Strategic Evasion of Centrality Measures—Supplementary Materials

Paper #6267

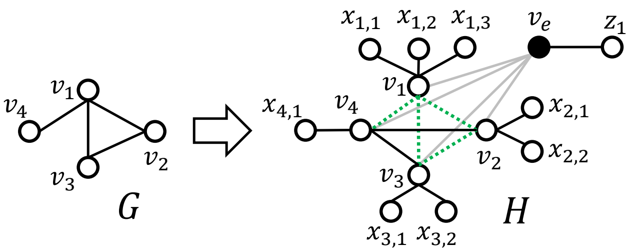

Figure S1: An illustration of the network used in the NP-completeness proof given the degree centrality.

S1 Proofs of Theoretical Results

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

Proof. The problem is trivially in NP, since after the addition of a given set of edges A^* and the removal of a given set of edges R^* 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 give a reduction from the NP-complete problem of Finding k-Clique, where the goal is to determine whether there exist k nodes in G that form a clique. Given an instance of the problem of Finding k-Clique, defined by some $k \in \mathbb{N}$ 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 $|N(v_i)|$ other nodes, denoted by $X = \{x_{i,1}, \ldots, x_{i,|N(v_i)|}\}$. Additionally, we create one node called v_e, as well as $k-2$ other nodes, namely $Z = z_1, \ldots, z_{k-2}$.
- **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, v_e) as well as an edge $(v_i, x_{i,j})$ for every $x_{i,j}$. Finally, for every node z_i we also create edge (z_i, v_e).

An example of such a network H is illustrated in Figure S1. Now, consider the instance $(H, v_e, b, c, d, \hat{A}, \hat{R})$ of the problem of Local Hiding where $H = (V', E')$ is the network we just constructed, v_e is the evader in V', $b = \frac{k(k-1)}{2}$, c is the degree centrality measure, $d = k$, $\hat{A} = E$, and $\hat{R} = \emptyset$.

From the definition of the problem we know that the edges to be added to H must be chosen from E, *i.e.*, edges from the network in the Finding k-Clique problem. Out of those edges, we need to choose a subset, $A^* \subseteq E$, as a solution to the Local Hiding problem. In what follows, we will show that a solution to the above instance of the Local Hiding problem in H corresponds to a solution to the problem of Finding k-Clique in G.

First, note that v_e has the highest degree in H, which is $n + k - 2$. Thus, in order for A^* to be a solution to the Local Hiding problem, 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, since 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 constructed), 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 edges in A^* are taken from G. □

Theorem 2. *The problem of Local Hiding is NP-complete given the closeness centrality measure.*

Proof. The problem is trivially in NP, since after the addition of a given A^*, and the removal of a given R^*, it is possible to compute the closeness 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 3-Set Cover problem. Let $U = \{u_1, \ldots, u_l\}$ denote the universe, and let $S = \{S_1, \ldots, S_m\}$ denote the set of subsets of the universe, where for every S_i we have $|S_i| = 3$. The goal is then to determine whether there exist k 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 all S_i nodes by S, the set of all u_i nodes by U, and the set of

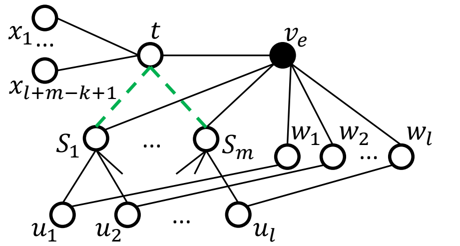

Figure S2: An illustration of the network used in the NP-completeness proof given the closeness centrality.

all w_i nodes by W. In addition, we create $l+m-k+1$ nodes denoted by $X = \{x_1, \ldots, x_{l+m-k+1}\}$. Lastly, we create two additional nodes, denoted by t and v_e.

- **The set of edges:** First, we create the edge (t, v_e). Then, for every node x_i we create an edge (x_i, t), for every node w_i we create the edges (w_i, v_e) and (w_i, u_i), and every node S_i we create an edge (S_i, v_e) and an edge (S_i, u_j) for every $u_j \in S_i$.

