On the Construction of Covert Networks

Marcin Waniek
University of Warsaw
vua@mimuw.edu.pl

Tomasz P. Michalak
University of Oxford
& University of Warsaw
tomasz.michalak@cs.ox.ac.uk

Talal Rahwan
Masdar Institute of Science and Technology
trahwan@gmail.com

Michael Wooldridge
University of Oxford
Michael.Wooldridge@cs.ox.ac.uk

ABSTRACT

Centrality measures are widely used to identify leaders of covert networks. We study how a group of such leaders can avoid being detected by such measures. More concretely, we study the hardness of choosing a set of edges that can be added to the network in order to decrease the leaders' ranking according to two fundamental centrality measures, namely degree, and closeness. We prove that this problem is NP-complete for each measure. We then study how the leaders can construct a network from scratch, designed specifically for them to hide in disguise. We identify a network structure that not only guarantees to hide the leaders to a certain extent, but also allows them to spread their influence across the network.

Keywords

Social Network Analysis, Centrality, Hiding in Networks

1. INTRODUCTION

Mapping terrorist networks is of vital importance to any counter-terrorism efforts. Not only does this help to understand their operational structure and *modus operandi*, but it also plays a key role in designing and implementing destabilization strategies [5, 13, 28]. One of the most common such strategies requires identifying individuals that are suspected to play central roles in the organization [11, 12]. To this end, *centrality measures*—metrics developed in graph theory to quantify the importance of nodes in networks—are often used in the analysis of covert networks [24, 26, 17]. Arguably, the three fundamental such measures are: (i) *Degree centrality*, which ranks each node based on the number of neighbours they it has; (ii) *Closeness centrality*, which ranks each node based on its average distance to other nodes; and (iii) *Betweenness centrality*, which ranks each node based on the relative number of shortest paths that go through that node.

Unfortunately, understanding how criminals organize themselves in a network is challenging at various levels [22, 32]: the data may be incomplete, the nature of the relationship between two criminals may be unclear, and the network may evolve continuously. The literature on this research problem generally agrees that criminals in general, and terrorists in particular, face a trade-off between *secrecy* and *efficiency* [29] though the way in which both factors are modelled differs. Overall, two approaches in this literature can be distinguished, which we briefly discuss next.

In the first approach, researchers study known topologies of historical or contemporary criminal networks, with the aim being to understand why particular structures have emerged [8, 9, 21]. Perhaps the most comprehensive study in this body of research is due to Kilberg [21], who analyzed an extensive dataset of more than 240 terrorist networks, and provided a classification of those networks based on their structure and functionality. Furthermore, using regression analysis, the author tried to quantify the degree to which the shape of terrorist networks is influenced by such variables as the GDP level of the target country, the political rights and civil liberties therein, and the inclination to attack police and military targets in that country.

In this article we contribute to the second approach in the literature, which is more theoretical in nature and aims to explain the structural properties of covert networks by developing explicit models of the terrorists' preferences and the different choices they face [10, 18, 23]. With such analyses, certain network topologies typically emerge as the result of modelling the terrorists as rational decision makers. A notable example of such a model is that of Lindelauf et al. [23], who consider the tradeoff between secrecy and operational efficiency of a terrorist network and borrows concepts from both game theory and graph theory to identify more fitting topologies. Arguably, it is less efficient if a message has to be passed many times from one person to another (i.e., the shortest path from the sender to the receiver is relatively long). Based on this, Lindelauf et al. defined efficiency as the (normalized) reciprocal of the total distance of the graph (i.e., the sum of shortest distances between any two nodes in the network). Secrecy, in turn, is defined for each node and is proportional to the fraction of the network that remains unexposed when this node is detected. Secrecy of the network is the sum of the secrecy scores over all nodes.

In this paper, we also propose a theoretical model to study the secrecy-efficiency tradeoff. However, our model differs from previous ones in a number of ways. Firstly, inspired by studies of real-life covert networks [6, 25], we take a *leader-centric approach*, i.e., we focus on the role played in terrorist networks by their leaders. In more detail, we investigate how

Appears in: *Proc. of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2017), S. Das, E. Durfee, K. Larson, M. Winikoff (eds.), May 8–12, 2017, São Paulo, Brazil.*

the topology of the network could be deliberately designed to keep the leader(s) identity hidden. In this context, while the previous literature on identifying leaders of terrorist networks typically assumed that such leaders are not aware of the techniques and methods used by law-enforcement agencies, we assume that this is not the case, i.e., in our model the terrorist leaders *strategically shape their network to shield themselves from detection by the centrality measures*. In fact, recent media reports and academic studies of criminal and terrorist organizations suggested that members of such organizations are becoming increasingly tech-savvy [30, 19]. Hence, their obliviousness with respect to the available social network analysis techniques should not be taken for granted.

As already argued, secrecy is not the only objective that the leaders of a terrorist network may have. Indeed, if they were concerned only with hiding themselves, they would simply cut most (if not all) of their connections in the network. This, however, would clearly impair the leaders' efficiency. In our model, the efficiency of the leaders is defined as their *influence over the network*. In other words, the leaders in our model face the trade off between hiding from centrality measures, and influencing the network members. Note that a node's influence over a network can be quantified according to various models, most notably the *Independent Cascade* model [16] and the *Linear Threshold* model [20].

In the first part of the paper, we focus on the computational aspects of modifying an existing network so as to shield the leaders from centrality analysis by hiding them among the followers. More in detail, we analyze the hardness of identifying a set of edges to be added between the followers so that the *ranking* of the leaders (based on one of the three main centrality measures) drops below a certain threshold. At first glance, this problem may appear to be easy at least for the degree centrality, which is mathematically uninvolved. Indeed, it is straightforward to decrease (*the value of*) degree centrality—it simply requires to cut edges [34]. Surprisingly, however, we find that our problem of decreasing the *ranking* of a node according to degree centrality is much more challenging. In particular, Theorems 1 state that the above problem is NP-Complete for degree centrality. Theorem 2, in turn, states the same for closeness centrality. This latter result is in line with the literature on modifying a network to increase centrality [7].

