Target Policy-making Under the Frame of Dark Networks

Strengths, Weaknesses & Opportunities

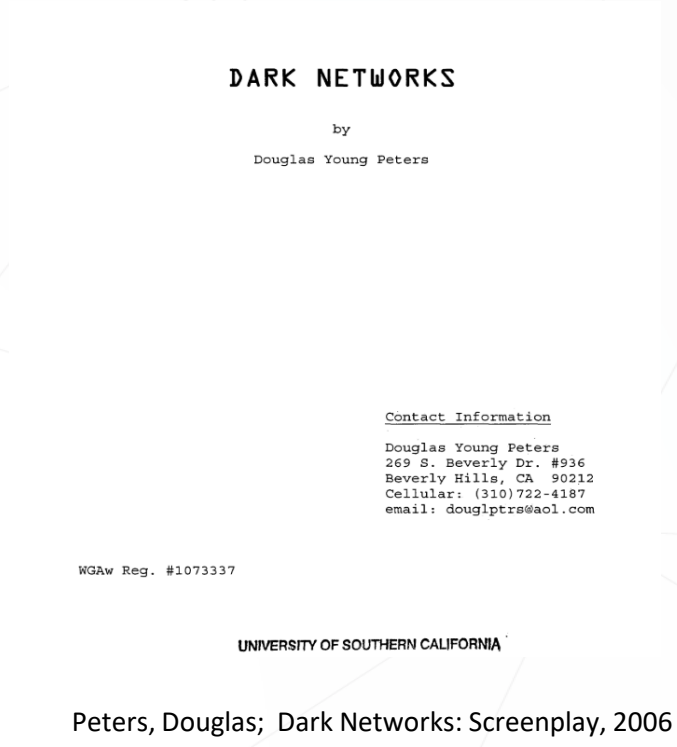
Joseph A. E. Shaheen
ORISE Intelligence Community Postdoc Fellow

Host Institution: Department of Computational & Data Science
George Mason University

This research was supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at George Mason University, administered by Oak Ridge Institute for Science and Education through an inter-agency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence.

The Origins of “Dark Networks”

- ***First instance*** of “Dark Networks”—Raab & Milward (2003) –Dark Networks as Problems
- Milward and Raab, 2006; Milward, 2006—Dark Networks as Organizational Problems
- Xu & Chen, 2008—The Topology of Dark Networks
- Bohannon, 2009—Investigating Networks: The Dark Side [***Science Editorial***]
- Keller; Atkinson; Roberts; Keegan (all 2010)
- 2010+ Deluge of “Dark Network” claims



DARK NETWORKS

by

Douglas Young Peters

Contact Information

Douglas Young Peters
269 S. Beverly Dr. #936
Beverly Hills, CA 90212
Cellular: (310)722-4187
email: douglptrs@aol.com