An example of the resulting network, G, is illustrated in Figure S2. Now, consider the following instance of the problem of Local Hiding, $(G, v_e, b, c, \hat{A}, \hat{R}, d)$, where G is the network we just constructed, v_e is the evader in G, $b = k$ (where k is the parameter of the 3-Set Cover problem), c is the closeness centrality measure, $d = 1$, $\hat{A} = \{(t, S_i) : S_i \in S\}$, and $\hat{R} = \emptyset$.

From the definition of the problem, we see that the only edges that can be added to the graph are those between t and the members of S, meaning that $A^* \subseteq \hat{A}$, where $\hat{A} = \{(t, S_1), \ldots, (t, S_m)\}$. 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 Local Hiding corresponds to a solution to the 3-Set Cover problem.

First, we will show that for every $v \in V \setminus \{t, v_e\}$ and every $A^* \subseteq \hat{A}$ we either have $c(G', v) < c(G', t)$ or have $c(G', v) < c(G', v_e)$, where $G' = (V, E \cup A^*)$. To this end, let $D(G', v)$ denote the sum of distances from v to all other nodes, *i.e.*, $D(G', v) = \sum_{w \in V \setminus \{v\}} d(v, w)$. Note that $D(G', v) = \frac{n-1}{c(G', v)}$. We will show that the following holds:

$$\forall_{v \in V \setminus \{t, v_e\}} \forall_{A^* \subseteq \hat{A}} \left(D(G', v) > D(G', t)\right) \\ \vee \left(D(G', v) > D(G', v_e)\right).$$

Let d_t denote $\sum_{u_i \in U} d(t, u_i) + \sum_{S_i \in S} d(t, S_i)$. Notice also that $k \leq m$. Next, we compute $D(G', v)$ for the different types of node v:

- $D(G', v_e) = 5l + 3m - 2k + 3$;
- $D(G', t) = 3l + m - k + 2 + d_t$;
- $D(G', x_i) = 6l + 3m - 2k + 3 + d_t > D(G', t)$;
- $D(G', w_i) = 8l + 5m - 3k + 2 > D(G', v_e)$;
- $D(G', u_i) \geq 9l + 4m - 3k + 2 > D(G', v_e)$ as $\sum_{S_j \in S} d(u_i, S_j) \geq m$;
- $D(G', S_i) \geq 7l + 4m - 2k - 4 > D(G', v_e)$ as $d(S_i, v_e) \geq 1$.

Based on this, either t or v_e has the highest closeness centrality, therefore $A^* \subseteq \hat{A}$ is a solution to the problem of Local Hiding if and only if $D(G', t) < D(G', v_e)$. This is the case when:

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

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 $d_t = |A^*| + 2(m - |A^*|) + 2|U_A| + 3(l - |U_A|)$ which gives us:

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

Since by definition $|U_A| \leq l$ and $|A^*| \leq k$, it is possible that $d_t < 2l + 2m - k + 1$ only when $|U_A| = l$ and $|A^*| = k$, *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 Local Hiding corresponds to a solution to the given instance of the 3-Set Cover problem, which concludes the proof. □

Theorem 3. *The problem of Local Hiding is NP-complete given the betweenness centrality measure.*

Proof. The problem is trivially in NP, since after the addition of a given set of edges A^*, and the removal of a given set of edges R^*, it is possible to compute the betweenness 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 3-Set Cover problem. Let $U = \{u_1, \ldots, u_l\}$ denote the universe, and let $S = \{S_1, \ldots, S_m\}$ denote the set of subsets of the universe, where for every S_i we have $|S_i| = 3$. The goal is then to determine whether there exist k 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 a single node denoted by u_i. We denote the set of all S_i nodes by S, and the set of all u_i nodes by U. In addition, we create α nodes denoted by $X = \{x_1, \ldots, x_\alpha\}$, where $\alpha = m^2 l(m + l + 2)$ and β nodes denoted by $Y = \{y_1, \ldots, y_\beta\}$, where $\beta = m^2 l(k + l + 2)$. Lastly, we create four additional nodes, denoted by t, v_e, w_1 and w_2.
- **The set of edges:** First, we create the edges (t, v_e) and (w_1, w_2). Then, for every node x_i we create an edge (x_i, t), and for every node y_i we create an edge (y_i, v_e). Moreover, for every node $S_i \in S$ we create the edges (S_i, v_e) and (S_i, w_1), as well as the edges (S_i, u_j) for every $u_j \in S_i$. We also create an edge (u_i, w_2) for every node $u_i \in U$. Finally, we create edges such that the nodes in X form a clique, and the nodes in Y form a clique, *i.e.*, we create an edge (x_i, x_j) for every $x_i, x_j \in X$, and an edge (y_i, y_j) for every $y_i, y_j \in Y$.