Given this hardness of modifying an existing network, we turn out attention to a different question, which is *how a terrorist network could be built from scratch so that the leaders are hidden and, at the same time, have a reasonable influence over the network members*. Here, the main idea is for the leaders to surround themselves with an "inner circle" of trustees, called "*captains*", whose role is to conceal the leaders, and to pass on their influence to the rest of the network. We identify one such network structure, and prove that every captain is guaranteed to be ranked higher than any of the leaders (according to the three standard centrality measures). In fact, "inner circles" have been identified in various real-life terrorist networks such as, e.g., Al-Quaeda [2] and IRA [33]. While we do not have access to data that confirms that those real-life "inner circles" have similar structure to the ones obtained in this article, we hope that our results shed more light on why such circles may exist in covert networks. In this context, charting the topology of covert networks became one of the key research directions.

2. PRELIMINARIES

In this section, we present some basic notation and concepts that will be used throughout the paper.

2.1 Basic Network Notation

Let $G = (V, E) \in \mathbb{G}$ denote a network, where $V = \{v_1, \dots, v_n\}$ is the set of n nodes and $E \subseteq V \times V$ is the set of edges. We denote an edge between nodes v_i and v_j by (v_i, v_j). In this article we consider *undirected* networks, in which E is a set of unordered pairs, *i.e.*, we do not discern between edges (v_i, v_j) and (v_j, v_i). We also assume that networks do not contain self-loops, *i.e.*, $\forall_{v_i \in V} (v_i, v_i) \notin E$.

A path in a network $G = (V, E)$ is an ordered sequence of distinct nodes, $p = \langle v_{i_1}, \dots, v_{i_k} \rangle$, in which every two consecutive nodes are connected by an edge in E. We consider the length of a path to be the number of edges in that path. We denote the set of all shortest paths between a pair of nodes, $v_i, v_j \in V$ by $\Pi_G(v_i, v_j)$. The distance between a pair of nodes $v_i, v_j \in V$, *i.e.*, the length of a shortest path between them, is denoted by $d_G(v_i, v_j)$. Furthermore, a network is said to be *connected* if and only if there exists a path between every pair of nodes in that network.

We denote by $N_G(v_i)$ the set of *neighbours* of v_i in G, *i.e.*, $N_G(v_i) = \{v_j \in V : (v_j, v_i) \in E\}$. Finally, we denote by $N_G(v_i, v_j)$ the set of common neighbours of nodes v_i and v_j, *i.e.*, $N_G(v_i, v_j) = N_G(v_i) \cap N_G(v_j)$.

To make the notation more readable, we will often denote two arbitrary nodes by v and w, instead of v_i and v_j. Moreover, we will often omit the network itself from the notation whenever it is clear from the context, *e.g.*, by writing $d(v, w)$ instead of $d_G(v, w)$. This applies not only to the notation presented thus far, but to also to all future notation.

2.2 Centrality Measures

The concept of *centrality* in human organizations was introduced by Bavelas [3]. Intuitively, a centrality measure is a function, $c : \mathbb{G} \times V \to \mathbb{R}$, that expresses the relative importance of any given node in any given network. Arguably, the three best-known centrality measures are *degree*, *closeness* and *betweeness* [15].

Degree centrality was introduced by Shaw [31]. It assumes that the importance of a node is proportional to the number of its neighbours. The normalized degree centrality of a node $v_i \in V$ in a network G is defined as follows:

$$c_{degr}(G, v_i) = \frac{|N(v_i)|}{n - 1}.$$

Closeness centrality, introduced by Beauchamp [4], quantifies the importance of a node in terms of shortest distances from this node to all other nodes in the network. As such, the most important node is the one with the shortest average path length to all other nodes. The normalized closeness centrality of a node $v_i \in V$ in a connected network G can be expressed as:

$$c_{clos}(G, v_i) = \frac{n - 1}{\sum_{v_j \in V} d(v_i, v_j)}.$$

Betweenness centrality was developed independently by Anthonisse [1] and Freeman [14]. This measure quantifies the importance of a given node in the context of network flow. In more detail, if we consider all the shortest paths in the network, then the more such paths that traverse through a given node, the more important the role of that node in the network. The normalized betweenness centrality of a node $v_i \in V$ in a connected network G can be expressed as:

$$c_{betw}(G, v_i) = \frac{2}{(n-1)(n-2)} \sum_{v_j, v_k \in V \setminus \{v_i\}} \frac{|\{p \in \Pi(v_j, v_k) : v_i \in p\}|}{|\Pi(v_j, v_k)|}$$

2.3 Models of Influence

The propagation of influence through the network is often described in terms of node activation. When a certain node is sufficiently influenced by its neighbours, it becomes "active". It then starts to influence any "inactive" neighbours, and so on. To initiate this propagation process, a set of nodes (known as the *seed set*) must be activated right from the start. Assuming that time moves in discrete rounds, we denote by $I(t) \subseteq V$ the set of nodes that are active at round t, implying that $I(1)$ is the seed set. The way influence propagates to inactive nodes depends on the influence model under consideration. Arguably, the two main models of influence are:

- *Independent Cascade* [16]: In this model, every pair of nodes is assigned an activation probability, $p : V \times V \to [0, 1]$. Then, in every round, $t > 1$, every node $v_i \in V$ that became active in round $t-1$ activates every inactive neighbour, $v_j \in N(v_i) \setminus I(t-1)$, with probability $p(v_i, v_j)$. The process ends when there are no newly activated nodes, *i.e.*, when $I(t) = I(t-1)$.

- *Linear Threshold* [20]: In this model, every node $v_i \in V$ is assigned a *threshold value*, t_{v_i}, which is sampled (according to some probability distribution) from the set $\{0, \ldots, |N(v_i)|\}$. Then, in every round, $t > 1$, every inactive node v_i becomes active, *i.e.*, becomes a member of $I(t)$, if $|I(t-1) \cap N(v_i)| \geq t_{v_i}$. The process ends when there are no newly activated nodes, *i.e.*, when $I(t) = I(t-1)$.

In either model, the influence of a node, v_i, on another node, v_j, is denoted by $\mathit{inf}_G(v_i, v_j)$ and is defined as *the probability that v_j gets activated given the seed set $\{v_i\}$*. We assume that $\mathit{inf}_G(v_i, v_i) = 0$ for all $v_i \in V$. The influence of v_i over the entire network G is then defined as $\mathit{inf}_G(v_i) = \sum_{v_j \in V} \mathit{inf}_G(v_i, v_j)$.

3. PROBLEM STATEMENT & ITS THEORETICAL ANALYSIS

In this section we state the main theoretical problem of this work and prove its NP-completeness.

