The Economics of Covert Community Detection and Hiding

Shishir Nagaraja
Computer Laboratory
15 JJ Thomson Avenue, Cambridge CB3 0FD, UK
shishir.nagaraja@cl.cam.ac.uk

April 26, 2008

Abstract

We present a model of surveillance based on the detection of community structure in social networks. We examine the extent of network topology information an adversary is required to gather in order to obtain high quality intelligence about community membership. We show that selective surveillance strategies can improve the adversary's resource efficiency. However, the use of counter-surveillance defence strategies can significantly reduce the adversary's capability. We analyze two adversary models drawn from contemporary computer security literature, and explore the dynamics of community detection and hiding in these settings. Our results show that in the absence of counter-surveillance moves, placing a mere 8% of the network under surveillance can uncover the community membership of as much as 50% of the network. Uncovering all community information with targeted selection requires half the surveillance budget where parties use anonymous channels to communicate. Finally, the most determined covert community can escape detection by adopting decentralized counter-surveillance techniques even while facing an adversary with full topology knowledge - by investing in a small counter-surveillance budget, a rebel group can induce a steep increase in the false negative ratio.

1 Introduction

The use of traffic analysis to trace communication and the parallel development of anonymous communication methods have both seen rapid strides in the past decade or so. As more and more governments move towards electronic censorship, resistance against traffic analysis becomes an important goal in information security.

While governments are engaged in violating user privacy for political reasons, enterprises do the same for economic reasons. However, the collaboration between the two is even more dangerous. In 2006, the EFF [Fou] asserted in a class action lawsuit that AT&T's Daytona system, a huge storehouse of telephone call records was the basis for NSA's domestic surveillance effort to mine records without a warrant.

Aggregation of personal data in electronic systems, especially with unrestricted access has long been considered as a significant risk to user privacy by security experts. Modern email service providers support a large number of users who are attracted by promises of free service and attractive storage capacity allowances. This results in the aggregation of a large amount of social network information within the administrative power of a very small number of people running the service. What sort of privacy risks does this pose and what does it cost users to defend themselves against them?

In the electronic world, conflicts of interest between parties, often turn on connectivity. In order for an adversary to remove a party opposing her interests in a network, she must use traffic analysis to trace communications followed by the arrest or removal of groups of individuals who appear to be significant nodes. Covert groups are aware of this and take steps to prevent their networks from being traced. Effective traffic analysis, in addition to placing network components under surveillance, involves the detection of covert groups. Covert groups such as secret societies do not exist in isolation, but are often connected to the rest of society through social links. An analysis of surveillance must therefore also measure the efficiency of accurately detecting complex structural information such as the presence of communities and their membership. An attacker in cahoots with the service provider locates a community of people suspected of 'undesirable' political activism. Police agencies in repressive regimes could thus obtain high-quality intelligence by combining traditional traffic analysis methods with advanced community detection algorithms developed by the complex networks community.

Our model of surveillance is inspired by the range of legal wiretapping legislations that have been recently passed to enable police agencies to collect ever larger amounts of traffic data. Our objective in this work is to analyze the impact on privacy by network externalities involved in computer insecurity. We take the view that adversary's capability to detection community boundaries and membership information constitutes a significant critical piece of information that is closer to operational intelligence than the mere discovery of nodes and inter-node relations.

Traffic data collected on a network components necessarily impacts the privacy of other components. This is a basic network externality. Several questions arise from this: How does surveillance impact the privacy of community membership information? Do selective surveillance strategies perform than the installation of random wiretaps? How does the extent of network topology knowledge affect the adversary's efficiency in performing surveillance? To what extent does the extent of community membership uncovered depend on the surveillance budget? What sort of counter-surveillance defences can communities use and what are the costs and resources involved?

Recent advances in the theory of networks have provided us with the mathematical and computational tools to understand such phenomena better. This paper starts to explore the tactical and strategic options open to combatants in such conflicts. What strategies can one adopt, when building a network, to provide good trade-offs between efficiency and resilience to detection? The problem of detecting hidden communities in large-scale networks has a long history in traffic analysis from secret societies evading detection to spy agencies data-mining call record information looking for criminals.

We answer these questions by performing experiments on a harvested social network obtained from email communication at a medium sized University. Our approach is to mimic the actions of a topology information seeking adversary, from data collection through to the analysis. Past work [DW06] carried out an illuminating study on the surveillance costs of monitoring mailing lists, and, discovering people and relations. They considered the static case, where the network does not adopt counter-surveillance to repel the adversary. We extend this work in two ways: First, we analyze the impact of surveillance on uncovering communities and community membership information; and second, we extend the static case to the dynamic where hard core communities carry out counter-surveillance manoeuvres to escape detection even under strong adversaries.

2 Community detection

2.1 Background

The problem of splitting a network into a number of sub-communities in some sort of a sensible way is not a new one. The first algorithm for graph partitioning, was to our knowledge, proposed by Kernighan and Lin [KL70]. Given an initial randomly selected partition, their greedy algorithm tries to minimize the size of the edge cut-set. A number of combinatorial optimization based graph partitioning techniques were directly based on or inspired by this algorithm. Another class of algorithms are based on geometric distance, in that nodes are placed on a grid before partitioning them using geometric graph metrics such as finding an appropriate bisecting plane leading to two approximately equal sized subsets. A detailed survey of these methods can be found in [Els05]. Many of these techniques have been researched in the computer science community for applications in parallel computing, in the optimal distribution of processes to CPUs, and, by the VLSI community in placement and routing design decisions [AK95]. A number of highly efficient techniques based on the spectral properties of graphs were invented [Fie73] with further contributions from Pothen et.al. [PSL90]. The recursive spectral partitioning method uses the eigen-vector corresponding to the second lowest eigen-value of the *Laplacian matrix* of the graph, to sort the vertices of the graph into two subsets. This method is then applied recursively to generate more communities.