An example of the resulting network, G, is illustrated in Figure S3. Now, consider the instance $(G, v_e, b, c, \hat{A}, \hat{R}, d)$ of the problem of Local Hiding, where G is the network we just constructed, v_e is the evader in G, $b = k$ (where k is the

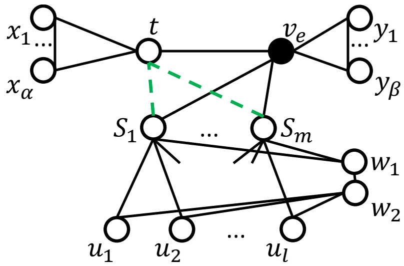

Figure S3: An illustration of the network used in the NP-completeness proof given the betweenness centrality.

parameter of the 3-Set Cover problem), c is the betweenness centrality measure, $d = 1$, $\hat{A} = \{(t, S_i) : S_i \in S\}$, and $\hat{R} = \emptyset$.

From the definition of the problem, one can see that the only edges that can be added to the graph are those between t and the members of S, meaning that $A^* \subseteq \hat{A}$, where $\hat{A} = \{(t, S_1), \ldots, (t, S_m)\}$. 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 Local Hiding corresponds to a solution to the 3-Set Cover problem.

First, we will show that for every node $v \in V \setminus \{t, v_e\}$ and every $A^* \subseteq \hat{A}$ we have $c(G', v) < c(G', t)$, where $G' = (V, E \cup A^*)$. To this end, let $B(v)$ denote the sum of percentages of shortest paths controlled by v between pairs of other nodes, *i.e.*, $B(v) = \sum_{w,w' \in V \setminus \{v\}} \frac{|\{p \in \Pi(w,w') : v \in p\}|}{|\Pi(w,w')|}$. Note that $B(v) = \frac{(n-1)(n-2)}{2} c(G', v)$ Next, we will show that the following holds:

$$\forall_{v \in V \setminus \{t, v_e\}} \forall_{A^* \subseteq \hat{A}} B(v) < B(t).$$

Since t controls all shortest paths between the nodes in X and those in $\{v_e, w_1, w_2\} \cup Y \cup S \cup U$, we have:

$$B(t) \geq \alpha(\beta+m+l+3) \geq m^4 l^3 (m+l+2) + m^2 l (m+l+2)^2$$

Moreover, since $\alpha = m^2 l(m + l + 2)$, $\beta = m^2 l(k + l + 2)$, and $k < m$, then: $\alpha + \beta < 2m^2 l(m + l + 2)$.

For nodes other than t we have:

- $B(x_i) = B(y_i) = 0 < B(t)$, since the nodes in $X \cup Y$ do not control any shortest paths.
- $B(w_1) \leq (\alpha+\beta+m+2) + \frac{m(m-1)}{2} + ml \leq 2m^2 l(m+l+2) + m^2 + m + ml < (2m^2 l + m)(m+l+2) < B(t)$, because w_1 controls some shortest paths between w_2 and nodes in $\{t, v_e\} \cup X \cup Y \cup S$ (there are $\alpha + \beta + m + 2$ such pairs), some shortest paths between pairs of nodes in S (there are at most $\frac{m(m-1)}{2}$ such pairs), and some shortest paths between nodes in U and nodes in S (there are at most ml such pairs).
- $B(w_2) \leq \frac{l(l-1)}{2} + l + ml < \frac{l^2+l}{2} + ml < B(t)$, because w_2 controls some shortest paths between pairs of nodes in U (there are at most $\frac{l(l-1)}{2}$ such pairs), some shortest paths between nodes in U and w_1 (there are at most l such pairs), and some shortest paths between nodes in U and nodes in S (there are at most ml such pairs).
- $B(u_i) \leq (\alpha + \beta + m + 2) + \frac{m(m-1)}{2} < B(t)$, because u_i controls some shortest paths between w_2 and nodes in $\{t, v_e\} \cup X \cup Y \cup S$ (there are $\alpha + \beta + m + 2$ such pairs), and some shortest paths between pairs of nodes in S (there are at most $\frac{m(m-1)}{2}$ such pairs).
- $B(S_i) \leq 3(\alpha+\beta+l+m+2)+l+2(\alpha+\beta+2) \leq 5(\alpha+\beta+l+m+2) \leq (10m^2 l+5)(m+l+2) < B(t)$, because S_i controls some shortest paths between the nodes in U that are connected to S_i and the nodes in $\{t, v_e\} \cup X \cup Y \cup S \cup U$ (there are at most $3(\alpha + \beta + l + m + 2)$ such pairs), some shortest paths between w_1 and the nodes in U (there are at most l such pairs), and some of the shortest paths between nodes in $\{w_1, w_2\}$ and nodes in $\{t, v_e\} \cup X \cup Y$ (there are at most $2(\alpha + \beta + 2)$ such pairs).

Therefore, either t or v_e has the highest betweenness centrality. Hence, $A^* \subseteq \hat{A}$ is a solution to the problem of Local Hiding if and only if $B(t) > B(v_e)$. We now compute the values of $B(t)$ and $B(v_e)$. We have that:

$$B(t) = \alpha(\beta + m + l + 3) + \sum_{\substack{S_i, S_j \in S: \\ (t,S_i) \in E \wedge (t,S_j) \in E}} \frac{1}{|N(S_i, S_j)|} + \sum_{S_i \in N(t)} \sum_{u_j \in U \setminus N(S_i)} \frac{|N(t, u_j)|}{|N(t, u_j)| + |N(v_e, u_j)| + 1}$$

as t controls all shortest paths between every pair (x_i, v) where $x_i \in X$ and $v \in V \setminus (X \cup \{t\})$ (there are $\alpha(\beta + m + l + 3)$ such pairs), one shortest path between each pair of nodes in $N(t) \cap S$, and the shortest paths between every pair (v, w) where $v \in N(t) \cap S$ and $w \in U : N(t) \cap N(w) \neq \emptyset$ (other paths run through v_e and nodes in S, or through w_1 and w_2). On the other hand, we have that:

$$B(v_e) = \beta(\alpha + m + l + 3) + \sum_{S_i, S_j \in S} \frac{1}{|N(S_i, S_j)|} + \sum_{S_i \notin N(t)} (\alpha + 1) + \sum_{u_i \in U : N(t, u_i) = \emptyset} (\alpha + 1) + \sum_{S_i \in S} \sum_{u_j \in U \setminus N(S_i)} \frac{|N(v_e, u_j)|}{|N(t, u_j)| + |N(v_e, u_j)| + 1}$$

as v_e controls all shortest paths between nodes in Y and all other nodes (there are $\beta(\alpha + m + l + 3)$ such pairs), one shortest path between each pair of nodes in S, paths between nodes in S and nodes in U, and all shortest paths between $\{t\} \cup X$ and nodes $\{S_i \in S : S_i \notin N(t)\} \cup \{u_i \in U : N(t, u_i) = \emptyset\}$. Thus, we have:

$$B(v_e) - B(t) = (\beta - \alpha)(m+l+3) + \sum_{\substack{S_i, S_j \in S: \\ (t,S_i) \notin E \vee (t,S_j) \notin E}} \frac{1}{|N(S_i, S_j)|} + \Delta SU + \sum_{S_i \notin N(t)} (\alpha+1) + \sum_{u_i \in U : N(t, u_i) = \emptyset} (\alpha+1)$$

where $0 < \Delta SU \leq ml$.