As mentioned in the introduction, we assume that the terrorist network is composed of two types of agents: the *leaders* and the *followers*. Furthermore, we assume that the leaders are aware that law-enforcement agencies may use centrality analysis to identify them. Thus, the leaders would like to strategically modify the existing network so that their centrality becomes lower than a certain predefined threshold $d \in \mathbb{N}$ that we refer to as *safety margin*. To achieve this objective, no more than $b \in \mathbb{N}$ modifications can be made to the network (b can be thought of as a "budget" to spend). Since removing edges would mean that existing communication link is severed, we assume that the network can be modified only by adding edges. Furthermore, since adding an edge to any leader increases this node's degree centrality, we assume that edges can be added only between followers.

Formally, we define *the problem of Hiding Leaders* as follows:

Definition 1 (Hiding Leaders). *This problem is defined by a tuple, (G, L, b, c, d), where $G = (V, E) \in \mathbb{G}$ is a network, $L \subset V$ is a set of leaders to be hidden, $b \in \mathbb{N}$ is a budget specifying the maximum number of edges that can be added, $c : \mathbb{G} \times V \to \mathbb{R}$ is a centrality measure, and $d \in \mathbb{N}$ is a chosen safety margin. Then, if we denote by $F = V \setminus L$ the set of "followers", the goal is then to identify a set of edges to be added to the network, $A^* \subseteq F \times F$, such that $|A^*| \leq b$ and the resulting network $G' = (V, E \cup A^*)$ contains at least d followers that each have a centrality score higher than that of any leader, i.e.:*

$$\exists_{F' \subseteq F} |F'| \geq d \wedge \forall_{f \in F'} \forall_{l \in L} c(G', f) > c(G', l)$$

Intuitively, the above problem should be easy to solve for the degree centrality measure. Indeed, adding an edge between any two (disconnected) followers, increases their degree centrality with respect to all the leaders. However, we prove below that the problem is in fact NP-complete for the degree centrality measure.

Theorem 1. *The problem of Hiding Leaders is NP-complete given the degree centrality.*

Proof. The problem is trivially in NP, since after the addition of a given A^* it is possible to compute the degree centrality for all nodes in polynomial time.

Next, we prove that the problem is NP-hard. To this end, we propose a reduction from the NP-complete problem of *Finding k-clique*. The decision version of this problem is defined by a network, $G = (V, E)$, and a constant, $k \in \mathbb{N}$, where the goal is to determine whether there exist k nodes in G that form a clique.

Let us assume that $k \geq 3$ (if $k = 2$ then the problem is trivial). Given an instance of the problem of *Finding k-clique*, defined by some $k \geq 3$ and a network $G = (V, E)$, let us construct a network, $H = (V', E')$, as follows:

- **The set of nodes:**: For every node, $v_i \in V$, we create a single node, v_i, as well as $n - 1 - |N_G(v_i)|$ other nodes, denoted by $X = \{x_{i,1}, \ldots, x_{i,n-1-|N_G(v_i)|}\}$. Additionally, we create one node called y, as well as $n + k - 1$ other nodes, namely $L' = l_1, \ldots, l_{n+k-1}$;

- **The set of edges:** We create an edge between two nodes $v_i, v_j \in V$ if and only if this edge was not present in G, *i.e.*, $(v_i, v_j) \in E' \iff (v_i, v_j) \notin E$. Additionally, for every v_i we create an edge (v_i, y) as well as an edge $(v_i, x_{i,j})$ for every $x_{i,j}$. We also create an edge (l_i, l_j) between every pair of nodes $l_i, l_j \in L'$, except for the edge (l_1, l_2). Finally, we create two additional edges, (l_1, y) and (l_2, y).

An example of such a H network is illustrated in Figure 1.

Now, consider the following instance of the problem of hiding leaders, (H, L, b, c, d), where:

- $H = (V', E')$ is the network we just constructed;

- $L = V' \setminus V$;

- $b = \frac{k(k-1)}{2}$;

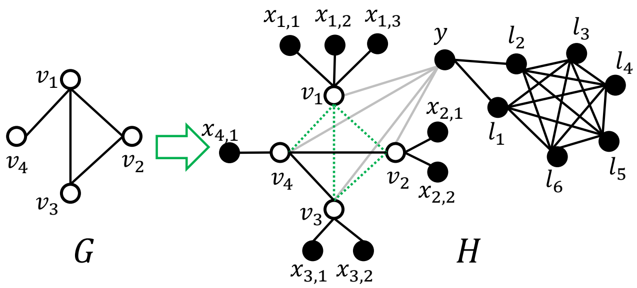

Figure 1: An illustration of the network used in the NP-completeness proof of the problem of Hiding Leaders given the degree centrality.

- c is the degree centrality measure;
- $d = k$.

Next, we reduce the problem of Finding k-cliques in G to the aforementioned instance of Hiding Leaders in H. To this end, from the definition of the problem of Hiding Leaders, we know that the edges to be added to H must be chosen from $F \times F$. Since in our instance we have: $F = V' \setminus L = V' \setminus (V' \setminus V) = V$, then the edges to be added to H must be chosen from $V \times V$. However, since the edges in $(V \times V) \setminus E$ are already present in H (see how H is created), then the edges to be added to H must be chosen from E. Out of those edges, we need to choose subset, $A^* \subseteq E$, as a solution to the problem. In what follows, we will show that a solution to the above instance of the Hiding Leaders in H corresponds to a solution to the problem of *Finding k-clique* in G.

First, note that each of the k nodes with the highest degree centrality in H must be a member of L'. This is because there are more than k nodes in L', each of which has a degree of $n+k-2$, while the degree of every node in $V' \setminus L'$ is smaller than $n + k - 2$. Thus, in order for A^* to be a solution to the problem of hiding leaders, the addition of A^* to H must increase the degree of at least k nodes in V such that each of them has a degree of at least $n + k - 1$ (note that the addition of A^* only increases the degrees of nodes in V, as we already established that $A^* \subseteq E$). Now since in H the degree of every node in V equals n (because of the way H is created), then in order to increase the degree of k such nodes to $n + k - 1$, each of them must be an end of *at least* $k - 1$ edges in A^*. But since the budget in our problem instance is $\frac{k(k-1)}{2}$, then the only possible choice of A^* is the one that increases the degree of *exactly* k nodes in V by *exactly* $k - 1$. If such a choice of A^* is available, then surely those k nodes would form a clique in G, since all $\frac{k(k-1)}{2}$ edges in A^* are taken from G. □

Having proven the NP-completeness of the problem given the degree centrality, we next prove its NP-completeness given the closeness centrality.