WGAw Reg. #1073337

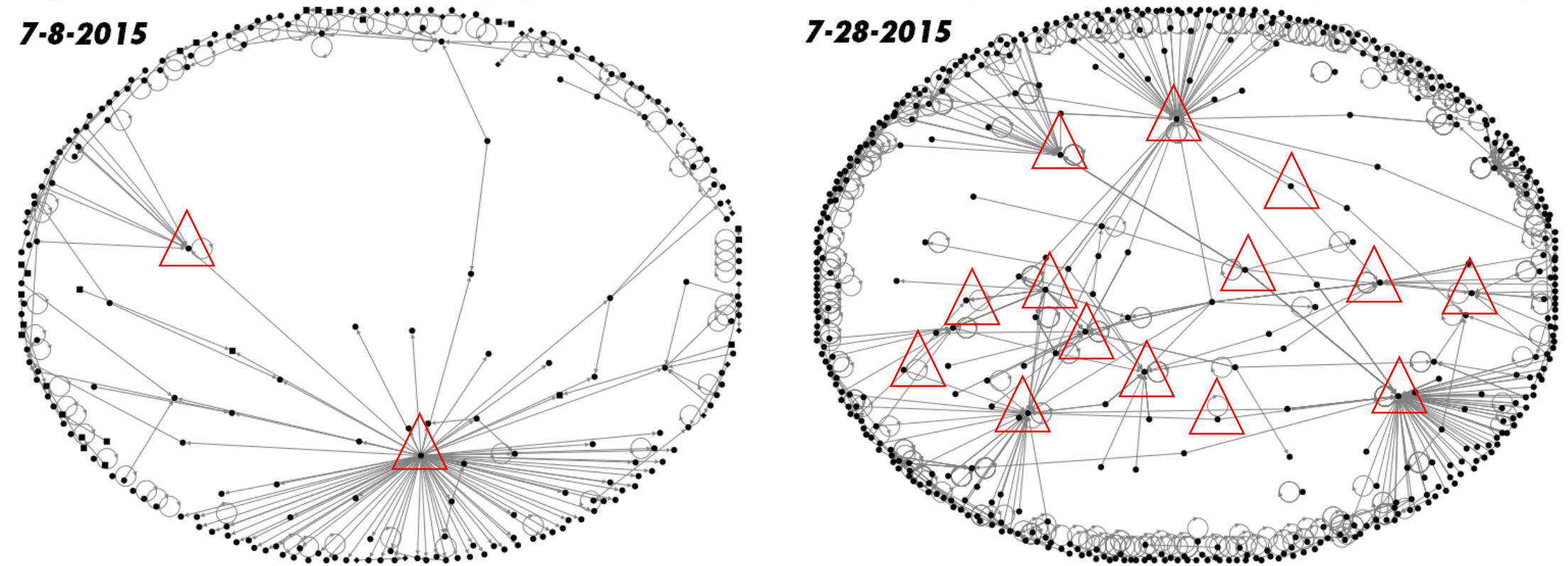
UNIVERSITY OF SOUTHERN CALIFORNIA

Peters, Douglas; Dark Networks: Screenplay, 2006

Salient Themes

- Dark Networks are ***different***. (Raab & Milward)
- ***Centrality*** is important (Xu and Chen)
- Remove the most central actor and someone else "moves in" (Tsevotat & Carley)
- ***SNA inappropriate*** for this domain (Valente)
- There are ***key players*** that can be identified; their removal would "break" the network (Borgatti)
- **Main Theme:** Here are a ***collection of methods drawn from classical SNA***… Now, let's talk about bad people. Occasionally, here is a method or two from the network resilience line of inquiry clothed in Network Science sprinkled with a few agent-based models.

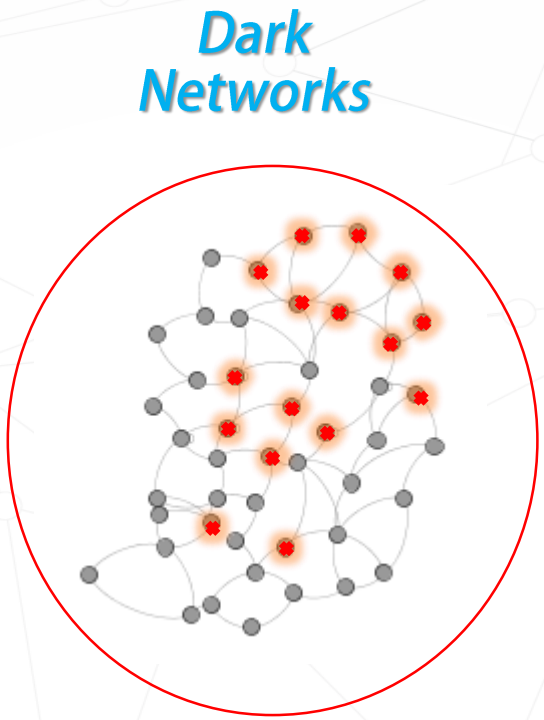
Example of Current Framework



Source: **Shaheen, J. A. E.** (2015). *Network of Terror: How Daesh Uses Adaptive Social Networks To Spread its Message*. *NATO Stratcom Centre of Excellence* (Vol. 1). Retrieved from https://www.stratcomcoe.org/network-terror-how-daesh-uses-adaptive-social-networks-spread-its-message

What is a Successful Framework?

- Should it explain? Predict?
- Predictive Power != Explanatory Power
- A collection of ***effective*** ideas, theories, tools (methodological) and principles.