Note that $B(v_e)$ decreases with $|A^*|$ and also decreases with $\left|\{u_i \in U : \exists_{S_j \in N(t)} u_i \in S_j\}\right|$. Next, we prove that:

(a). If $|A^*| = k$ and for every $u_i \in U$ there exists $S_j \in N(t)$ such that $u_i \in S_j$, then $B(v_e) < B(t)$;

(b). If $|A^*| = k$ and there exists $u_i \in U$ such that for every $S_j \in N(t)$ we have $u_i \notin S_j$, then $B(v_e) > B(t)$.

Regarding point (a), we have:

$$B(v_e)-B(t) = (\beta-\alpha)(m+l+3) + (m-k)(\alpha+1) + \sum_{\substack{S_i,S_j\in S:\\(t,S_i)\notin E\vee(t,S_j)\notin E}} \frac{1}{|N(S_i,S_j)|} + \Delta SU$$

Now since $|\{S_i, S_j \in S : (t,S_i) \notin E \vee (t,S_j) \notin E\}| = \frac{m(m-1)-k(k-1)}{2} = \frac{(m-k)(m+k-1)}{2}$, and $|N(S_i,S_j)| \geq 2$, then we have:

$$B(v_e)-B(t) \leq (\beta-\alpha)(m+l+3) + (m-k)(\alpha+1+\frac{\Delta SU}{m-k}+\frac{m+k-1}{4}).$$

By substituting the values of α and β, and observing that $\Delta SU < ml$ and $k < m$, we get:

$$B(v_e)-B(t) < m^2l(k-m)(m+l+3) + (m-k)(m^2l(m+l+2)+1+ml+2m-1),$$

which gives us:

$$B(v_e)-B(t) < (k-m)m^2l + (m-k)(ml+2m) = (k-m)m(ml-l-2) < 0.$$

Hence, if $|A^*| = k$ and for every $u_i \in U$ there exists $S_j \in N(t)$ such that $u_i \in S_j$, then $B(v_e) < B(t)$.

Regarding point (b), since there exists $u_i \in U$ such that for every $S_j \in N(t)$ we have $u_i \notin S_j$, then:

$$B(v_e)-B(t) \geq (\beta-\alpha)(m+l+3)+(m-k)(\alpha+1)+(\alpha+1) + \sum_{\substack{S_i,S_j\in S:\\(t,S_i)\notin E\vee(t,S_j)\notin E}} \frac{1}{|N(S_i,S_j)|} + \Delta SU.$$

Since $\sum_{\substack{S_i,S_j\in S:\\(t,S_i)\notin E\vee(t,S_j)\notin E}} \frac{1}{|N(S_i,S_j)|} > 0$ and $\Delta SU > 0$, then we have:

$$B(v_e)-B(t) > (\beta-\alpha)(m+l+3) + (m-k+1)(\alpha+1).$$

By substituting the values of α and β we get:

$$B(v_e)-B(t) > m^2l(k-m)(m+l+3)+(m-k+1)(m^2l(m+l+2)+1)$$

which gives us:

$$B(v_e)-B(t) > m^2l(k-m)+m^2l(m+l+2) = m^2l(k+l+2) > 0$$

Hence, if $|A^*| = k$ and there exists $u_i \in U$ such that for every $S_j \in N(t)$ we have $u_i \notin S_j$, then $B(v_e) > B(t)$.

Thus, the solution to the problem of Local Hiding corresponds to a solution to the given instance of the 3-Set Cover problem, which concludes the proof. □

S2 Data sets

We considered three terrorist networks, namely:

- *WTC*: This is the network of terrorists responsible for the WTC 9/11 attack [Krebs, 2002];
- *Bali*: The network of terrorists behind the 2002 Bali attack [Hayes, 2006];
- *Madrid*: This is the network of terrorists responsible for the 2004 Madrid train bombing [Hayes, 2006].

We also considered three standard models for random networks (for each model, we generate 50 networks consisting of 30 nodes):

- *Scale-free* networks, generated using the Barabasi-Albert model [Barabási and Albert, 1999]. We set the number of links added with each node to be 3;
- *Small-world* networks, generated using the Watts-Strogatz model [Watts and Strogatz, 1998]. In our experiments, we set the average degree to be 10;
- *Random graphs* generated using the Erdos-Renyi model [Erdős and Rényi, 1959]. In our experiments, the expected average degree is set to be 10.