THEOREM 2. *The problem of Hiding Leaders is NP-complete given the closeness centrality.*

PROOF. The problem is trivially in NP, since after the addition of a given A^* it is possible to compute the closeness centrality for all nodes in polynomial time.

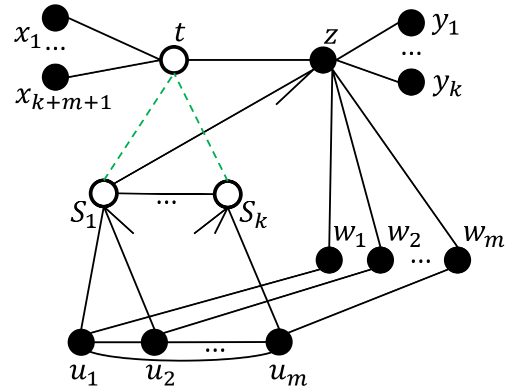

Figure 2: An illustration of the network used in the NP-completeness proof of the problem of Hiding Leaders given the closeness centrality.

Next, we prove that the problem is NP-hard. To this end, we propose a reduction from the NP-complete *3-Set-Cover* problem. The decision version of this problem is defined by a universe $U = \{u_1, \dots, u_m\}$ and a collection of sets $S = \{S_1, \dots, S_q\}$ such that $\forall_i S_i \subset U \land |S_i| = 3$, where the goal is to determine whether there exist $a \leq b$ elements of S the union of which equals U.

Given an instance of the *3-Set-Cover* problem, let us construct a network G as follows:

- **The set of nodes:**: For every $S_i \in S$, we create a single node denoted by S_i, and for every $u_i \in U$, we create two nodes denoted by u_i and w_i. We denote the set of every S_i node by S, the set of every u_i node by U, and the set of every w_i node by W. In addition, we create $k+m+1$ nodes denoted by $X = \{x_1, \dots, x_{k+m+1}\}$, and k nodes denoted by $Y = \{y_1, \dots, y_k$. Lastly, we create two additional nodes, denoted by t and z;
- **The set of edges:** First, we create the edge (t, z). Then, for every node x_i we create an edge (x_i, t); for every node y_i we create an edge (y_i, z), for every node w_i we create the edges (w_i, z) and (w_i, u_i), and every node S_i we create an edge (S_i, u_j) for every $u_j \in S_i$. After that, we create b edges, $(z, S_1), \dots, (z, S_b)$. Finally, we create edges such that the nodes in S form a clique, and those in U also form a clique. That is, we create an edge (u_i, u_j) for every $u_i, u_j \in U$ and an edge (S_i, S_j) for every $S_i, S_j \in S$.

An example of the resulting network, G, is illustrated in Figure 2.

Now, consider the following instance of the problem of hiding leaders, (G, L, b, c, d), where:

- G is the network we just constructed;
- $L = \{z\} \cup X \cup Y \cup W \cup U$;
- b is the parameter of the *3-Set-Cover* problem (where the goal is to determine whether there exist $a \leq b$ elements of S the union of which equals U);
- c is the closeness centrality measure;
- $d = 1$.

From the definition of the problem of Hiding Leaders, we see that the only edges that can be added to the graph are edges between t and the members of S, *i.e.*, $A^* \subseteq \hat{A}$, where $\hat{A} = \{(t, S_1), \ldots, (t, S_k)\}$. Notice that any such choice of A^* corresponds to selecting a subset of $|A^*|$ elements of S in the *3-Set-Cover* problem. In what follows, we will show that a solution to the above instance of Hiding Leaders corresponds to a solution to the *3-Set-Cover* problem.

First, we show that for every $v \in V \setminus \{t, z\}$ and every $A^* \subseteq \hat{A}$ we either have $c(G', v) < c(G', t)$ or $c(G', v) < c(G', z)$, where $G' = (V, E \cup A^*)$. To this end, we show that the following holds, where $D(G', v) = \frac{n-1}{c(G',v)} = \sum_{w \in V \setminus \{v\}} d(v, w)$:

$$\forall v \in V \setminus \{t, z\} \forall_{A^* \subseteq \hat{A}} D(G', v) > D(G', t) \vee D(G', v) < D(G', z)$$

Let d_t denote $\sum_{u_i \in U} d(t, u_i) + \sum_{S_i \in S} d(t, S_i)$. In what follows, we compute $D(G', v)$ for different types of v. While doing so, the expression on the right-hand side of each equality (or inequality) will have exactly seven terms: the 1st equals $d(v, z)$; the 2nd equals $d(v, t)$; the 3rd equals $\sum_{x_i \in X} d(v, x_i)$; the 4th equals $\sum_{y_i \in Y} d(v, y_i)$; the 5nd equals $\sum_{w_i \in W} d(v, w_i)$; the 6th equals $\sum_{u_i \in U} d(v, u_i)$; the 7th equals $\sum_{S_i \in S} d(v, S_i)$. This will hold for every type of v, except for the case when $v = x_i$:

- $D(G', z) = (0) + (1) + (2(k+m+1)) + (k) + (m) + (2m) + (b + 2(k-b)) = 5m + 5k - b + 3$;
- $D(G', t) = (1) + (0) + (k+m+1) + (2k) + (2m) + (\sum_{u_i \in U} d(t, u_i)) + (\sum_{S_i \in S} d(t, S_i)) = 3m + 3k + 2 + d_t$;
- $D(G', x_i) = (2) + (1) + (2k + 2m) + (3k) + (3m) + (m + \sum_{u_i \in U} d(t, u_i)) + (k + \sum_{S_i \in S} d(t, S_i)) = 6m + 6k + 3 + d_t$ $>$ D(G',t);
- $D(G', y_i) = (1) + (2) + (3(k+m+1)) + (2(k-1)) + (2m) + (3m) + (2b + 3(k-b)) = 8m + 8k - b + 4 > D(G', z)$;
- $D(G', w_i) \geq (1) + (2) + (3(k+m+1)) + (2k) + (2(m-1)) + (1 + 2(m-1)) + (2k) = 7m + 7k + 3 > D(G', z)$ because we have $d(w_i, S_j) \geq 2$;
- $D(G', u_i) \geq (2) + (2) + (3(k+m+1)) + (3k) + (1 + 2(m-1)) + (m-1) + (k) = 6m + 7k + 5 > D(G', z)$ because we have $d(u_i, t) \geq 2$ and $d(u_i, S_j) \geq 1$;
- $D(G', S_i) \geq (1) + (1) + (2(k+m+1)) + (2k) + (2m) + (3 + 2(m-3)) + (k-1) = 6m + 5k > D(G', z)$; because we have $d(S_i, z) \geq 1$ and $d(S_i, t) \geq 1$.