Claim: "Surprise" as Risk Management

- Analytical tools for security policy-making ***should enrich***…security policy-making.
- In the absence of additional network information that can plainly identify dark actors (such as behavioral), one must, through an analytical tool ***identify the risk*** (e.g. structural, processual) of central actors becoming "dark".
- ***Analytical conclusions*** can then be made, and interventions, such as "Hardening" can then be employed.
- There is a branch of applied mathematics that offers a measure of surprise: ***Information Entropy***

Differential network entropy reveals cancer system hallmarks

James West[1,2,3], Ginestra Bianconi[4], Simone Severini[2,3] & Andrew E. Teschendorff[1,3]

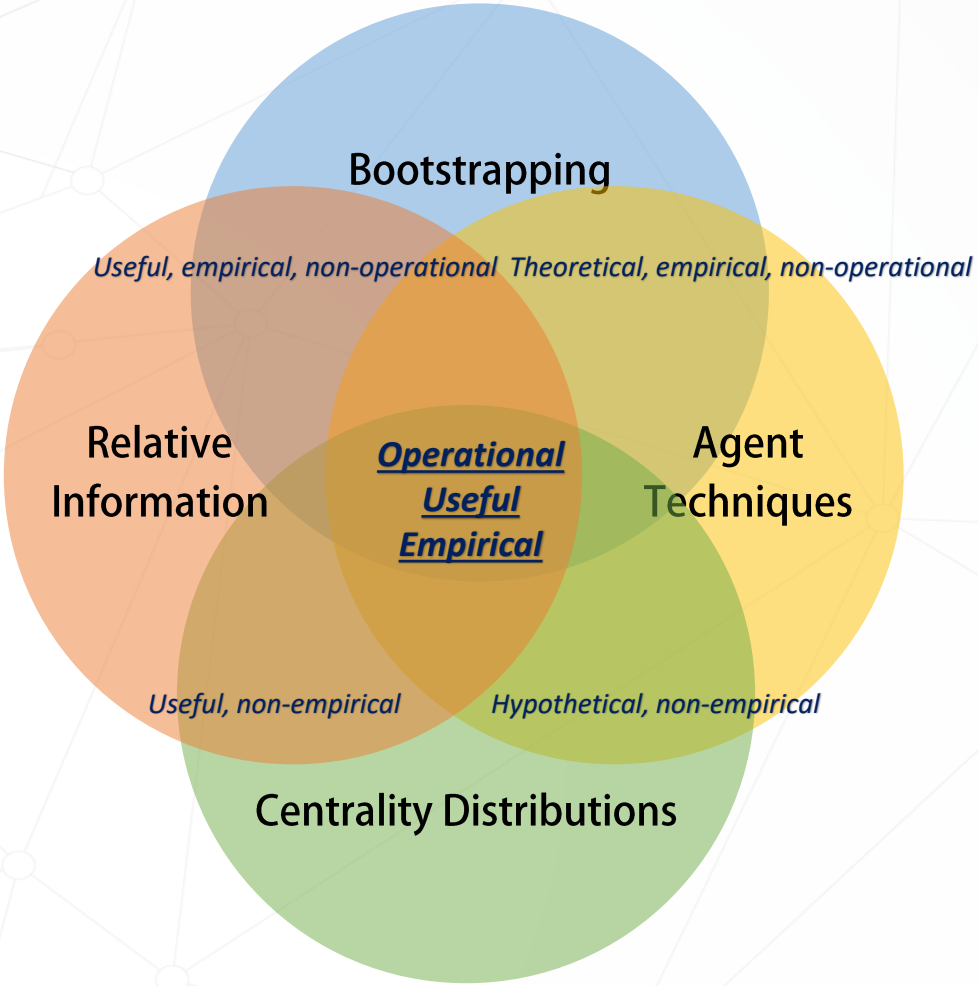
Hardening

Simply Put: Instead/addition of/to who should we target for node deletion, the framework should answer who causes maximum "surprise" if they were to be a "Dark Actor"

