

# Detection of Small Covert Networks Embedded in Large Networks

Carl A. B. Pearson <sup>1</sup>    Burton H. Singer <sup>1</sup>    Edo Airolidi <sup>2</sup>

<sup>1</sup>Emerging Pathogens Institute, University of Florida

<sup>2</sup>Harvard University

May 22, 2013

# TODO FUNDING INFORMATION

# Overview

- ▶ Definitions,
- ▶ A Model to Reflect Those,
- ▶ Implementation for a Particular Case: Salafi Jihadi Network,
- ▶ Detecting Groups in this Model,
- ▶ Some Detection Results for that Implementation, and
- ▶ Flaws, Extensions, and Outlook

## What is *Covert*?

a *covert network* is a sub graph where interaction information is some combination of unavailable, unreliable, or (mostly) indistinguishable from the enclosing graph structure

## ...or Operationally

A relatively small, organized group of conspirators, masking their existence via communication discipline and taking advantage of a noisy background.

For this particular talk: Salafi Jihadi network as described by Sageman, et al.

Note: not that group's more recent focus on *leaderless jihad*.

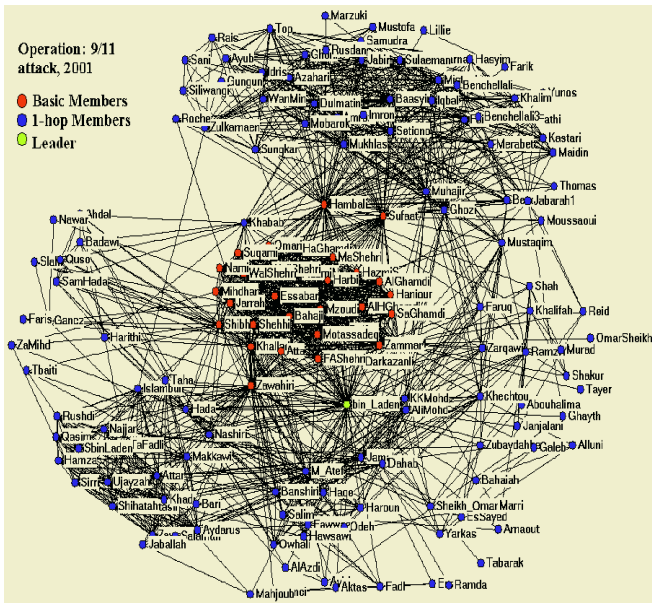


Fig. 4. The 1-hop Network of the September 11<sup>th</sup> Attack

## Salient Features

- ▶ highly interconnected subordinate groups, and
- ▶ bridging middle managers,
- ▶ communications masked with some tradecraft,
- ▶ missing from picture: vast background population

# Simple Implementation addressing a Salafi Jihadi-like Network

- background population** many small cliques, which are recursively cliqued into single graph
- covert leader** stochastically added to cliques, outgoing connections to a random member of each of the covert groups
- subordinates** few, medium size cliques with connections between clusters
- communications** simple message content *Good* vs. *Bad*

## ... or Symbolically

- ▶ a structured population,  $P$ ,
- ▶ covert leader,  $H$ ,
- ▶ subordinate covert groups,  $\{C_i\}$ ,
- ▶ stochastic behavior model for intra- and inter-group messages,
- ▶ drawn from a set vocabulary,  $V$



## Aside: Sales Pitch

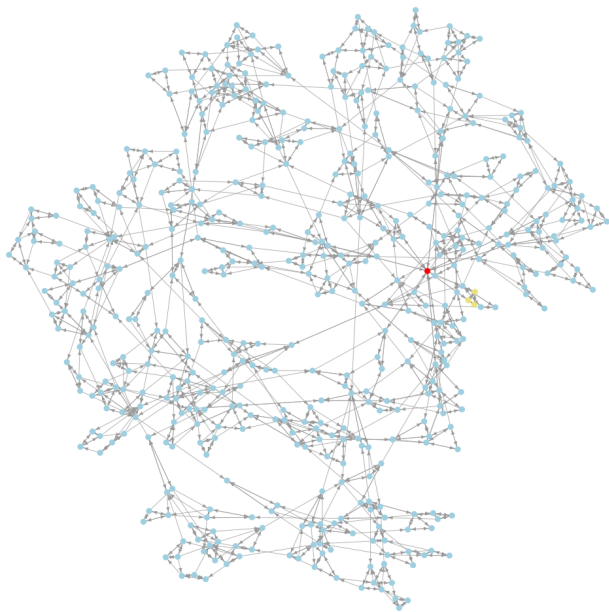
Scala-based Implementation available for review/remix:

<https://github.com/pearsonca/scala-commsim>

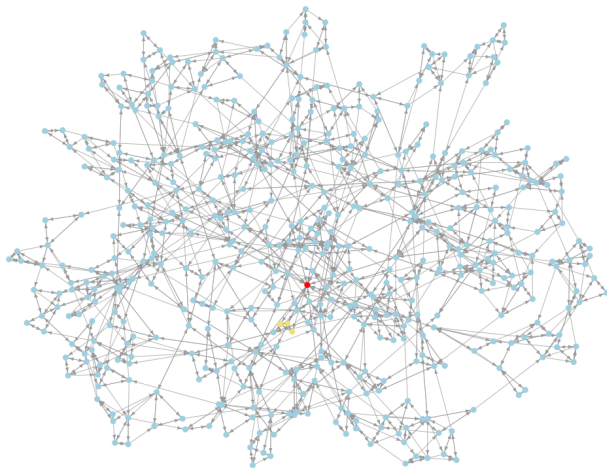
Actively moving from closed, non-Scala implementation to that repository. Please request changes, point out bugs, etc.