In recent years, the physics community has also been active in solving the problem, with a host of techniques based on the spectral properties of graphs, for a recent survey see Danon et.al. [DDDA05]. Kirkpatrick et.al. [KGV83] suggest another class of algorithms based on statistical mechanics, which unlike greedy methods, is not easily trapped in a local minima. Guimera and Amaral [GA05] propose further algorithms in this class.

Social scientists, who have been working on the problem for much longer, have developed a parallel suite of algorithms such as, block-modelling based on clustering [WF94], centrality based methods [WF94, New03] where high centrality components highlight community borders, and hierarchical clustering [WF94] based on classifying nodes based on similarity indices.

2.2 Problem formulation

The attacker is a passive adversary interested in collecting intelligence from traffic analysis of messages exchanged between nodes of the network. Over a period of time the attacker builds a graph where vertices correspond to the nodes of the network and edges correspond to an exchange of messages between them. He then applies the community detection algorithm in order to uncover the community structure of the graph. The adversary's capability and surveillance budget determine his ability to discover all the communities as well as correctly associate each node with its community in the target network.

We first study the efficiency of surveillance under different adversary models. In a real life situation covert communities (defenders) are aware that of being watched and respond appropriately. Their goal is to escape detection by blending in with the rest of the network. Naturally, we extend our analysis to measure the efficiency of such defensive manoeuvres.

We now give a comprehensive definition of the problem we are trying to solve. The graph $G = (V, E)$ of vertices $V = \{v_i | i = 1, \ldots, n\}$, some of which are connected by edges in $E = \{e_{i,j} |$ there is an edge between v_i and $v_j\}$. Given a subgraph(V', E'), where $V' \subset V$, then $E' \subset E$ contains all the edges that join the vertices in V'.

The problem of community detection is to find a graph partition (communities) (V_1, V_2)

of V, such that, $V_1 \cup V_2 = V$ and $V_1 \cap V_2 = \emptyset$. Further partitions are uncovered by recursively applying the solution to each of the resulting partitions.

So, what criteria must be applied to achieve *good* community detection? Techniques based in sociology concentrate on carrying out the above partitioning without constraints on the sizes of V_1 and V_2, whilst trying to find a minimal edge cut-set using an appropriate network centrality criteria, such as edge betweenness centrality [Fre78]. Min-cuts work well where the community divisions are sharp and well defined by high betweenness edges. Such techniques are trivially bypassed by placing redundancy around high betweenness edges. Hence, min-cuts as a community detection tool are too narrow in scope to discover embedded covert groups.

A much more promising candidate is based on the idea of *modularity*, proposed by Newman and Girvan [NG04] and further developed in [New06]. Modularity is defined as the difference between the *expected* number of edges in a group and the actual number of edges within a group. Instead of using min-cuts or high-centrality components to differentiate between groups, optimizing modularity, is the criteria chosen for uncovering topological structures such as clubs. Quoting Newman [New06], "A good division of a network into communities is not merely one in which there are few edges between communities; it is one in which there are fewer than expected edges between communities". The intuition behind this algorithm is clearly much broader, compared to edge min-cut based methods.

This method is known to be highly accurate in detecting community structure in a variety of social and biological networks [New06]. In addition, the intuition behind the algorithm in locating communities is much more broader than looking for edge min-cuts. For these reasons, we shall use this algorithm to explore the dynamics between community detection and community hiding techniques.

2.3 Spectral partitioning based on Modularity

We now describe Newman's modularity optimizing graph partitioning technique. The intuition behind modularity detection is that real world networks, such as social networks, often tend to be composed of modules or clubs. Informally, a module is a subgraph whose nodes are more likely to be connected to one another than to the nodes outside the subgraph. A variety of approaches have been used in existing network theory literature[NG04, CNM04, GK06].

In this paper our focus will be on the Newman modularity based community detection method [New06]. What makes it especially interesting in the context of surveillance, is its performance and accuracy of community detection even under significant random errors in the input topology - a linear increase in random noise in the network topology results in a linear increase in number of false positives, as opposed to an exponential increase false positive rate in the case of min-cut based methods.

Modularity reflects the extent, relative to a random network, to which edges are formed within communities rather than across them. Newman uses a modularity score, to assess the quality of any assignment of nodes to the same community, hence, identifying community membership becomes a modularity maximization problem.

Let A_{ij} be the adjacency matrix of graph $G(V, E)$. $m = |E|$ is the total number of edges in the graph. A_{ij} is set to 1 if an edge exists between nodes i and j, and 0 otherwise. Degree of node i is given by d_i. P is the matrix containing the expected number of edges between any pair of nodes, if edges are placed randomly. We choose P_{ij} as follows:

$$P_{ij} = \frac{d_i d_j}{2m} \tag{1}$$

The modularity matrix B, is given by $B_{ij} = A_{ij} - P_{ij}$. Each element of B is the difference between the number of edges e_{ij} between i and j and the *expected* number of edges between them. The vector of eigen-values of the modularity matrix is known as the eigen spectrum of the matrix. This eigen spectrum $(\lambda_n \ldots \lambda_1)$ of matrix B has a fundamental relationship with the extent of club like nature of the given network.