Methodology

- Information Entropy is a ***representation of the information contained within*** a closed system. For the majority of parametric distributions, a closed form solution exists
- Relative Entropy or the ***Kullback-Leibler Divergence*** is a measure of ***information loss or gain***, or the divergence of one distribution's information content from another
- The ***Configuration Model*** is a generalized random graph model that relies on fixing the degree distribution of a random network – so called ***degree sequence***. Computationally, it uses a ***bootstrapping*** technique along with a vast amount of theoretical and mathematical tools, allowing for the permutation/perturbation of degree sequences (and distributions) while measuring a variable or effect of interest.
- Agent-based modeling is a set of techniques that connect behaviors at the micro-states to properties of the system (we won't need that today)

→ ***Network ensemble comparison***



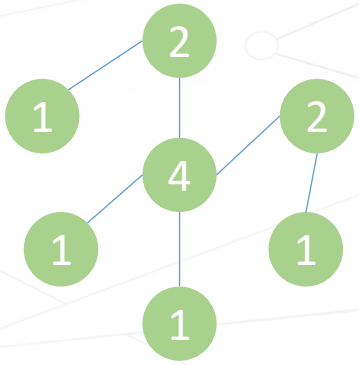
Methodology

1. Given some graph ***G***, with ***n*** nodes and ***m*** edges, and some fixed degree sequence $k = \{k_1, k_2, \ldots, k_n\}$ and some normalized centrality measure $c = \{c_1, c_2, \ldots, c_n\}$ let us calculate the ***change in entropy D*** (KL Divergence) of ***G*** given some new degree-preserving configuration (matching) ***G'.***.
2. Furthermore, we must find a ***re-wiring method*** that can approximate the contribution of some node ***i*** to the $D_{KL}(\boldsymbol{G} \,||\, \boldsymbol{G'})$ for each configuration

Methodology

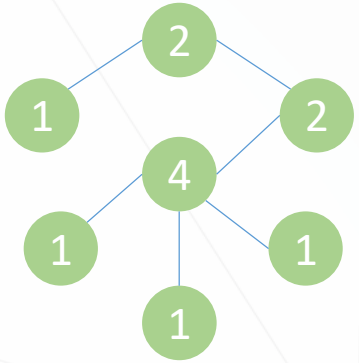
- MCMC Sampling
- Assume simple and undirected graphs
- We will use degree-preserving rewiring (sometimes known as ***double-edge-swap vertex labeled rewiring***.)
- Rewiring will allow for self-loops
- Applied to configuring the whole network 2 stubs at a time.

Original Network



Degree preserving

Configuration/Match



$$C(x) = \frac{N}{\sum_y d(y,x)} \qquad D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right).$$

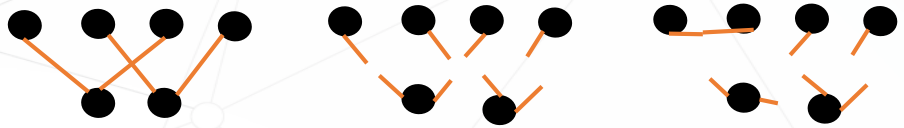
Rely on theoretical derivations from Zichao, L. I., Mucha, P. J., & Taylor, D. (2018). Network-ensemble comparisons with stochastic rewiring and von neumann entropy. SIAM Journal on Applied Mathematics, 78(2), 897–920. https://doi.org/10.1137/17M1124218*

Methodology

- We run into an issue if we want to know the ***specific contribution*** of $D_{KL,i}$ *(c)—or* the entropy loss/gain given the re-wiring of a specific node.
- Remember: We'd like to know the given amount of topological ***"surprise"*** given a fixed degree sequence as we re-wire the network
- Heuristic Solution: Since the minimum number of rewires needed to sample a single configuration/match is $\sum_i k_i = \frac{2m}{2} = m$ (reconfigured 2 at a time), and that option is not available to us, because we need the node-level entropy change, ***we'll need to reconfigure 4 at a time***, or a double dyad and so at a rate of $\sum_i k_i = \frac{m}{2}$ .But that still doesn't solve the attribution problem
- We must modify the double-edge-swap vertex labeled rewiring method and introduce the ***double-dyad-swap vertex labeled rewiring*** method.