Also, this presentation:

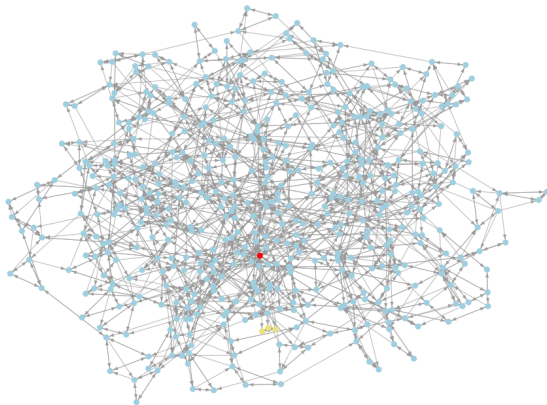
<https://github.com/pearsonca/sunbelt13-presentation>



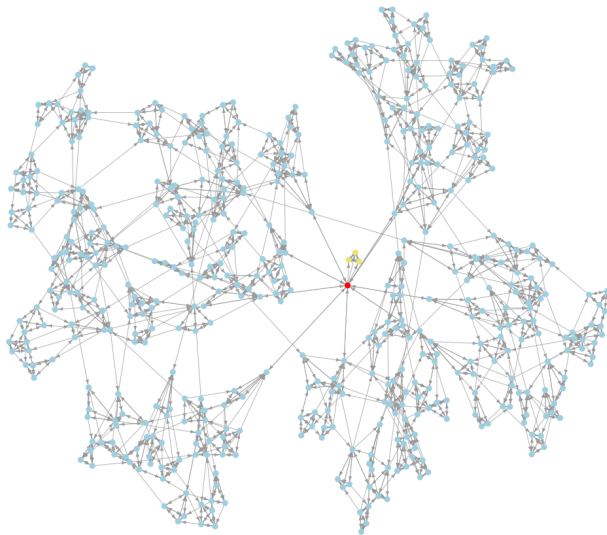
3 clique, 1% remix



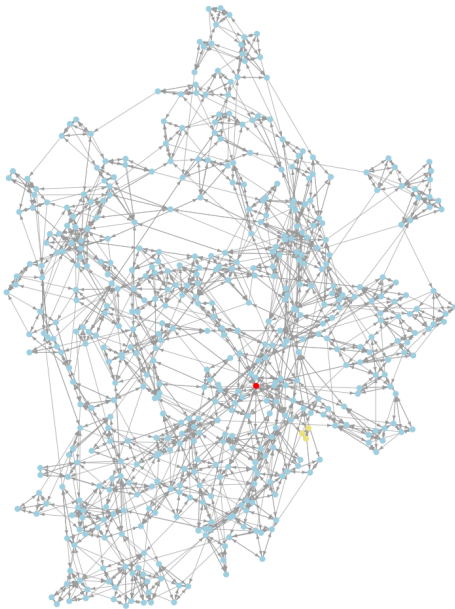
3 clique, 10% remix



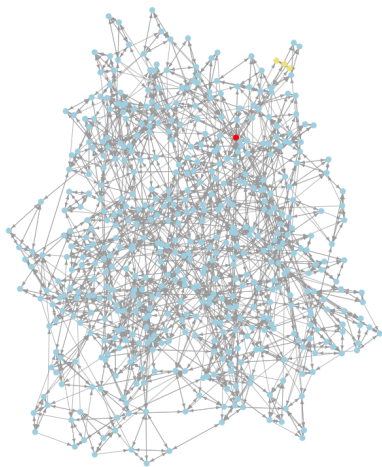
3 clique, 30% remix



4 clique, 1% remix



4 clique, 10% remix



4 clique, 30% remix

# Steady State, Structural View vs. Real Time



# Real Time Challenges to Detection

- ▶ population vs. covert group communication network initially unknown,
- ▶ potentially limited resources for monitoring those communications,
- ▶ thus gathered information unreliable / incomplete,
- ▶ and risk trade-offs: FPR & TPR vs. action by group

# Our Model: The Observer

An algorithmic description of

- ▶ the data limitations (e.g., random suppression or transformation of signals), and
- ▶ detection strategy(ies)

# Some Simple Strategies

- ▶ pure content: pick up everyone that has sent and received a *Bad* message
- ▶ pure structural: pick up highest degree person and all people below median
- ▶ mixed structural and content

# Appropriate Measures?

Assume that any given plot has some critical amount of planning-related communication.

But what else?

- ▶ true positive rate,
- ▶ false positive rate,
- ▶ resource investment

# Results For These Modes

TODO series of plots

# Flaws, Extensions, and Outlook

- ▶ limited vocabulary – add message diversity, require content detection as well,
- ▶ unsophisticated Observer model and strategies – add resource model, shifting strategies
- ▶ background / foreground structural generations – new generators, fitting to live traffic