**High-level problem definition**

"How to win the battle against terrorism"

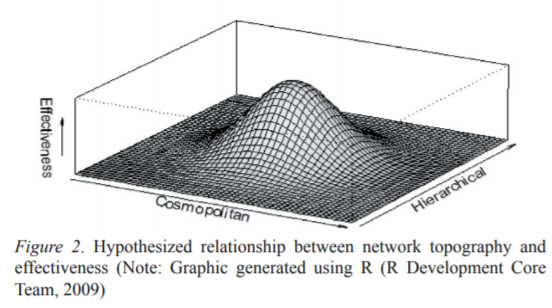
**Detailed problem definition**

What vulnerabilities exist in the terrorist relations network and how can we exploit this to break it apart?

**Some ideas of how to define vulnerability (different colors are different tasks):**

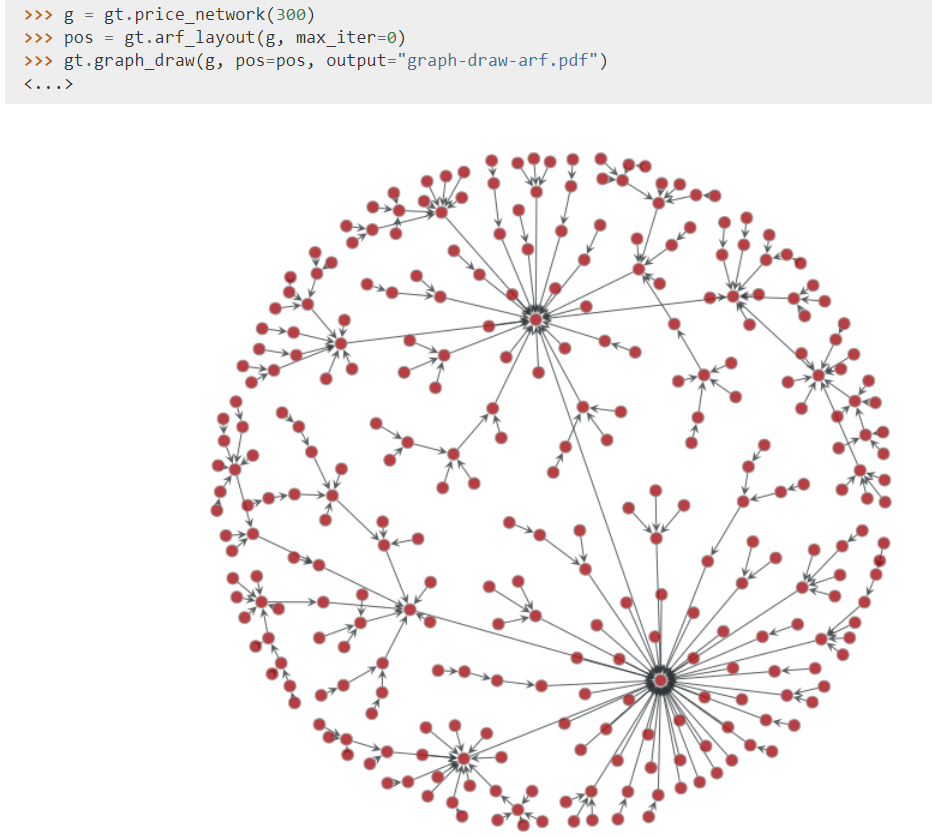
* Invert the network first.
* Generate **feature vectors** for each node (terrorist) which includes things like degree, clustering coeff, betweenness centrality, relations, etc. and conduct clustering to identify communities in this way. Does removing a few of these clusters (or terrorists) make our network random (and thus effectively destroy its effectiveness)?
* **Number of communities** (in our largest component) and **their connectivity** with each other.
* **Minimum number of links or nodes to cut** to disconnect these communities.
* Evaluating **smoothness of randomly placed Dirac impulses** across the entire network (smoothness implies well communicating network)
* Invert the network first so that nodes represent terrorists and edges relations. Compute **probability transition matrix** to reveal **which terrorists have the highest probabilities**; perhaps they are the most important people in the network in the sense that everyone else is most likely to end up “meeting” these terrorists, check also the entire network in this case. What happens when we remove (assassinate) these terrorists?
* Betweenness centrality: number of shortest paths that pass through a node. Invert the network first. Identify the top 5 (or top 10) nodes with the highest betweenness centrality. What happens when we remove these terrorists?
* Closeness centrality: reciprocal of the sum of the lengths of shortest paths between a node and all other nodes in the network.
* Invert the network first. Can we **classify the label (relation) of the edges**? If we can, it means we can accurately identify the type of relations between any new terrorist who enters the network with the rest.