double-edge-swap vertex labeled rewiring

1. *Original* 2. *Stub network* 3. *Rewiring 1 dyad*



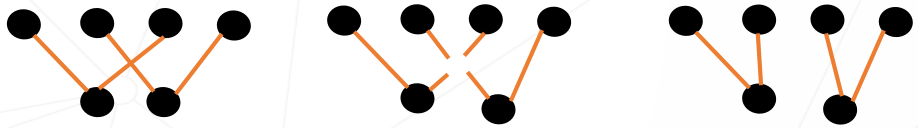
Ok try to calculate D_{kl} now please. ***Dude, I can't. I'm not done yet.***

$$N(\{k_i\}) = \prod_i k_i! \qquad N(\{k_i\}) = \frac{\prod_i k_i!}{\prod_{i<j} A_{ij}! \prod_i A_{ii}!!}$$

Departing from the Configuration Model

*double-**dyad**-swap vertex labeled rewiring*

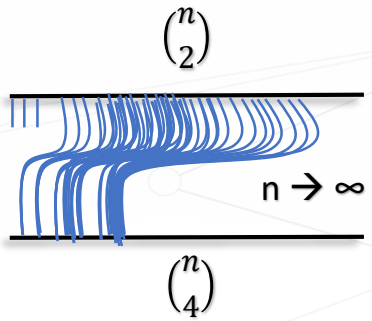
1. *Original* 2. *Stub network* 3. *Rewiring 2 dyads*

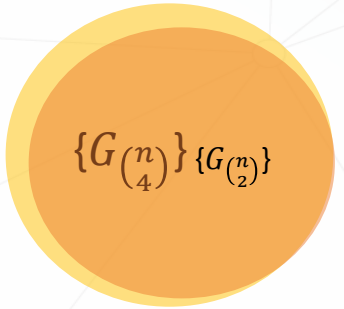


$$N(\{k_i\}) = \prod_i k_i!$$

Methodology

- We have ***departed from the configuration model***: Our rewiring method fundamentally contrasts the network generation mechanism that the configuration model relies on → Now it's just bootstrapping.
- Based on ***cardinality principles*** of both sequence iterations, $\binom{n}{2}$ (CM) will differ from $\binom{n}{4}$ so we should estimate the relative sampling rate and adjust by simply sampling more
- Note that the graph set $\{G_{\binom{n}{2}}\} \subseteq \{G_{\binom{n}{4}}\}$ but what we care about is the ***rate of sampling***. We want to cover enough of the former while implementing the later.





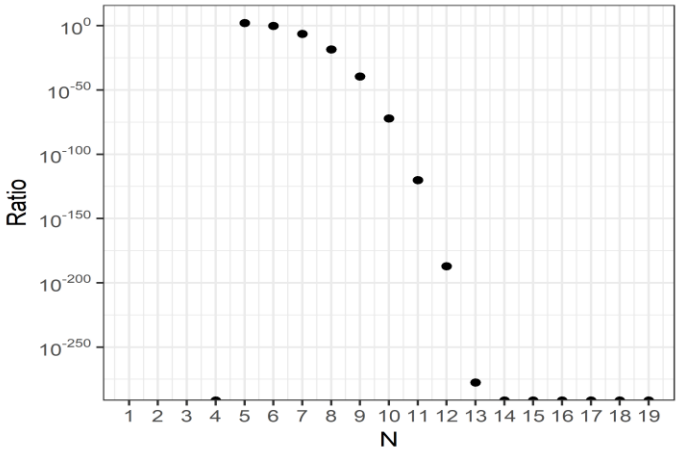
$$\{G_{\binom{n}{2}}\} \subseteq \{G_{\binom{n}{4}}\} \quad for\ n \geq 4$$

$$P\left(G_{\binom{n}{2}}\right) = p^m(1-p)^{\binom{n}{2}-m} \quad \text{Also...} \quad P\left(G_{\binom{n}{4}}\right) = p^m(1-p)^{\binom{n}{4}-m}$$