The modularity (Q) of the network is the sum B_{ij} over all pairs of vertices (i, j) falling in the same group: $Q = \frac{1}{4m} \sum_{ij} [A_{ij} - P_{ij}] s_i s_j$, where $s \in \{0, 1\}$ represents the membership bit-vector. To find a good partition, Q must be maximized, achieved by replacing s by the leading eigenvector corresponding to the leading eigenvalue of B. For this we calculate the normalized eigenvectors and eigenvalues of B. Matrix B plays the similar role in graph partitioning as the Laplacian matrix does in Laplacian spectral graph partitioning [PSL90, Fie73]. The second eigenvector is chosen as the leading eigenvector, taken to represent the community membership s, to maximize Q, with all vertices with $s \geq 0$ forming one partition and $s < 0$ forming the second partition. Where there is no good partition, all eigenvalues for B will be negative, in which case $(1, 1, 1....)$ is the leading eigenvector and all the vertices will be grouped together.

3 Network model and dataset

To understand the dynamics of community hiding and detection with the modularity spectral partitioning method, we need to first define our network model that serves as the base for our analysis framework.

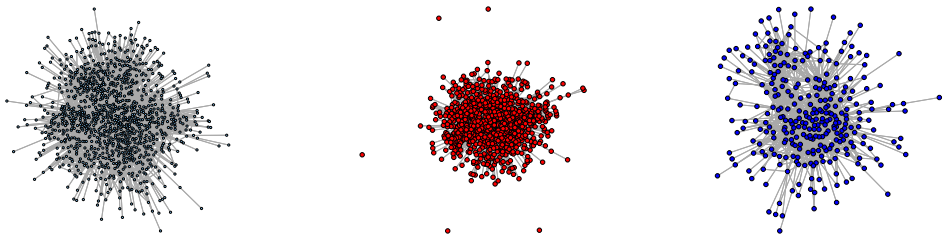
We consider social networks comprising of people and their relations. The social network is represented by a graph $G(V, E)$, where people are represented as nodes, while relationships between people are represented as edges. Sets V, E are the set of all nodes and edges, respectively. Each edge is associated with an integer weight which is an indicator of the quantity of information exchanged between the two end-points.

In analyzing community structure our main goal is to study the detection of communities under conditions of incomplete information and dynamics of community hiding. This allows us to focus our analysis on the shared aspects of security at the boundaries of communities rather than pairwise security or network-level security.

3.1 Email communication network

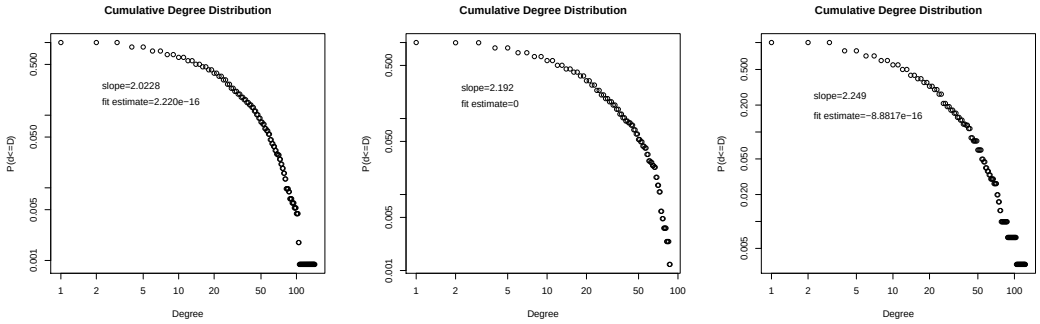
Our network dataset comprises of a social network harvested from email exchanges within a University of 1700 researchers, graduate students and staff. Each email address was mapped to a person. We discarded all email messages where either the sender or the receiver email address was not a University email address. This means we have left out relationships where two persons at the University might be connected via an outsider, and this could impact our results. We disregarded unidirectional email messages which removed bulk email messages as well as any spam that was not replied to. We added an edge between every two nodes that had sent at least one message in each direction. The weight of the edge was set as the sum total of messages exchanged between two nodes.

Next, we extracted the largest connected component or giant component consisting of 1133 people with 10903 edges, which will be the focus of our attention, see fig 1. Initial community detection reveals two partitions, G_M, with 831 nodes and 6807 edges which we shall consider our main network and G_C a covert group of 302 nodes and 2574 edges. These correspond to the two main domains within the dataset.



(a) Giant connected component N=1133, $|E|$=10903 (b) Partition G_M: N_M=831, $|E|$=6807 (c) Partition G_C: N_C=302, $|E|$=2574

Figure 1: Email communication network dataset



(a) Giant connected component N=1133, $|E|$=10903 (b) Partition G_M: N_M=831, $|E|$=6807 (c) Partition G_C: N_C=302, $|E|$=2574

