Detecting Covert Groups Embedded in a Population

Carl A. B. Pearson[1,2,*], 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 background population, except 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 [1], though we will point out where assumptions can be modified to identify of kinds of groups against a background population.

Model

Sageman et al. identified the structure of the Salafi networks to be a few key individuals with links to a large group of lieutenants – the middle management of terror – that in turn each connected to several tightly clustered subordinate groups that execute plots. The lieutenants typical integrate with 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.

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 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 $P_n \in \mathbf{P}$

The background population comprises multiple distinct communities, bridged by random connections. Individuals are members of multiple communities, divided among multiple dimensions – e.g., family, religion, work. Most of these connections are a bi-directional.

TODO which community formation algorithm?

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.

TODO algorithm for **H** in communities? Draft: pick a larger than typical number of communities of membership, then add **H** to that that many communities. Possibly preferentially to certain community types.

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 consider only triads, leaving the features of larger groups (e.g., more opportunities to violate communication tactics) to represented by other model parameters. The C_i have no other structured communication channels.

Integrating $\mathbf{P} \cup \mathbf{H} \cup \mathbf{C}$

TODO need this section? the random interconnection of background population should be accomplished by multiple-community formation algorithm.

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 all their community memberships equally, their messaging activity occupies an inconsequential period of time during any iteration, and the iteration time is such that multiple real communicate events (e.g., a few calls between individuals) can be treated as a single continuous event. Thus, during each iteration, each individual $P_i \notin \mathbf{C}_n \cap \mathbf{H}$ (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 every to each community it is in, and possibly more. That is, for each community **H** is a member of, **H** will send $1 + \binom{n}{k_i - 1}$ messages, where k_i is the number of connections **H** has within that community. These messages will always be "good" messages.

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.

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

The members of each C_i are largely silent. 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; also consider: more likely to send to people with a link to them?

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 $\mathbf{O}$. 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 $\mathbf{O}$ entails forging messages from a false $\mathbf{H}$, 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.

References

[1] Qin J, Xu J, Hu D, Sageman M, Chen H (2005) Analyzing terrorist networks: A case study of the global salafi jihad network. In: Kantor P, Muresan G, Roberts F, Zeng D, Wang FY, et al., editors, Intelligence and Security Informatics, Springer Berlin Heidelberg, volume 3495 of *Lecture Notes in Computer Science*. pp. 287–304.