Via Newman (2010)

$$ratio = \frac{P(G_{\binom{n}{2}} \cap G_{\binom{n}{4}})}{P(G_{\binom{n}{4}})} = \frac{P(G_{\binom{n}{2}})}{P(G_{\binom{n}{4}})} = \frac{p^m(1-p)^{\binom{n}{2}-m}}{p^m(1-p)^{\binom{n}{4}-m}} = how\ much\ more\ sampling$$

$$some\ algebra + making\ use\ of\ stirling's\ approximation + \text{solving numerically}$$



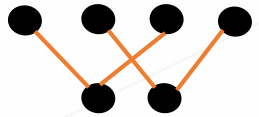
Note that what we're actually measuring is:

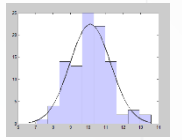
$$P\left(G_{\binom{n}{2}} | G_{\binom{n}{4}}\right) = \frac{P(G_{\binom{n}{2}} \cap G_{\binom{n}{4}})}{P(G_{\binom{n}{4}})}$$

Methodology

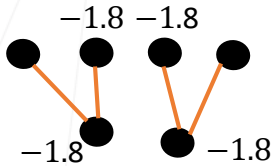
- Now, we can bootstrap appropriately and ***attribute KL divergence to a set of 4 nodes***/2 dyads/2 edges.
- We can calculate statistics on the aggregate and ***attribute entropy loss to a small number of nodes and dyads***.
- On average, we will be able to measure the information loss, or "surprise" on the whole, given the ***permutation of any centrality measure*** (or property of any node for that matter) on a fixed degree sequence—this later part is critical.
- We can identify nodes and dyads that are surprisingly important (by any statistic we choose) to the whole distribution (say centrality) ***while holding their degree constant***.
- *Comment: not sure if we should normalize by node degree or not.*

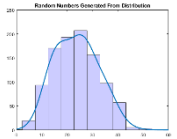
$$C(x) = \frac{N}{\sum_y d(y,x)}$$





rewire



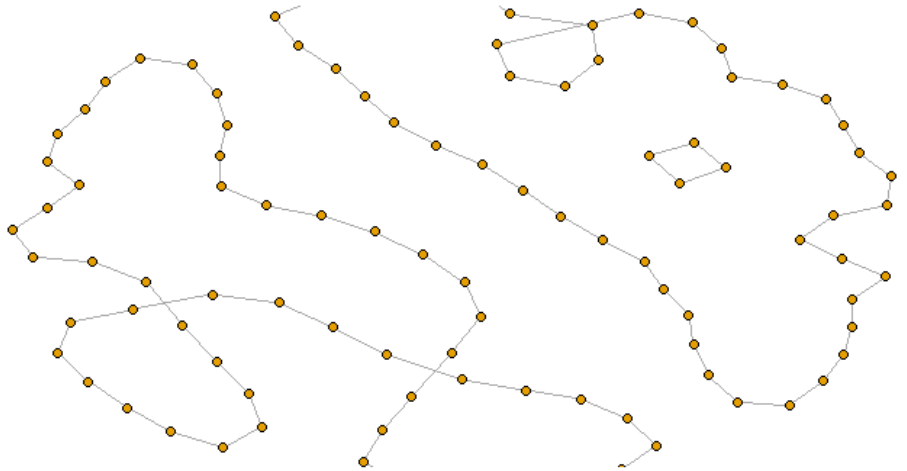


$$D_{KL}(P||Q) = \sum_{x \in \mathrm{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) = -7.2\ bits$$