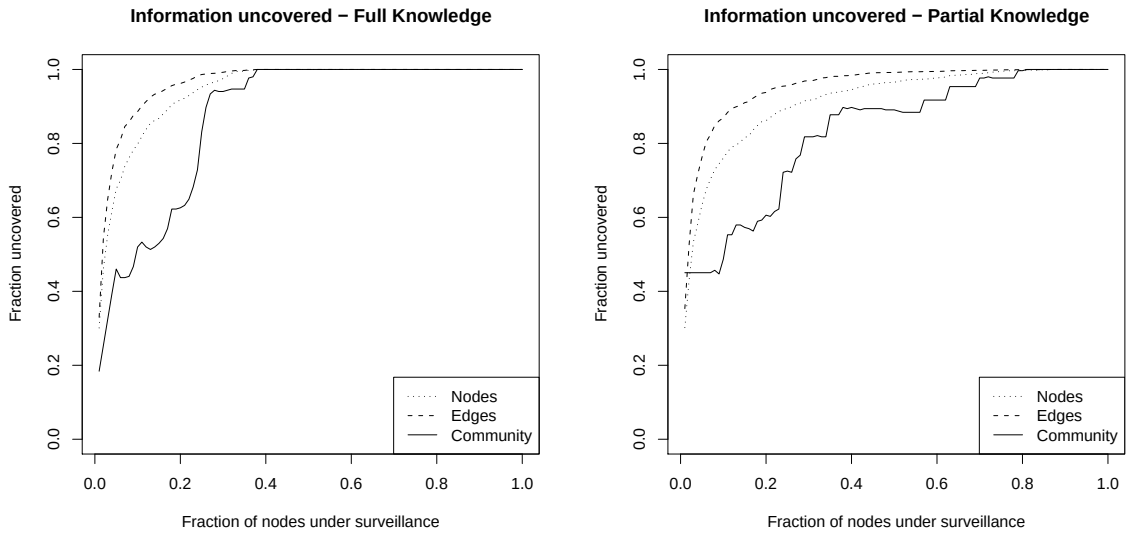
Figure 2: Degree distributions

3.2 Network characteristics

Fig- 2 shows the statistical characteristics of our networks. We used the visual representation in this diagram to choose the starting value of our observed power-law distribution X. We chose $x_{min} = 5$, and calculated the slope α of the distribution using the approximation formula of [CSN07], based on the method of maximum likelihood estimation [CBN94]: $\alpha \approxeq 1 + N[\sum_{i=1}^{N} ln\frac{x_i}{x_{min}-\frac{1}{2}}]^{-1}$

The giant connected component comprising both communities seems (by visual inspection of fig- 2) to have a power-law degree distribution with a slope of $\alpha = 2.0228$. The individual communities also seem to be scale-free networks [AB02] with $\alpha_M = 2.192$ and $\alpha_C = 2.249$.

Having determined the scaling parameter α of the seemingly power-distribution, we must now determine whether our observed data is a good fit for a power-law distribution at all. For this we first generate synthetic data for a power-law distribution using the parameter α we have determined and then compare the synthetic and observed distributions using the Kolmogorov-Smirnov test [PTVF92]. The difference between the observed and generated distributions turns out to be very close to zero in all the above three graphs fig- 2, indicating that a power-law distribution is indeed a good fit for the degree-distributions observed in our dataset.



(a) (b)

Figure 3: Effects of surveillance on community detection efficiency

4 Effects of surveillance on community detection accuracy

In our model, the adversary's goal is to accurately determine the membership of each community in the network. Our first experiment attempts to measure the efficiency of community detection by different surveillance strategies. The adversary is limited by a surveillance budget, which limits his ability of gathering topology information. This in turn affects the success of surveillance goals, and we wish to measure how the adversary's success varies as the fraction of the network under direct surveillance increases from partial direct surveillance to full direct surveillance .

Although there are several communities and sub-communities in our dataset, we specifically concentrate on the two large communities while ignoring the extent of surveillance success with respect to any sub-communities involved.

The first strategy involves the adversary with complete topology knowledge. The adversary starts to first put those nodes under surveillance, that are the most central within the network, progressively working his way through this list in the descending order of significance. There are several centrality scores that could be used [WF94]. The most appropriate from the perspective of traffic routing is the betweenness centrality measure devised by Freeman [Fre78]. It is designed to capture how crucial a node or an edge is to the process of routing data along shortest paths. The betweenness centrality C_b^v of a node v may be defined as the number of all pairs shortest paths that pass through v:

$$C_B^v = \sum_{x \in V} \sum_{y \neq x \in V} \frac{\sigma_{xy}(v)}{\sigma_{xy}}$$

Where σ_{xy} is the number of shortest paths between nodes x and y.

The effectiveness of surveillance using this strategy is shown in fig- 3.a. After placing 8% of the nodes under direct surveillance under the centrality strategy, the adversary is aware of 76% of the nodes and 85% of the edges within the network. This partially confirms one of the results of an earlier study by Danezis and Wittneben [DW06] in the

same adversary model. However, the community membership of only 50% of the nodes was correctly identified. The adversary then makes further progress: by selectively placing 28% of the nodes under surveillance, he can correctly identify the membership of 95% of the nodes, whilst uncovering 96% of nodes and almost 99% of edges. The adversary then needs to put 38% of the network under direct surveillance to obtain 100% information about all nodes and edges in the network. The success of correct community identification is also obviously at 100% of the nodes, at this point.

A much more interesting adversary threat model (inspired by the work of [DW06]) is where the adversary has only partial information about the network's topology, in choosing surveillance targets. Perhaps, as obtained by observing traffic on anonymous communication networks [PH00, DDM03, DMS04, Nag07]. Here, the adversary can only observe the aggregate total of traffic passing between a given node and the rest of the network. He has no information about whom this traffic is destined for. The property of *communication unlinkability* offered by current anonymous communication infrastructure only hides the communication endpoints, not the existence of communication. We now measure the adversary's accuracy in discovering community membership under the case of partial information. The adversary can observe the sum total of traffic going in or out of each node, represented as an edge weight in our model. He then lists the nodes in descending order of traffic volumes and places an increasing fraction of nodes under surveillance in each successive time period.

Fig- 3.b shows the fruits of the adversary's efforts in the second threat model. The lack of full topology knowledge particularly dents the adversary's ability to effectively spy on the network: When 8% of the nodes are spied upon, the community membership of 45% of the nodes is correctly identified whilst uncovering 73% of nodes and 84% of edges. However, while 90% of the nodes and 96% of the edges are known to the adversary by putting 28% of the nodes under surveillance, the membership of only 58% is accurately known to him!

In terms of pure numbers the adversary is able to acquire a significant quantity of topology knowledge by putting 28% of the network under surveillance in both threat models. However, community membership knowledge is much more harder to gain. In the partial case, it requires the adversary to put almost 80% of the nodes under direct surveillance to enable him to accurately pin down the community memberships of 99% of the nodes. This is in stark contrast to the upper bound provided by the case of the fully knowledgeable adversary who only needs to place 37% of the nodes under surveillance to gain the same amount of membership information.

