Detecting Covert Groups Embedded in a Population

Carl A. B. Pearson[1,*], Edo Airoldi[2], Edward Kao[2], Burton Singer[1],

1 Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

2 Statistics, Harvard University, Cambridge, MA, USA

∗ **E-mail: cap10@ufl.edu**

Abstract

We specify a graph-based model of populace-wide communications, with an embedded, relatively small module representing a clandestine group. The members of this group behave similarly to the background population, except that they also pass special messages in furtherance of a plan. We parametrize this model based on cell phone data sets.

Using simulated message traffic on this network, we benchmark various strategies, a particular set of which we call an *Observer*, for detecting the clandestine group. We measure several Observers for their performance in terms of detection rate and accuracy measures (e.g, Receiver Operator Characteristic) relative to statistical features of the general population, the clandestine group, and their respective communication behaviors.

Finally, we consider the implications of *forged* messages. In the basic model, we consider incomplete information about the communications, but the available information is always accurate. In this extension, we allow the Observer and the clandestine group to forge messages. We again measure various Observer performance traits relative to properties of the observed network.

Introduction

For investigators ranging from anthropologists to law enforcement, the need to identify groups which wish to remain anonymous can be paramount. In particular, the need for intelligence organizations to identify terrorist cells and defuse their violent plots is a matter of increasing import. As such, we will use the extant evidence about Salafi jihad networks as our motivating case [?], though we will point out where assumptions can be modified to identify different kinds of groups against a background population.

Model

Sageman, Qin, et al. identified the structure of the Salafi networks to comprise a few key individuals with links to a large group of lieutenants – the middle management of terror – that are each connected to several tightly clustered subordinate groups that execute plots. The lieutenants typically integrate with the regular population, while the subordinate groups are largely cloistered.

To represent the three components – the background population, the lieutenants, the subordinate clusters – we generate the graph from clusters with the features of each of these. Vertices are people ($\mathbf{P} = \{P_1, P_2, \ldots P_k\}, n(\mathbf{P}) = k$), with a directed edge from P_i to P_j if person i initiates communication with person j. Communication takes the form of messages of a simplified sort: a binary "good" or "bad" signal. Full instructions from a lieutenant to the subordinate groups that will implement a terrorist event consist of a cumulation of several bad signals. Additionally, bad signals can be transmitted by anyone in the general population, though these play no role in the plot.

In the following sections, we provide the details of generating the groups, assembling them into a whole, and finally their communication behavior. For our simulations, we focus on a population that contains a single lieutenant coordinating multiple subordinate clusters, though we acknowledge that more realistic scenarios would typically entail tracking multiple plotting groups.

The Background Population, the $P_n \in \mathbf{P}$

Individuals in the background population are members of multiple communities, divided among multiple dimensions – e.g., family, religion, work. Most of these connections are bi-directional.

For our simulation, we assume that each of the background individuals is a member of three independent community "dimensions"[1]. We form each of these dimensions by generating community structures according to a community size distribution and formation algorithm, until the total number of vertices in a dimension equals the total population. We then randomly assign individuals to one single community in each dimension, and then "flatten" the resulting graph by merging any edges that are duplicated across dimensions.

To form a community, we sample the size distribution to determine community size. We then form

[1] **TODO: could this assumption be informed? I'm making it for simplicity, but I think its plausible that someone has done research about people's group identities, the extent to which those are non-indepedent, and concluded about the range of independent groups a person can be called a member of**

completely connected triads up to that size, adding the potentially remaining 1 or 2 vertices to triad (or two) and forming a completely connected quartet. We then take these cliques 3 at a time, and treating them like vertices, form more completely connected triads by choosing a random member for each clique to bidirectionally connect (left over cliques are treated similarly to left over vertices). This joining proceeds iteratively until there are fewer than three objects at a scale to connect. We then consider each unjoined pair of vertices and form a directed link with low probability r_p.

A Lieutenant, the H Vertex

H has community affiliations like most members of the population. However, **H** is a member of more communities than the typical individual in the population given the need to gather information, identify recruits, etc. Finally, **H** is completely connected to the members of the clusters, but those connections are only directed from **H** to the cluster members.

H is added to a number of communities sampled from a distribution and then randomly connected with members in that community with probability c_o (to that member) and c_i (from that member). This distribution and these probabilities should be set relative to background population connection structures such that the **H** is a high outlier for both the in and out degree distributions of the population.[2]

The Subordinates, $C_i \in \mathbf{C}$

Each C_i is a bi-direction clique, comprising a small number of individuals. In our simulations, we sample from a binomial clique size. The C_i have no other structured communication channels.