**Network resilience:**

* **Motivation:** The approach of identifying and removing key players can be limited because it does not consider other important properties of the network such as resilience or adaptive capacity that are directly linked to the overall success of the network.
* ***Provincial-Cosmopolitan* trade off:** “[An ideal mix of weak and strong ties appears to provide benefits at the individual level as well as the organizational level](https://my.nps.edu/documents/104382430/104582412/Everton+2012+%28Topography%29.pdf/c6c7a7c1-3178-4bfb-aeaf-aa6120bdca2e)”[[1]](#footnote-1). Weak ties often connect disconnected groups and hence are important for organizational survival. Examples:
  + When it comes to finding their current job, people have used personal ties particularly weak ties (i.e. acquaintances) far more than strong ties (i.e. close friends)[[2]](#footnote-2).
  + There exists an inverted U relationship between the degree of embeddedness (i.e. strong ties) and the probability of firm failure[[3]](#footnote-3).
  + The inverted U relationship also exists between the extent to which creative teams producing Broadway musicals are connected and cohesive vs. the financial success of the shows[[4]](#footnote-4).
  + **Metrics to estimate the provincial-cosmopolitan dimension**:
    - Average degree centrality + its variance
    - Local clustering coefficient normalized by that of a random network
  + When we visualize our network after color coding the nodes according to their cluster assignments, can we visually identify weak ties between these clusters? If so, are these ties something we can potentially remove i.e. if they are “congregate”, we may be able to achieve this by destroying their common facility.
* ***Heterarchical-Hierarchical* trade off:** “[An optimal level of centralization (hierarchy) exists](https://my.nps.edu/documents/104382430/104582412/Everton+2012+%28Topography%29.pdf/c6c7a7c1-3178-4bfb-aeaf-aa6120bdca2e)” for the overall success of the network, much like the ratio of strong to weak ties.
  + **Metrics to estimate the heterarchical-hierarchical dimension:**
    - Closeness centrality + its variance across the network
    - Between centrality + its variance across the network
  + Is our network relatively centralized?
    - If yes, an actionable strategy would be a misinformation campaign that breeds distrust in the network and hence higher level of centralization, which will in fact make it less effective.
    - If no, we can amplify this by removing key players and making the network even more decentralized.
* [Image source](https://my.nps.edu/documents/104382430/104582412/Everton+2012+%28Topography%29.pdf/c6c7a7c1-3178-4bfb-aeaf-aa6120bdca2e)
  + Cosmopolitan = degree of sparsity (lack of strong ties)
  + Hierarchical = degree of centralization (opposite of network)

**Visualization ideas:**

* Visualize both the terrorists as edges and nodes
* Python graph\_tool ARF layout to represent our inverted network:



* GEPHI

**Important reminder:**

We want to have something to hit every box in the data science process below. Some examples include:

**Data processing**: network inversion

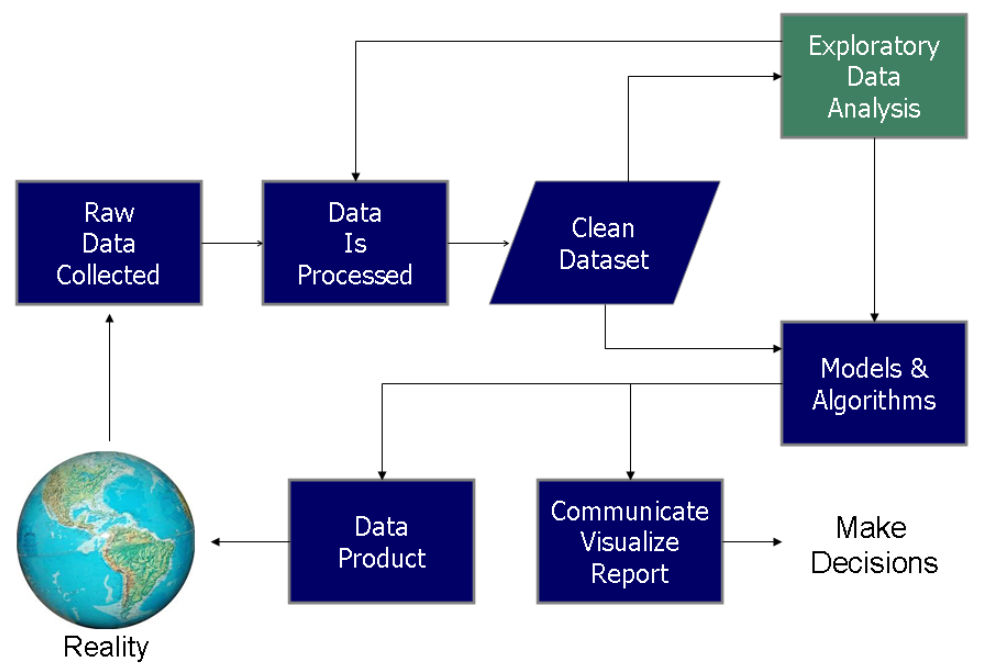
* Every pair of linked nodes in the original graph must have one name in common.
* If above is true, we can invert our network.
* Find number of disconnected components (first eigenvalue). Is there one dominant community so large that it is sufficient to study this one only? Otherwise, work with several communities separately.

**Data cleaning**: detecting empty names and replacing them with unique datetime strings (we discovered that a terrorist always has a unique datetime string, and even when a node is missing the name it always has a datetime string attached to it)

**Exploratory data analysis**: computing network properties (betweenness centrality, #communities, smoothness, etc.)

**Models/algorithms**: clustering task

**Data product**: “So what are the vulnerable points of the network?”, “How will these inform decisions or actions for fighting terrorism?”



1. [Link to paper](https://my.nps.edu/documents/104382430/104582412/Everton+2012+%28Topography%29.pdf/c6c7a7c1-3178-4bfb-aeaf-aa6120bdca2e) [↑](#footnote-ref-1)
2. [Link to paper](https://www.cs.umd.edu/~golbeck/INST633o/granovetterTies.pdf) [↑](#footnote-ref-2)
3. [Link to paper](https://www.researchgate.net/publication/243765837_The_Sources_and_Consequences_of_Embeddedness_for_the_Economic_Performance_of_Organizations_The_Network_Effect) [↑](#footnote-ref-3)
4. [Link to paper](https://www.kellogg.northwestern.edu/faculty/uzzi/ftp/uzzi%27s_research_papers/0900904.pdf) [↑](#footnote-ref-4)