4.1 Discussion

Does this mean that the use of anonymous communications increases the work load of the adversary by almost 100%? Sadly, this is true only when the adversary needs to uncover the membership information of all nodes within the network. This could be case when she is interested in uncovering, say a terrorist network, where the risk of a terror attack may be considerable unless every node is removed from the network. However, a network of political dissidents might not be so lucky, detection followed by threats of arrest and torture of 50% of the organization could lead to its collapse. Our results show that in both the adversary models, putting 8% of nodes under direct surveillance compromises the community membership information of almost 50% of the nodes.

Since close to 80% of the population must be monitored to detect all the communities, it means that in the short run, government surveillance budgets are more likely to cause

harm to privacy than to uncover hardened terrorist cells.

Our results also confirm as well as contradict some of the results of an earlier study carried out by Danezis and Wittneben [DW06]. While we can confirm the trend in the number of edges and nodes exposed by indirect surveillance, the actual values are quite different. We attribute the significant deviation of adversarial success rates between our studies to the difference in the extent of skewness of the degree distributions of the two datasets. While they adopt a model where the community diameter is one, hence placing any one node in the community under surveillance exposes the entire community to the adversary, which is not the case in our dataset. This is due to structural differences between social networks based on mailing lists versus those on email networks. However, the adversary's efficiency in community membership assignment not only markedly lower but has a slower growth rate. This is not surprising given the relatively complex structural characteristics of the information the adversary is trying to uncover.

5 The effect of counter-surveillance defences

So far, we have measured and discussed the ability of an adversary to perform surveillance limited by topology knowledge and the surveillance budget. To this, we now add the variable of a counter-surveillance budget. Real world networks are dynamic in nature. The use of network topology to attack networks motivates covert communities to, in turn, find ways of hiding their presence by appropriately modifying their topology. We hence wish to measure the error tolerance of community detection techniques. We assume that the smaller of the two main communities in our dataset is a covert group. We then wish to study the strategy space of various random and selective graph topology manipulations that minimize the ability of the adversary in acquiring community membership knowledge for a given surveillance budget.

We adopt the following defence model: The members of the covert community adopt one of a number of defence strategies involving topological rewiring limited by a counter-surveillance budget. The adversary then runs several community detection algorithms to deduce group membership. We then determine her success by comparing results with the actual group membership. We consider strategies that require only local knowledge of topology, full topology knowledge and a hybrid between the two extremes.

Several topological manipulation options are open to the defending covert groups, such as *self-decapitation* where the covert group removes a selection of critical nodes from itself; *spurious intermediaries* where nodes are added as intermediaries between group members; and finally *edge-removal*, the removal of communication links between members of the community. However, such actions can substantially change the nature of information flow within the network hence be fairly disruptive to the dynamics of its operation. Therefore, we shall concentrate on the least disruptive form of topological manipulation by analyzing various strategies of *edge-addition* alone. The network defenders (nodes) add edges according to various strategies as a fraction of the total number of nodes in the covert group. We will then assess the efficiency of defence by calculating the corresponding change in the adversary's surveillance budget.

5.1 Measuring community hiding

We define the success of community detection with the following quantitative metric:

On running a community detection algorithm, we obtain n' communities $V' = (V_1 \cup \cdots \cup V_{n'}), n' \geq 1$.

The miss-ratio R, is defined as the fraction of nodes that are mis-detected as part of a community different from the one it belongs to.

$$R = \sum \left(\frac{|V_i| - |V_i \cap V_j'|}{|V_i|} \right)$$

where V_i maps to V_j' if $\forall j, |V_i \cup V_j'| - |V_i \cap V_j'|)$ is minimized.

Since we consider a single embedded covert community V_c in G, then the miss-ratio is defined as

$$R_c = \frac{|V_c| - |V_c \cap V_c'|}{|V_c|}$$

R_c, the percentage of covert nodes that are mis-detected as part of the main network which consists of the set of nodes $V - V_c$, is the false negative rate.

5.2 Counter-surveillance strategies

A naive defence for the covert group is to invest in communication links with the main network, where both end points are randomly chosen.
The RandomCovert(RNDc)-RandomMainstream(RNDm) edge addition strategy requires group members to toss a biased coin, so that a fraction of covert nodes select and engage a target node in the mainstream network. The number of edges added are limited by the cost of creating a social link.

A more sophisticated option involves targeted selection where defenders have full topology knowledge. This gives us two more strategies: High-centrality nodes in the covert network G_c connect with high-centrality nodes in the larger group G_m (HBc-HBm strategy); similarly high-degree nodes in the covert group invest in establishing contact with high-degree nodes in the larger group.

An hybrid strategy based on the knowledge of popular hub nodes in the larger network consists of a fraction of covert nodes volunteering to establish contact with well known (high betweenness centrality or high degree) nodes in the larger group. This constitute the RNDc-HBm and RNDc-HDm strategies. The reverse strategies of HBc-RNDm and HDc-RNDm are similarly defined.

5.3 Evaluating counter-surveillance defences

Let us first consider the case of the adversary with full topology knowledge and an unlimited surveillance budget, which will provide us an upper bound in our results.

The first strategy we analyzed for hiding G_C was the naive RNDc-RNDm strategy: edge addition with random end points selected from either group. Figure 4 shows the performance of this strategy of inducing errors, indicated by the blue line with an 'x' motif. RNDc-RNDm performs quite poorly. The gains in miss-ratio are a pathetic 10% when 600 additional edges (approximately 2 edges per covert node) are added, or 20% when 1000 additional edges (or 3.31 edges per covert node) are added. Given that the average degree of a covert node is 8.5 edges, 3.31 more edges is a steep 39% increase in edge resources per node for hiding 20% of the covert network, which is fairly unacceptable. This shows that the modularity community detection method we are using is quite resilient to random noise, which is important in the real-world scenario of noisy information (graph topology information) collection.