Therefore, either t or z has the highest closeness centrality. Since $z \in L$ and $t \in F$, then $A^* \subseteq \hat{A}$ is a solution to the problem of problem of Hiding Leaders if and only if $D(G', t) < D(G', z)$. This is the case when:

$$d_t < 2m + 2k - b + 1.$$

Let $U_A = \{u_i \in U : \exists_{S_j \in S} u_i \in S_j \wedge (t, S_j) \in A^*\}$. We have that $d_t = |A^*| + 2(k - |A^*|) + 2|U_A| + 3(m - |U_A|)$ which gives us:

$$d_t = 3m - |U_A| + 2k - |A^*|$$

Since by definition $|U_A| \leq m$ and $|A^*| \leq b$, it is possible that $d_t < 2m + 2k - b + 1$ only when $|U_A| = m$ and $|A^*| = b$, *i.e.*, $\forall_{u_i \in U} \exists_{S_j \in S} u_i \in S_j \wedge (t, S_j) \in A^*$. This solution to the problem of Hiding Leaders corresponds to a solution to the given instance of the *3-Set-Cover* problem, which concludes the proof. □

The main idea behind the technique we used to prove Theorems 1 and 2 was to introduce certain sets of nodes (namely X in the proof Theorem 1, and X, Y in the proof of Theorem 2) such that the difference in centrality between leaders and followers can only be reduced by adding edges in a way that solves the corresponding NP-complete problem. While this idea worked well for degree and closeness centrality, things become more complicated when dealing with betweenness centrality. Specifically, while the addition of a single node usually has a slight impact on the degree and closeness centrality of other nodes, the same is not true for betweenness centrality. This is because a new node introduces at least $n-1$ new shortest paths that can be controlled by other nodes, which may end up changing their betweenness centrality significantly.

4. CAPTAIN NETWORK

In the previous section, we proved the NP-completeness of modifying an existing network in order to hide its leaders. However, in certain cases, the leaders are to develop a new terrorist network (e.g. a subnetwork in a foreign country) rather than to modify an existing one. In this section we show that it is possible to efficiently create a network from scratch, designed specifically to hide its leaders without limiting their ability to influence the other nodes in the network. We call this the "*captain*" network. Here is how it works. First the leader nodes, L, form a clique, to provide the best possible communication among them. Each leader $l_i \in L$ is then assigned a group of k "*captains*", $C = \{c_{i,1}, \ldots, c_{i,k}\}$, which are connected to that leader. All captains are then connected into a complete $|L|$-partite graph. A captain, $c_{i,j}$ servers two purposes: the first is to conceal the leaders in L, by ensuring that it is ranked higher than each of them (according to the three standard centrality measures); the second purpose of $c_{i,j}$ is to pass on the influence of l_i to the rest of the network. The remaining nodes, the set of which is $X = \{x_1, \ldots, x_m\}$, are each connected to one captain from each group. Note that the set of followers in this network is $F = X \cup C_1 \cup \ldots \cup C_h$ Figure 3 illustrates a sample captain network with $|L| = 3$, whereas Algorithm 1 summarized the steps that create such a network.

Note that if the above steps are followed given a single leader, the result would be a tree structure. While a tree is a fairly common organizational structure, it may not provide adequate disguise of the leader, especially if the leader is identified as a root of the tree. With this in mind, whenever there is a single leader, we create two groups of captains to avoid the tree structure. The resulting structure is illustrated in Figure 4.

Next, we prove that every captain has a higher centrality value than any of the leaders.

Theorem 3. *Given a captain network, let $r = \left\lfloor \frac{m}{k} \right\rfloor$ denote the minimal number of connections that a captain, $c_{i,j}$, has with nodes from X. Then, if either ($h \geq 2$ and $r \geq 1$), or ($h = 1$ and $k < \sqrt{|F| + 1} - 1$), then all captains have greater degree, closeness and betweenness centrality than any of the leader nodes.*

Proof. Starting with degree centrality and multiple leaders, the degree of a leader node, l, is $c_{degr}(G, l) = \frac{h+k-1}{n-1}$,

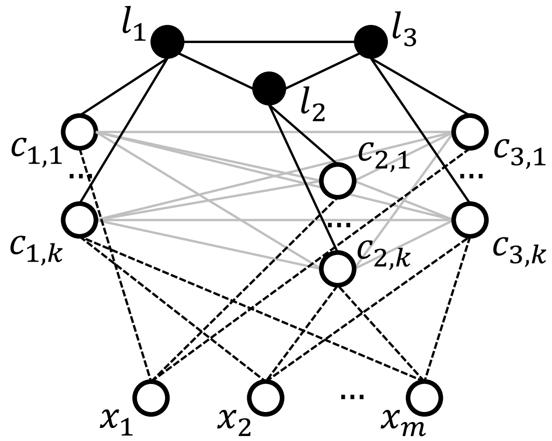

Figure 3: An illustration of a captain network with $|L| = 3$. Edges including leaders are depicted as solid black lines; edges between captains are depicted as gray lines; edges between captains and other nodes are depicted as dotted lines.

since it is only connected to other leaders and captains from its group. On the other hand, the degree of a captain, $c_{i,j}$, is $c_{degr}(G, c_{i,j}) \geq \frac{1+k(h-1)+r}{n-1}$, since it is connected to one of the leader nodes, to all captains from other groups, and to at least r other nodes from X. As such, we have:

$$c_{degr}(G, c_{i,j}) - c_{degr}(G, l) \geq \frac{1 + k(h-1) + r - (h + k - 1)}{n-1}$$

which gives us:

$$c_{degr}(G, c_{i,j}) - c_{degr}(G, l) \geq \frac{(h-2)(k-1) + r}{n-1}$$

Therefore, since $h \geq 2$, $k \geq 1$, and $r \geq 1$, we have that $c_{degr}(G, c_{i,j}) > c_{degr}(G, l)$ for any $c_{i,j}$.

