses on integrating temporal information into network modeling.

**Talal Rahwan** received the Ph.D. degree in computer science from The University of Southampton, Southampton, U.K., in 2007. He is currently an Associate Professor of computer science and the Director of the Data Science and AI Lab, New York University Abu Dhabi, Abu Dhabi, UAE. His research interests include data science, computational social science, game theory, and artificial intelligence. He was the recipient of the Dean's Award for Early Career Researcher from the University of Southampton. His Ph.D. thesis earned the British Computer Society's Distinguished Dissertation Award, which annually recognizes the most outstanding Ph.D. thesis in computer science, U.K. He was selected by the IEEE Computer Society as one of the 10 most promising, young Artificial Intelligence (AI) researchers in the world. His work appeared in main academic journals, including *Nature Communications*, *Nature Human Behaviour*, and *Nature Machine Intelligence*. His research was the recipient of the coverage from international media outlets, including The Boston Globe, WIRED, Scientific American, and Nature Middle East.