Having learnt that the naive strategy is of little use in hiding the covert group in a real world network, we proceeded to apply the next set of techniques, namely the purely

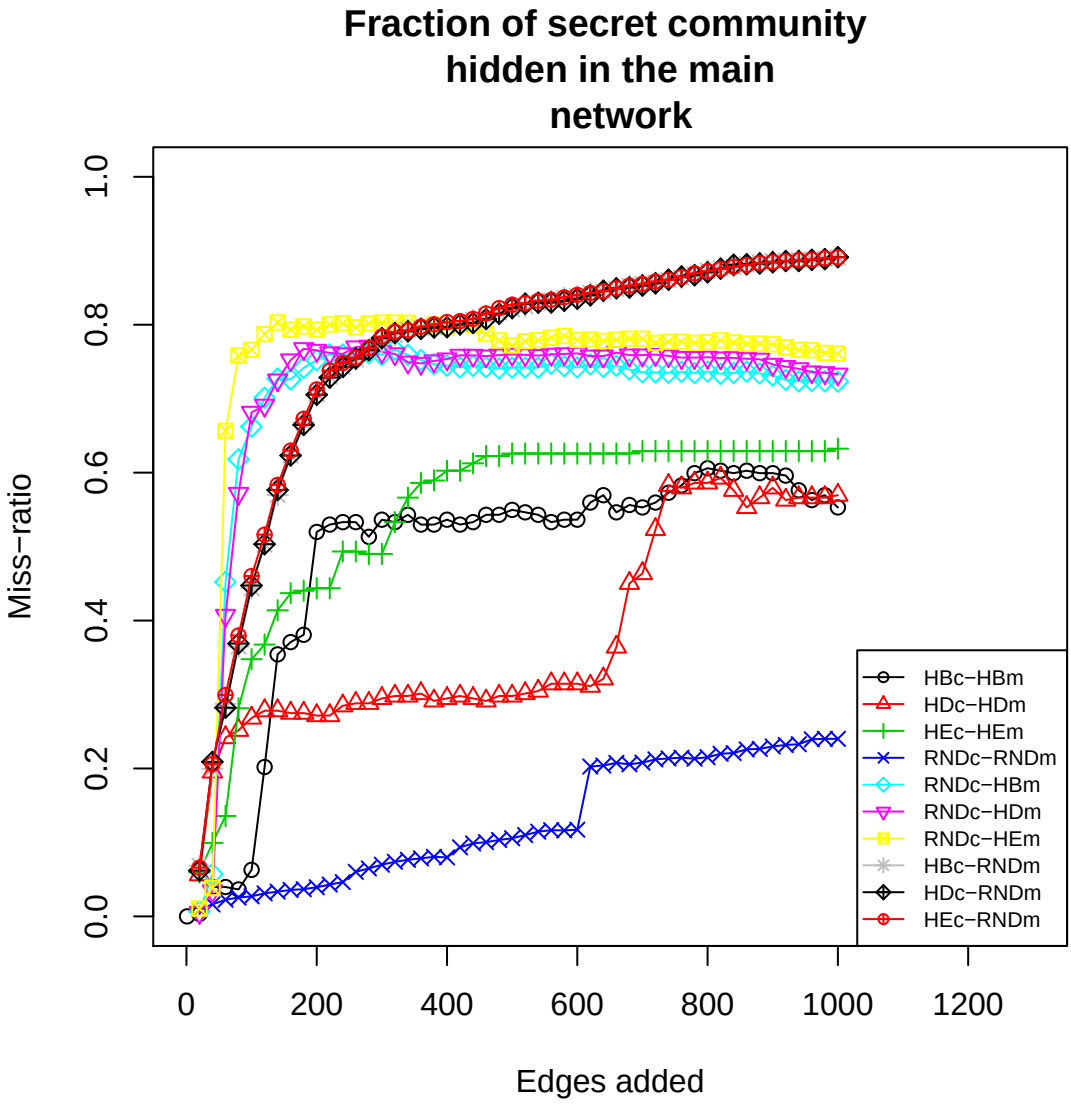
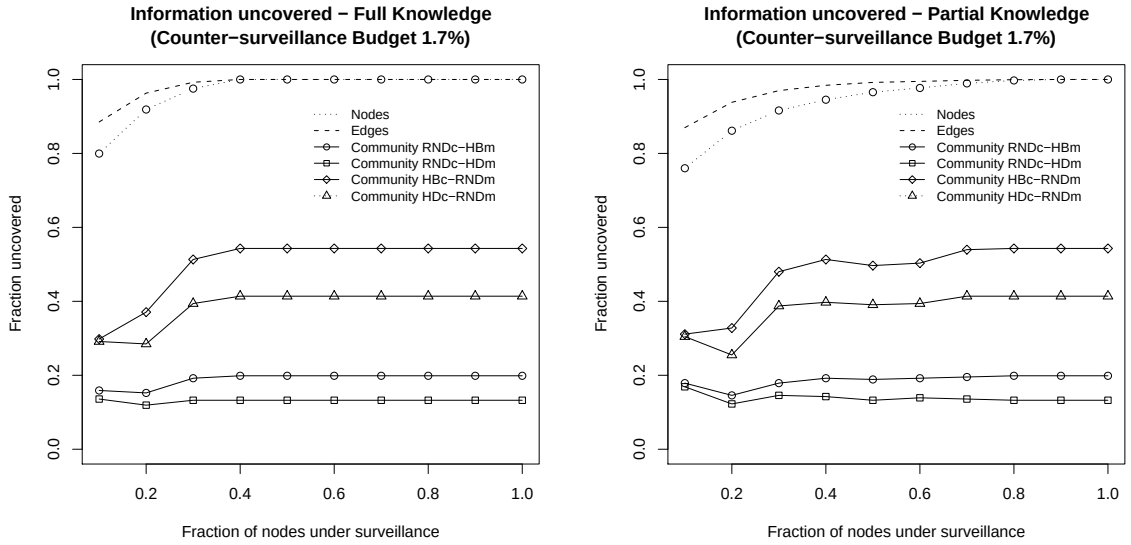