As for the case with a single leader, the degree of the leader node, l, is $c_{degr}(G, l) = \frac{2k}{n-1}$, since it is only connected to captains from both groups. On the other hand, the degree of a captain $c_{i,j}$, is $c_{degr}(G, c_{i,j}) \geq \frac{1+k+r}{n-1}$, since it is connected to the leader node, to all captains from other groups, and to at least r members. As such, we have:

$$c_{degr}(G, c_{i,j}) - c_{degr}(G, l) \geq \frac{1 + k + r - 2k)}{n-1}$$

which gives us:

$$c_{degr}(G, c_{i,j}) - c_{degr}(G, l) \geq \frac{1 + r - k}{n-1}$$

Therefore, since $r = \left\lfloor \frac{m}{k} \right\rfloor$, we have that $c_{degr}(G, c_{i,j}) > c_{degr}(G, l)$ for $k < \sqrt{|F| + 1} - 1$.

Moving on to closeness centrality, for any given node, v, this centrality depends inversely on the sum of the lengths of shortest paths from v to every other nodes, *i.e.*, $\sum_{w \in V} d(v, w)$. For every leader and every captain, the distance to every other node is either 1 or 2. More precisely, for every $v \in V$, we have: $\sum_{w \in V} d(v, w) = 1|N(v)| + 2(n - |N(v)|) = 2n - |N(v)|$. Consequently, whenever all captains have greater degree centrality than all leaders, they must also have greater closeness centrality. Since we have already proven this fact

Algorithm 1 The construction of a captain network with multiple leaders

Input: The set of leaders $L = \{l_1, \ldots, l_h\}$, the set of followers $F = \{f_1, \ldots, f_{|F|}\}$, the number of captains in each group, i.e., k (where $1 \leq k \leq \frac{|F|}{|L|}$).
Output: The set of edges E that constitutes the captain network

$h' \leftarrow \max(2, h)$
for $i = 1, \ldots, h'$ **do**
  for $j = 1, \ldots, k$ **do**
    $c_{i,j} \leftarrow f_{(i-1)k+j}$
    $C_i \leftarrow C_i \cup \{c_{i,j}\}$
$X \leftarrow F \setminus \bigcup_{i=1}^{h'} C_i$
for $l_i, l_j \in L$ **do**
  $E \leftarrow E \cup \{(l_i, l_j)\}$
for $l_i \in L$ **do**
  for $c_{i,j} \in C_i$ **do**
    $E \leftarrow E \cup \{(l_i, c_{i,j})\}$
if $h = 1$ **then**
  for $c_{2,j} \in C_2$ **do**
    $E \leftarrow E \cup \{(l_1, c_{2,j})\}$
for $C_i \neq C_j$ **do**
  for $c \in C_i$ **do**
    for $c' \in C_j$ **do**
      $E \leftarrow E \cup \{(c, c')\}$
$j \leftarrow 0$
for $x \in X$ **do**
  for $i = 1, \ldots, |L|$ **do**
    $E \leftarrow E \cup \{(x, c_{i,j})\}$
  $j \leftarrow (j + 1) \mod k$

for the degree centrality, then this implies that $c_{clos}(G, c_{i,j}) > c_{clos}(G, l)$.

Finally, regarding betweenness centrality, let $\zeta(v)$ denote: $\sum_{u,w \in V \setminus \{v\}} \frac{|\{p \in \Pi(u,w) : v \in p\}|}{|\Pi(u,w)|}$. Then the betweenness centrality of a node $v \in V$ can be written as:

$$c_{betw}(G, v) = \frac{2}{(n-1)(n-2)} \zeta(v).$$

For a network with multiple leaders, every leader node l belongs to one of $(h-1)k + 1$ shortest paths between pairs of captains from its group (alternative shortest paths run through captains from other groups), as well as one of $k+1$ shortest paths between each captain from its group and all other leaders (alternative shortest paths run through captains from the group of the chosen leader). Since the leader node l belongs to no other shortest paths, we have:

$$\zeta(l) = \frac{k(k-1)}{2((h-1)k+1)} + \frac{k(h-1)}{k+1}$$

Having analyzed $\zeta(l)$, let us now analyze $\zeta(c_{i,j})$ for a captain, $c_{i,j}$. In particular, since $c_{i,j}$ belongs to one of $(h-1)k + 1$ shortest paths between pairs of captains from all other groups, as well as one of $k+1$ shortest paths between each captain from other groups and the leader of its group, we have:

$$\zeta(c_{i,j}) > \frac{(h-1)k(k-1)}{2((h-1)k+1)} + \frac{k(h-1)}{k+1}$$

Therefore, we have that $\zeta(c_{i,j}) > \zeta(l)$, which results in $c_{betw}(G, c_{i,j}) > c_{betw}(G, l)$.

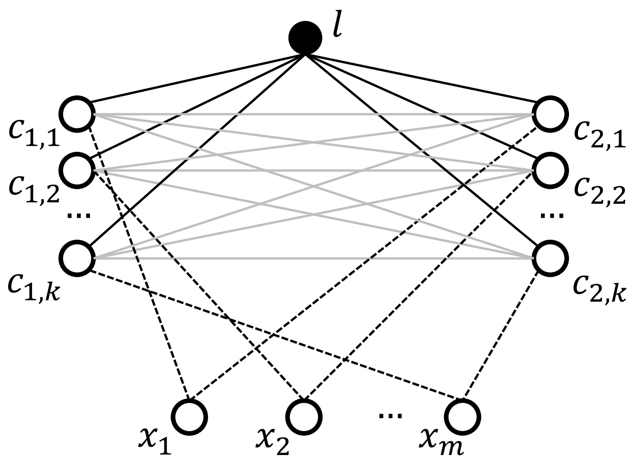

Figure 4: An illustration of a captain network with one leader. Edges including leaders are depicted as solid black lines; edges between captains are depicted as gray lines; edges between captains and other nodes are depicted as dotted lines.

For a network with a single leader, the leader node l belongs to one of $k+1$ shortest paths between pairs of captains from each group (alternative shortest paths run through captains from other group). Since the leader node l belongs to no other shortest paths, we have:

$$\zeta(l) = \frac{k(k-1)}{(k+1)}$$

Having analyzed $\zeta(l)$, let us now analyze $\zeta(c_{i,j})$ for a captain, $c_{i,j}$. In particular, since $c_{i,j}$ belongs to one of $k+1$ shortest paths between pairs of captains from the other group, and to only shortest path between member nodes connected to it and captains from the other group, we have:

$$\zeta(c_{i,j}) > \frac{k(k-1)}{2(k+1)} + rk = \frac{k(k-1) + 2rk(k+1)}{2(k+1)}$$

Therefore, we have that $\zeta(c_{i,j}) > \zeta(l)$, which implies that $c_{betw}(G, c_{i,j}) > c_{betw}(G, l)$. □

As for the followerâĂŹs centrality values, they will be lower than those of the captain. Intuitive argument is that the follower has fewer neighbours than the captain that she is connected to, and she lies on no shortest paths, as her neighbours form a clique.