Message Passing Behavior

O understands the network by monitoring message traffic between individuals. For this analysis, we consider messages with binary state only: the message is either "good" or "bad".

The background population generate these messages according to simplifying assumptions about the real world: they have no preference for their community memberships beyond how many members of a community they connect to, their messaging activity occupies an inconsequential period of time during any iteration, and the iteration time is such that multiple real events (e.g., a few calls between individuals) can be treated as a single continuous event. Thus, during each iteration, each individual $P_i \notin \mathbf{C}_n \cap \mathbf{H}$

[2] **TODO what distribution? math to enforce prob limits?**

(1) activates its out degrees with probability ρ_m – *i.e.*, a person does a binomial sample of the available channels – and then (2), P_i sends a single message to each active channel. These messages are "bad" with a low probability p_b.

TODO equations for P_i outgoing messages, probability of sending a bad one.

Like the P_i, **H** abides by the simplifying assumptions about the real world, with one small perturbation. **H** is a member of many communities, and strategically cultivates and exploits these memberships. To model that, **H** will send at least one message to each community he is in, and possibly more. So, for each community **H** is a member of, **H** sends $1 + \binom{n}{k_i - 1}$ messages, where k_i is the number of connections **H** has within that community.[3]

Any given iteration, **H** may also issue directives to the subordinate groups. These messages will always be "bad" messages, but (1) are sent with low probability h_b and (2) are sent to only one member of any particular C_i, since any C_i member can be assumed to instantly disperse this information to the others.

TODO equations for **H** outgoing messages, probability of sending a bad one.

The members of each C_i are largely "silent" – which is to say, their communication is direct and is largely untraceable. Rarely, however, they will break their direct communication discipline, or otherwise be observed to interact. Each iteration, they may communicate with one other member of their C_i (chosen uniformly) with low probability c_m; these messages have a relatively high probability of being "bad", c_b. If any member of the clique received a message from **H** the previous iteration, one member (chosen uniformly) may communicate a bad message to another C_j (which j chosen uniformly, which member of C_j chosen uniformly) with probability c_o. This models the largely untraceable communication among members in a C_i and between C_j's.

TODO equations for C_i outgoing messages.

Finally, there is a low probability ρ_r of an individual sending a random message to an individual they do *not* have an outgoing link to each iteration. For each individual that will send one of these messages, they recipient is selected uniformly from the candidate recipients.

TODO equation

[3] **TODO: good vs. badness of these messages? Original: "These messages will always be "good" messages." but certainly if he's actively recruiting, or testing the waters, or looking to gather sensitive intel, etc. these could all be "bad" messages.**

Observers

A particular **O**, for a particular scenario, only observes the message traffic as it comes along. **O** does directly see any structural features of the population graph, nor does **O** know certainly when messages are among normal individuals or the plotters, let alone within a particular community. Different **O**'s may make different assumptions about these features, and use those features to target their monitoring and even adjust their beliefs about those features according to the message traffic that occurs.

For all of the strategies we consider, our **O**'s make some limiting assumptions consistent with those we put in the model, specifically that there is a single **H**, connected to multiple C_i. The **O** knows that both the plotters and background population can send "bad" messages.

Additionally, **O** makes other assumptions about the structure of background communities, the messaging rate of various parties, and other features of the graph. These assumptions are not necessarily correct, but **O** still makes them as starting guess.

Strategy 1

Strategy 2

Strategy n

Results Ignoring Decoys

Considering Decoys

There are two fundamental sorts of deception available in the model: deception by $\mathbf{H} \cap \mathbf{C}$ and deception by **O**. Deception by the plotters includes sending "bad" messages to the background, and, assuming the background is sympathetic, having those bad messages echoed along. Or forging bad messages between members of the background. Deception by **O** entails forging messages from a false **H**, in hopes of hitting one of the C_i and seeing them respond accordingly.

Discussion

Appendices

Parametrizing the Graph

TODO a-b-c calculation to tune to afghan cell phone data or other parameter studies?

Implementation, the DarkNet API, and Extension

TODO description of software package. where to get source, how to write extensions. Propose some extensions: give individuals a vocabulary assigned from a distribution, have them randomly assemble messages from the vocabulary, some “words” of which are “bad”. Allow multiple communication channel types. Include a community dimension and have “within community X” as part of the message information.