Figure 4: Fraction of hidden covert nodes by modularity spectral partitioning method



(a) Full knowledge (b)

Figure 5: Effects of counter-surveillance on adversarial community detection

centrality based strategies. The betweenness centrality strategy HBc-HBm (black line with circle motif) and HDc-HDm (red line with triangle motif) are average performers, with a peak miss-ratio of 50%, for an additional .66 edges per covert node (or 10% extra covert network edge resources). In addition, HDc-HDm also requires relatively larger amount of resources before being able to deliver a false negative rate above 30%.

Finally we look at the four hybrid strategies of edge addition, combining random node selection in one community with strategic selection in the other. Of the three strategies involving strategic selection in the main partition (requiring knowledge of popular nodes), RNDc-HBm (blue line with a diamond) and RNDc-HDm (pink line and triangle motif) perform equally well, with 70% of the covert nodes hidden from detection with only an additional 0.08 edges per node (or 1% of covert network edges).

The final three hybrid strategies involve local topology knowledge: HBc-RNDm indicated by the '*' motif and HDc-RNDm indicated by solid diamond. They work almost as well as the previous hybrid strategies on this network, however they require almost twice as many edges to offer the same level of protection. The two strategies have the same fault injection rate, which points to the fact that the covert network follows a scale-free regime where the degree of a node is as good an indicator of centrality as is betweenness centrality. Even more interestingly, these strategies offer a false negative rate of 93% of the covert group mis-detected as part of the larger community. This indicates that strategies with local knowledge can be more efficient at community hiding than strategies dependent on global topology information!

Our counter-surveillance defence tactics show that, if the defenders have full topology knowledge, then at a cost of 3 additional communication links per covert node, the modularity of the graph can be manipulated to the extent of raising false negative detection ratio upto 90%, in the adversary's membership calculations. In both cases, of partial and full topology knowledge, the defenders can drive up the false negative detection ratio to 80%, with a investment of only 0.01 edges per covert node or approximately 1% of existing covert network edge resources.

Figure 5 shows the efficiency of surveillance when counter-surveillance tactics are deployed by covert communities. Strikingly, while people and relations are discovered rapidly, the percentage of community membership information available to the attacker remains limited to 12% in the best performing counter-surveillance tactic. The RNDc-HDm strategy, where random members in the rebel group connect with hub nodes in the larger community, is attractive as a counter strategy in several ways. First, hub nodes with high degrees of connectivity are well known [WF94]; and second, since covert group members bearing responsibility for implementing the strategy do so in a decentralized manner and are easily replaceable if removed [NA06]. Further, we observe that the use of anonymous communication channels by the covert group does not change their privacy gains against community detection. This means that anonymous communication systems that promise unlinkability fall short of the required level of security, full unobservability is required to prevent the adversary's success. As current anonymous communication networks do not provide this property, the best defences lie in modifications at the fundamental level of network topology.

6 Previous work

Past work by Danezis and Wittneben [DW06] highlighted the privacy compromising network externalities involved in computer insecurity. If Alice's computer gets hacked into, then the adversary also obtains some information about Alice's neighbors in her social network. Their work considered the risk of privacy invasion by the discovery of people and relations, due to indirect surveillance.

We extend this work by determining the lower bound of surveillance required to uncover communities of larger diameter - complex subgraphs where spying on one member does not reveal information about all people and relations in the community. To this end, we consider three different surveillance models. The first, inspired by the above example, considers the global (passive) adversary who can observe all communication. The second model is inspired by current anonymous communication networks, where the adversary is able to observe only the traffic volumes of all users. The third model is of a local adversary who can monitor only a small set of people at a time.

While the adversary is interested in discovering as many communities as possible, community members themselves are keen to prevent the adversary from discovering community membership. We consider surveillance techniques based on graph theory that the adversary uses to discover community membership and counter-surveillance techniques that make it difficult for the adversary to do so. We apply complex network theory to surveillance that better enables the adversary to discover community structure. We show how counter-surveillance defences allow hardened covert cells to escape detection even in the strong threat model of a fully knowledgeable adversary and discuss the costs involved in doing so.

7 Conclusions

We have studied the network externalities of computer insecurity from a new angle, the detection of community structure and membership. In this paper, we have analyzed the interplay of surveillance and counter-surveillance tactics at the level of network topology. We have some concrete results to present. We have shown that while structural elements of a network such as nodes and edges are easily discovered when small fractions of the network are placed under surveillance, discovering community structure information requires the

adversary to invest in a significantly higher surveillance budget.

We have also shown that, regardless of whether network members communicate through an anonymous communications channels, placing 8% of the network under selective surveillance based on traffic volume is enough to compromise the community membership information of at least 45% of the nodes in the network. Our results also show that where the adversary is interested in understanding the community membership information of a far higher fraction of the nodes, the use of anonymous communication networks can increase the adversary's cost by almost 100% (80% of nodes under surveillance to uncover 99% community membership information).