5. SIMULATION RESULTS

As stated in Theorem 3, a captain network can indeed conceal its leaders as far as centrality is concerned. On the other hand, as far as influence is concerned, we evaluate the network empirically to see how the different parameters affect the influence of the leaders. To this end, given a captain network with 400 nodes, we varied the parameters of the network, either k (the size of each captain group) and q (the number of captains from each group, connected to any given member) for a network with a single leader, or k (the size of each captain group) and h (the number of leaders) for a network with multiple leaders. For every pair of parameters, we measure the difference in centrality between a leader node, and any given captain (the greater the difference, the greater the leaders' disguise), and measured the influence of a leader to see how this influence is affected by the disguising process. When measuring the influence, we use either the Independent Cascade model with probability 0.15 on each edge, or the Linear Threshold model with the threshold value sampled uniformly at random.

The results are depicted in Figures 5 and 6. Both figures should be read as follows. The x-axis represents the number of captains in each group. The y-axis represents the number of leaders of the network. The more intense the color in Figure 5, the higher the difference in centrality between a leader node and a captain, and the safer the leader. The more intense the color in Figure 6, the higher the influence of a leader node.

Roughly speaking, the results can be categorized into three categories:

- *small k*: This yields relatively high levels of disguise in terms of degree, closeness, and betweenness. On the other hand, it yields rather low levels of Independent-Cascade influence and Linear-Threshold influence;

- *large k and small h*: This yields relatively low levels of disguise in terms of degree, closeness and betweenness. On the other hand, it yields relatively high levels of Linear-Threshold influence, but not Independent-Cascade influence;

- *large k and large h*: This yields relatively high levels of disguise in terms of degree and closeness, but not betweenness. On the other hand, it yields relatively high levels of Independent-Cascade influence, but not Linear-Threshold influence.

6. DISCUSSION & CONCLUDING REMARKS

The model studied in this paper offers new insights into the secrecy-efficiency tradeoff faced by the covert organizations. The novelty of our approach comes from our definition of secrecy, which assumes that the members of a terrorist network act strategically to evade detection by centrality measures. Indeed, it is well established that centrality measures belong to the key social network analysis (SNA) tools used to analyse covert networks. Unfortunately, centrality measures—like most other SNA tools—were designed to analyse social networks among members of the general public, rather than among adroit members of covert organizations who are well aware of the possibility of attracting unwanted attention from the authorities. However, recent findings—for, instance, with respect to ISIS—strongly suggest that such an assumption is too far-fetched.

Our work constitutes a step to analyse such issues and a contributed to the line of research on strategic analysis of social networks [27]. In particular, we showed that choosing a set of edges to add to the network in order to decrease the leaders' ranking (according to both degree and closeness centrality measures) is NP-complete. While this is a "negative" result from the computational point of view, it is in fact rather positive news for law-enforcement agencies.

The above hardness results are general in the sense that they were obtained without any considerations of the "efficiency" part of the aforementioned secrecy-efficiency trade-

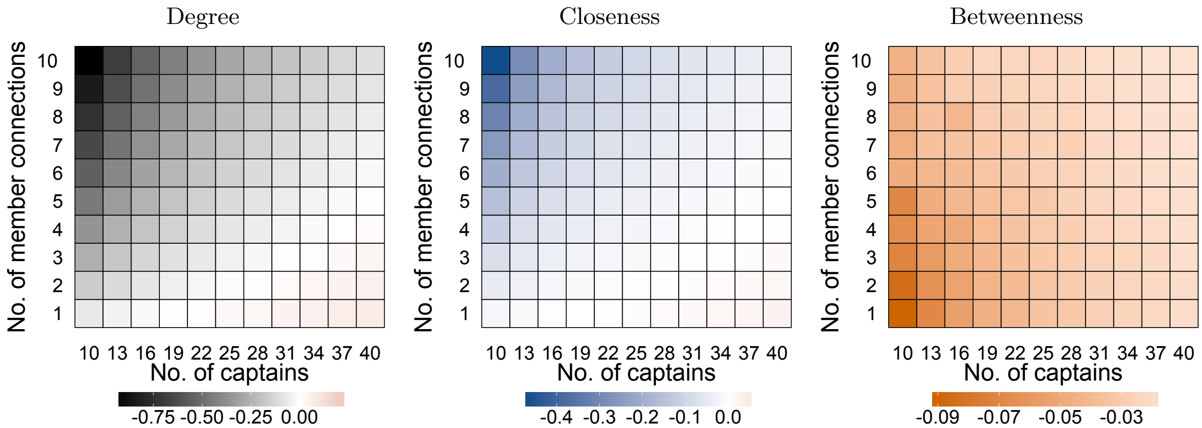

Figure 5: Given a captain network of 400 nodes, with different number of captains in each group (the x-axis) and number of leaders (the y-axis), the figure depicts the difference in centrality between a leader and a captain.

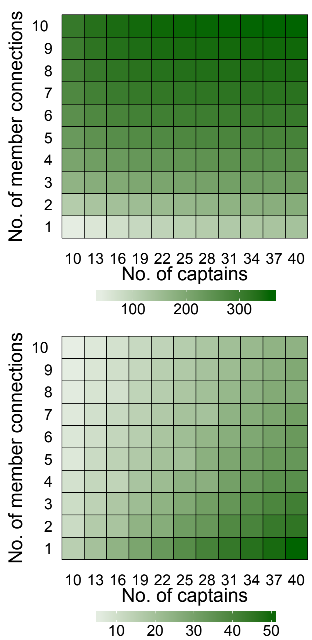

Figure 6: Given a captain network of 400 nodes, with different number of captains in each group (the x-axis) and number of leaders (the y-axis), the figure depicts the influence value of a leader. Top figure - IC Influence; bottom figure - LT Influence.

off. We introduced such efficiency into the model by investigating how the leaders could construct a network from scratch so that they are adequately hidden from the three fundamental centrality measures, and adequately influential at the same time.