Finally, we considered anonymized fragments of three social networks, namely Facebook, Twitter and Google+. These fragments are taken from SNAP—the Stanford Network Analysis Platform [Leskovec and Mcauley, 2012].

References

[Barabási and Albert, 1999] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[Erdős and Rényi, 1959] Paul Erdős and Alfréd Rényi. On random graphs i. *Publ. Math. Debrecen*, 6:290–297, 1959.

[Hayes, 2006] Brian Hayes. Connecting the dots can the tools of graph theory and social-network studies unravel the next big plot? *American Scientist*, 94(5):400–404, 2006.

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

[Leskovec and Mcauley, 2012] Jure Leskovec and Julian J Mcauley. Learning to discover social circles in ego networks. In *Advances in neural information processing systems*, pages 539–547, 2012.

[Watts and Strogatz, 1998] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *nature*, 393(6684):440–442, 1998.

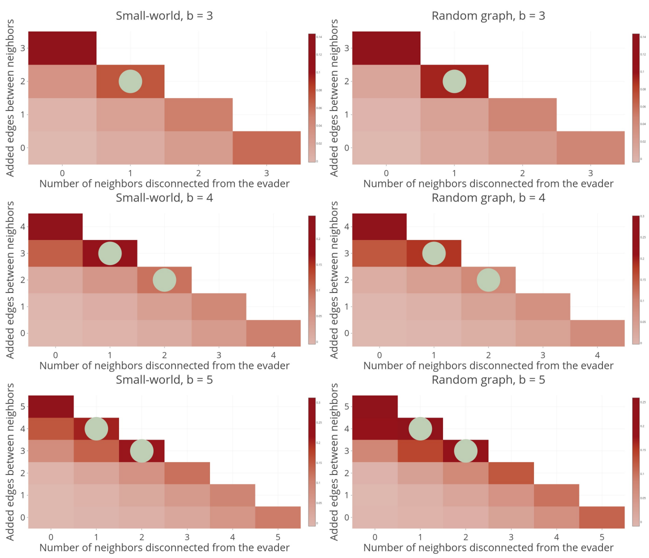

Figure S4: Given Small-world and Random-graph networks, and given budgets 3, 4, and 5, the figure depicts the evader's average payoff for different strategies. Specifically, the x-axis represents the number of network modifications whereby the evader is disconnected from one of this neighbours; the y-axis represents the number of network modifications whereby an edge is added between two of the evader's neighbours; the color intensity in every cell represents the evader's average payoff, taken over all the strategies in which the network modifications match the corresponding x and y coordinates. A circle indicates the fact that at least one version of ROAM lies in the corresponding cell.

