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 and demonstrate 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 messages in furtherance of some plot. We parametrize this model based on cell phone data sets.

The purpose of this model is to provide a test framework for various methods of detecting these clandestine groups, in real time, before they achieve the intended plot. We refer to collection of such methods as an *Observer*. We propose various Observer models and measure their performance relative to statistical features of the population, plotters, and their respective communication behaviors.

Finally, we consider the implications of *decoy* messages. In the basic model, we consider missing – but not misleading or false – communications. If, instead, the plotters or the Observer can issue forged messages, the problem becomes substantially more complex.

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

We represent a population as a directed graph. 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 contact with person j. Our simulation tool allows for multiple edges from one vertex to another, representing multiple avenues of communication, subject to different levels of monitoring. In these analyses, we do not exploit having multiple edges between individuals, and only use this capability to distinguish between monitored and unmonitored channels.

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 local groups that execute plots. We refer to these lieutenants as "hubs" or the single $\mathbf{H}$ vertex in our population graphs. In the practical cases we present, we will consider our terrorist groups to consist of a single hub no higher leadership element. In addition to general interactions with the population at large, the $\mathbf{H}$ will have connections to one or more small terrorist clusters; we refer to these clustered groups as $\mathbf{C}_n$, enumerating the clusters from 1: $\mathbf{C}_1$, $\mathbf{C}_2$, $\ldots \mathbf{C}_k$ when the $\mathbf{H}$ has k subordinate clusters.

As to the explicit formation of these networks, [1] suggests

Message Generation

$\mathbf{O}$ 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".

We use a simple model for generating these messages between the background population. During each iteration, for each individual $P_i \notin \mathbf{C}_n \cap \mathbf{H}$:

- P_i activates its out degrees with probability ρ_m – *i.e.*, a person does a binomial sample of the available channels,
- P_i sends a single message to each active channel, and
- this message is "bad" with a low probability ρ_b.

The $\mathbf{H}$ and $\mathbf{C}_n$ have their own messaging behavior:

- $\mathbf{H}$ behaves like a typical module member, but never sends bad messages, except to the $\mathbf{C}_n$ at a low rate h_b, and

- the members of $\mathbf{C}_k$ will a send single "good" message per iteration, with low probability c, to another member of $\mathbf{C}_k$. If any member of a $\mathbf{C}_k$ received a "bad" message from $\mathbf{H}$ in the previous interval, these messages will instead be "bad". Additionally, when sending a "bad" message, members of a $\mathbf{C}_k$ may instead randomly send the message outside the cluster with probability c_o, to one of the other $\mathbf{C}_j$ with uniform probability.

Finally, all of $\mathbf{P} \cap \mathbf{C}_n \cap \mathbf{H}$ may send a message with probability ρ_r to any other member of the network (with uniform probability).

0.1 Subtitle

Plain text.

0.2 Another subtitle

More plain text.

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.