The network that we construct from scratch has a group of leaders forming a clique (to assure efficient communication among them), and has a well-defined core "captains" who are densely connected among themselves and who act as intermediaries between leaders and other members of the organization. It is known that such "inner circles" exist in some real-life terrorist networks such as, e.g., Al-Quaeda [2] and IRA [33]. Unfortunately, we did not have any access to real-life data that we could use for comparison. Nevertheless, we hope that our results shed more light on why such circles may exist in covert networks.

Our model can be extended in various directions. First, we assume that the "*evaders*" (i.e., the members of covert organization) are strategic whereas the "*seeker*" (who is using centrality measures to identify key terrorist) is not, i.e., he or she is unaware of any potential strategic efforts by the evaders. It would be interesting to see new SNA tools, and centrality measures in particular, that are immune (at least to some extent) against such evasion techniques.

Second, although our captain networks appear to be effective in terms of influence (i.e., they are empirically shown to grant the leaders a reasonable level of influence), they do not provide any worst-case guarantees on solution quality in this regard. This problem constitutes another direction for future research.

Finally, another interesting direction is to investigate whether there exist special classes of networks for which the problem of hiding leaders can easily be solved.

Acknowledgements

Marcin Waniek was supported by the Polish National Science Centre grant 2015/17/N/ST6/03686. Tomasz Michalak was supported by the European Research Council under Advanced Grant 291528 ("RACE") and by the Polish National Science Centre grant 2014/13/B/ST6/01807. Michael Wooldridge was supported by the European Research Council under Advanced Grant 291528 ("RACE").

REFERENCES

[1] J. M. Anthonisse. The rush in a graph. *Amsterdam: University of Amsterdam Mathematical Centre*, 1971.
[2] V. Barber. The evolution of al qaeda's global network and al qaeda core's position within it: A network analysis. *Perspectives on Terrorism*, 9(6), 2015.
[3] A. Bavelas. A mathematical model for group structures. *Human organization*, 7(3):16–30, 1948.

[4] M. A. Beauchamp. An improved index of centrality. *Behavioral Science*, 10(2):161–163, 1965.

[5] K. M. Carley. Destabilization of covert networks. *Computational & Mathematical Organization Theory*, 12(1):51–66, 2006.

[6] J. T. Chatagnier, A. Mintz, and Y. Samban. The decision calculus of terrorist leaders. *Perspectives on Terrorism*, 6(4-5), 2012.

[7] P. Crescenzi, G. D'Angelo, L. Severini, and Y. Velaj. Greedily improving our own centrality in a network. In *Proceedings of the 14th International Symposium on Experimental Algorithms - Volume 9125*, pages 43–55, New York, NY, USA, 2015. Springer-Verlag New York, Inc.

[8] N. Crossley, G. Edwards, E. Harries, and R. Stevenson. Covert social movement networks and the secrecy-efficiency trade off: The case of the {UK} suffragettes (1906-1914). *Social Networks*, 34(4):634 – 644, 2012.

[9] F. Demiroz and N. Kapucu. Anatomy of a dark network: the case of the turkish ergenekon terrorist organization. *Trends in organized crime*, 15(4):271–295, 2012.

[10] W. Enders and X. Su. Rational terrorists and optimal network structure. *Journal of Conflict Resolution*, 51(1):33–57, 2007.

[11] S. F. Everton. Network topography, key players and terrorist networks. *Connections*, 2009.

[12] S. F. Everton and N. Roberts. Strategies for combating dark networks. *Paper presented at the Sunbelt XXIX: The Annual Meeting of the International Network of Social Net work Analysis.*, 2011.

[13] J. D. Farley. Breaking al qaeda cells: A mathematical analysis of counterterrorism operations (a guide for risk assessment and decision making). *Studies in Conflict & Terrorism*, 26(6):399–411, 2003.

[14] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

[15] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.

[16] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 9(3):1–18, 2001.

[17] I. Hamed, M. Charrad, and N. B. Ben Saoud. *Which Centrality Metric for Which Terrorist Network Topology?*, pages 195–208. Springer International Publishing, Cham, 2016.

[18] R. Janssen and H. Monsuur. Stable network topologies using the notion of covering. *European Journal of Operational Research*, 218(3):755–763, 2012.

[19] N. F. Johnson, M. Zheng, Y. Vorobyeva, A. Gabriel, H. Qi, N. Velasquez, P. Manrique, D. Johnson, E. Restrepo, C. Song, and S. Wuchty. New online ecology of adversarial aggregates: Isis and beyond. *Science*, 352(6292):1459–1463, 2016.

[20] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[21] J. Kilberg. A basic model explaining terrorist group organizational structure. *Studies in Conflict & Terrorism*, 35(11):810–830, 2012.

[22] V. E. Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

[23] R. Lindelauf, P. Borm, and H. Hamers. The influence of secrecy on the communication structure of covert networks. *Social Networks*, 31(2):126–137, 2009.

[24] R. Lindelauf, H. Hamers, and B. Husslage. Cooperative game theoretic centrality analysis of terrorist networks: The cases of jemaah islamiyah and al qaeda. *European Journal of Operational Research*, 229(1):230–238, 2013.

[25] J. Magouirk, S. Atran, and M. Sageman. Connecting terrorist networks. *Studies in Conflict & Terrorism*, 31(1):1–16, 2008.

[26] T. P. Michalak, T. Rahwan, O. Skibski, and M. Wooldridge. Defeating terrorist networks with game theory.

[27] T. P. Michalak, T. Rahwan, and M. Wooldridge. Startegic social network analysis. In *AAAI*, 2017.

[28] I.-C. Moon. *Destabilization of adversarial organizations with strategic interventions*. ProQuest, 2008.

[29] C. Morselli, C. Giguère, and K. Petit. The efficiency/security trade-off in criminal networks. *Social Networks*, 29(1):143 – 153, 2007.

[30] A. Nordrum. Pro-ISIS Online Groups Use Social Media Survival Strategies to Evade Authorities, 2016.

[31] M. E. Shaw. Group structure and the behavior of individuals in small groups. *The Journal of Psychology*, 38(1):139–149, 1954.

[32] M. K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social networks*, 13(3):251–274, 1991.

[33] R. Stevenson and N. Crossley. Change in covert social movement networks: The 'inner circle' of the provisional irish republican army. *Social Movement Studies*, 13(1):70–91, 2014.

[34] M. Waniek, T. P. Michalak, T. Rahwan, and M. Wooldridge. Hiding individuals and communities in a social network. *CoRR*, abs/1608.00375, 2016.