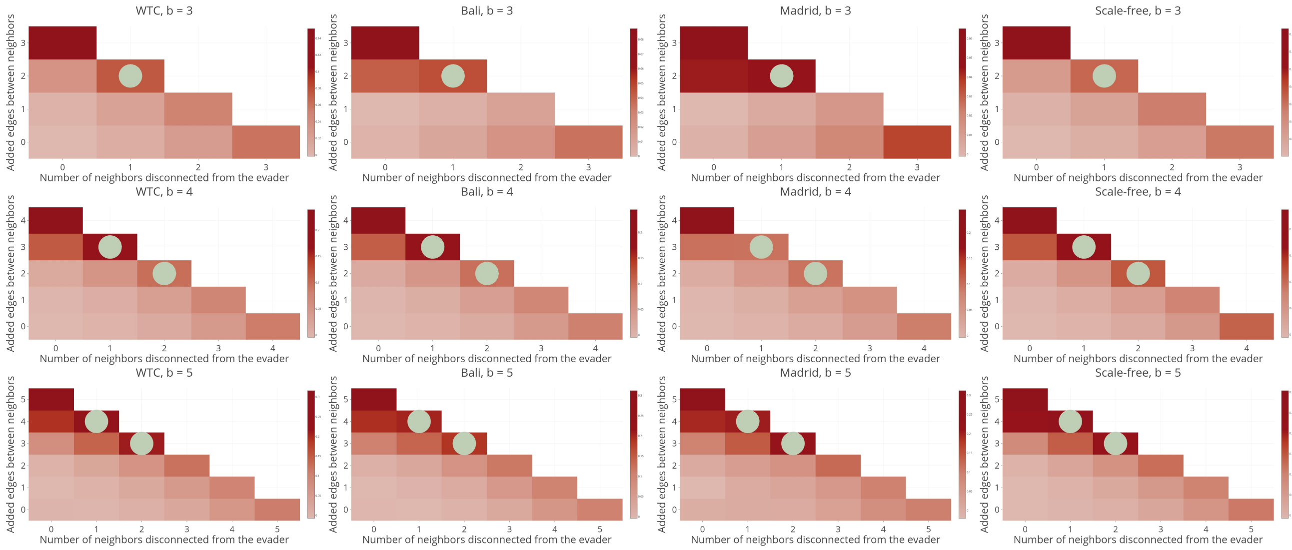

Figure S5: Given three terrorist networks (WTC, Bali, and Madrid) and Scale-free networks, and given budgets 3, 4, and 5, the figure depicts the evader's average payoff for different strategies. Specifically, the x-axis represents the number of network modifications whereby the evader is disconnected from one of this neighbours; the y-axis represents the number of network modifications whereby an edge is added between two of the evader's neighbours; the color intensity in every cell represents the evader's average payoff, taken over all the strategies in which the network modifications match the corresponding x and y coordinates. A circle indicates the fact that at least one version of ROAM lies in the corresponding cell.

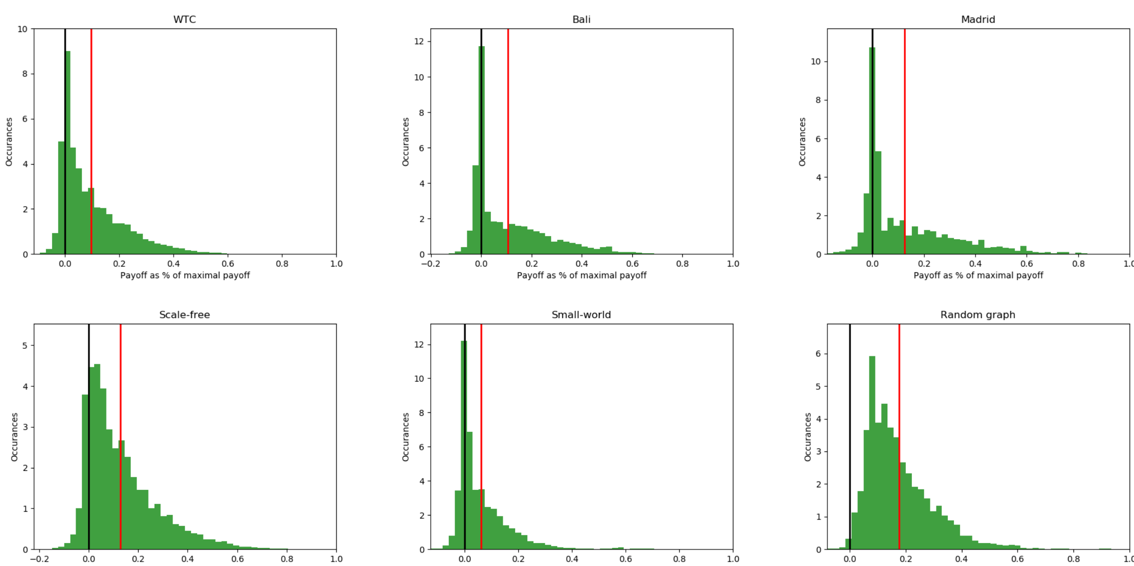

Figure S6: The distributions of the evader's payoffs for budget 3. Values are provided for evader type $\phi = 0.5$ and averaged over the seeker's equilibrium strategies. For each network, the red and black lines denote the average payoff and 0, respectively.