Further, we have analyzed the dynamics of community detection and hiding. First, we have shown that naive strategies of edge addition between randomly selected pairs of nodes from either partition have limited community hiding capability. A linear increase in edge resources for naive surveillance defences, only leads to a small linear increase in the false negative rate of covert network detection. Pure centrality based edge formation strategies, where high centrality covert nodes connect to high centrality vertices in the rest of the network, perform somewhat better. One the average, if one edge per covert node is added between the two communities, the fraction of the hidden covert network is 57%. This is both hugely expensive and risky as it increases the profile of central nodes within the covert network.

Hybrid strategies involving a combination of the above strategies work best - Edges are added between non-critical covert nodes and high centrality nodes in the main network; Specifically, up to 70% of the covert group went undetected with a counter-surveillance budget of only an additional .082 edges per covert node. A variant strategy that associates high centrality covert nodes to randomly chosen nodes in the main delivers a more striking result: Up to 93% of the covert community remained hidden if the covert network were willing to invest in a (high) counter-surveillance budget of 3.31 edges per covert node, while almost 80% could be hidden with additional 1 edge per node!

In the study of network resilience, community detection plays an important role in uncovering crucial structural information. In the context of topological attack and defence, such intelligence can either be used to crackdown on criminal groups or by covert groups facing police repression; Or even users of a large email provider who wish to hide their existence from automated tools that sift through large amounts of traffic end-point information with the aim of filtering and shortlisting traffic for content analysis. Our results show that surveillance attacks based on exploiting partial information as well as full information can be successfully repelled if the target group carries out selective topological rewiring. Such rewiring merely requires partial topological knowledge and effectively injects a large number of false positives even in the case of a fully knowledgeable adversary with a unlimited surveillance budget. This certainly does not mean that the communities are immune from detection, but merely that attacks based on high-level traffic analysis can be repelled, forcing the adversary to invest in cryptanalysis technology in order to better detect spurious links which he then removes. Additionally, our analysis only considers aggregate traffic and does not mine the available data for the full extent of traffic information, such as timing of messages which may lead the adversary to use better selection strategies in the face of counter-surveillance defences.

Our study confirms results from previous work showing that invasion of privacy of most people is cheap and easy - by placing a mere 8% of nodes under surveillance upto 45% of community membership information can by uncovered, regardless of the use of current anonymous communication technology. Targets that wish to minimize risk can take counter-surveillance steps. By getting obscure and easily replaceable members of their

group to connect to relatively well known members of the public, covert groups can go undetected even with higher surveillance budgets - A counter-surveillance budget comprising only 0.082 edges per covert node (additional 25 edges), when applied strategically even in the scenario of partial topology information, can result in upto 70% of the covert community going undetected even when 99% of the network is under direct surveillance.

References

[AB02] R. Albert and A. Barabási. Statistical mechanics of complex networks, 2002.

[AK95] Charles J. Alpert and Andrew B. Kahng. Recent directions in netlist partitioning: a survey. *Integr. VLSI J.*, 19(1-2):1–81, 1995.

[CBN94] D. R. Cox and O. E. Barndorff-Nielsen. *Inference and Asymptotics (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, March 1994.

[CNM04] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks, August 2004.

[CSN07] Aaron Clauset, Cosma R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data, Jun 2007.

[DDDA05] Leon Danon, Albert Díaz-Guilera, Jordi Duch, and Alex Arenas. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, 9:8–+, September 2005.

[DDM03] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a type iii anonymous remailer protocol. In *IEEE Symposium on Security and Privacy*, pages 2–15, 2003.

[DMS04] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, August 2004.

[DW06] George Danezis and Bettina Wittneben. The economics of mass surveillance and the questionable value of anonymous communications. In Ross Anderson, editor, *Proceedings of the Fifth Workshop on the Economics of Information Security (WEIS 2006)*, Cambridge, UK, June 2006.

[Els05] Ulrich Elsner. Graph partitioning - a survey. *MONARCH - Dokumenten- und Publikationsservice [http://archiv.tu-chemnitz.de/cgi-bin/interfaces/oai/oai2.pl] (Germany)*, 2005.

[Fie73] Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(98):289–305, 1973.

[Fou] Electronic Frontier Foundation. At&t class action suit: Att-nsa faq.

[Fre78] Linton C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1978.

[GA05] Roger Guimera and Luis Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, February 2005.

[GK06] V. Gol'dshtein and G. A. Koganov. An indicator for community structure. *ArXiv Physics e-prints*, July 2006.

[KGV83] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220, 4598(4598):671–680, 1983.

[KL70] B.W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell Systems Technology J.*, 49(2):292–370, 1970.

[NA06] Shishir Nagaraja and Ross Anderson. the topology of covert conflict. In Tyler Moore, editor, *Pre-Proceedings of The Fifth Workshop on the Economics of Information Security*, June 2006.

[Nag07] Shishir Nagaraja. Anonymity in the wild: Mixes on unstructured networks. In Nikita Borisov and Philippe Golle, editors, *Privacy Enhancing Technologies*, volume 4776 of *Lecture Notes in Computer Science*, pages 254–271. Springer, 2007.

[New03] Mark Newman. Mixing patterns in networks. *Phys. Rev. E*, 67(2):026126, Feb 2003.

[New06] Mark Newman. Modularity and community structure in networks. *PNAS*, 103(23):8577–8582, June 2006.

[NG04] Mark Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 69(2), 2004.

[PH00] Andreas Pfitzmann and Marit Hansen. Anonymity, unobservability, and pseudonymity: A consolidated proposal for terminology. Draft, July 2000.

[PSL90] Alex Pothen, Horst D. Simon, and Kang-Pu Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM Journal on Matrix Analysis and Applications*, 11(3):430–452, 1990.

[PTVF92] William Press, Saul Teukolsky, William Vetterling, and Brian Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2nd edition, 1992.

[WF94] S. Wasserman and K. Faust. *Social network analysis*. Cambridge University Press, Cambridge, 1994.