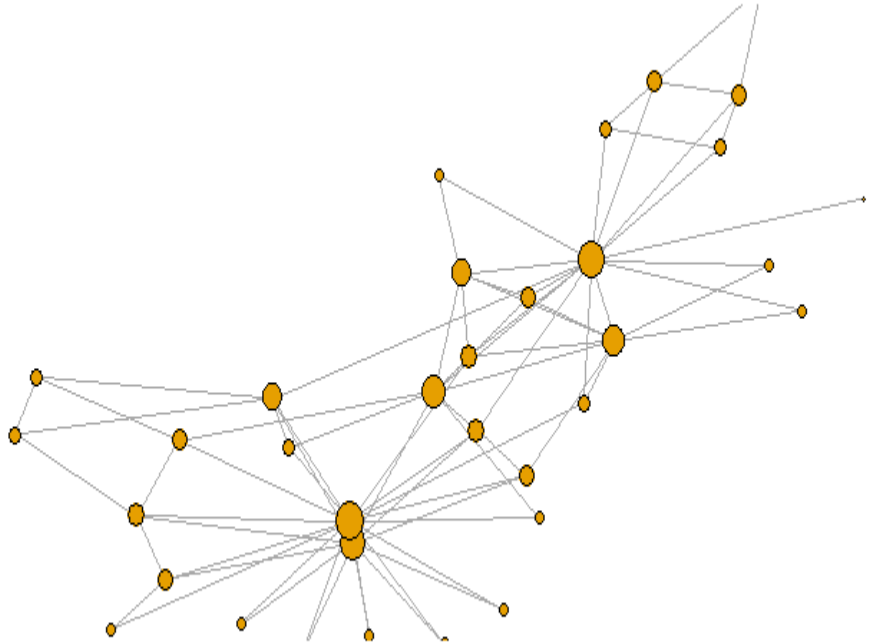
All the Pieces

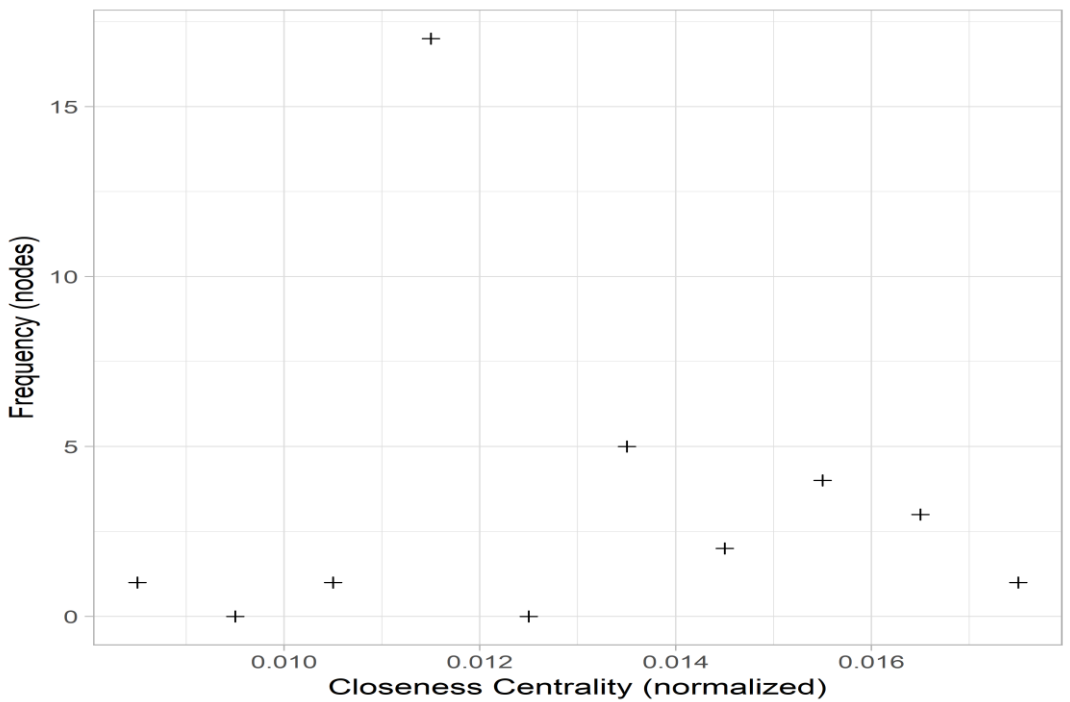
- A framework that allows for ***diverse interventions***—Targeting, hardening etc. instead of only ***node deletion***
- A meaningful scalar statistic—***Entropy***—that fits right into security and target policy-making
- An empirical framework to borrow from—the *Configuration Model*
- A ***permutation/bootstrapping*** method that covers enough sample space, and allows attribution of aggregate level properties to individual nodes
- An expandable ***Dark Networks 2.0*** framework that could easily integrate several other toolsets, including temporal, longitudinal, and agentized methods (consider permuting according to an activation rate, simulating time, and consider permuting according to some "rule").

Theoretical Test: Ring Networks

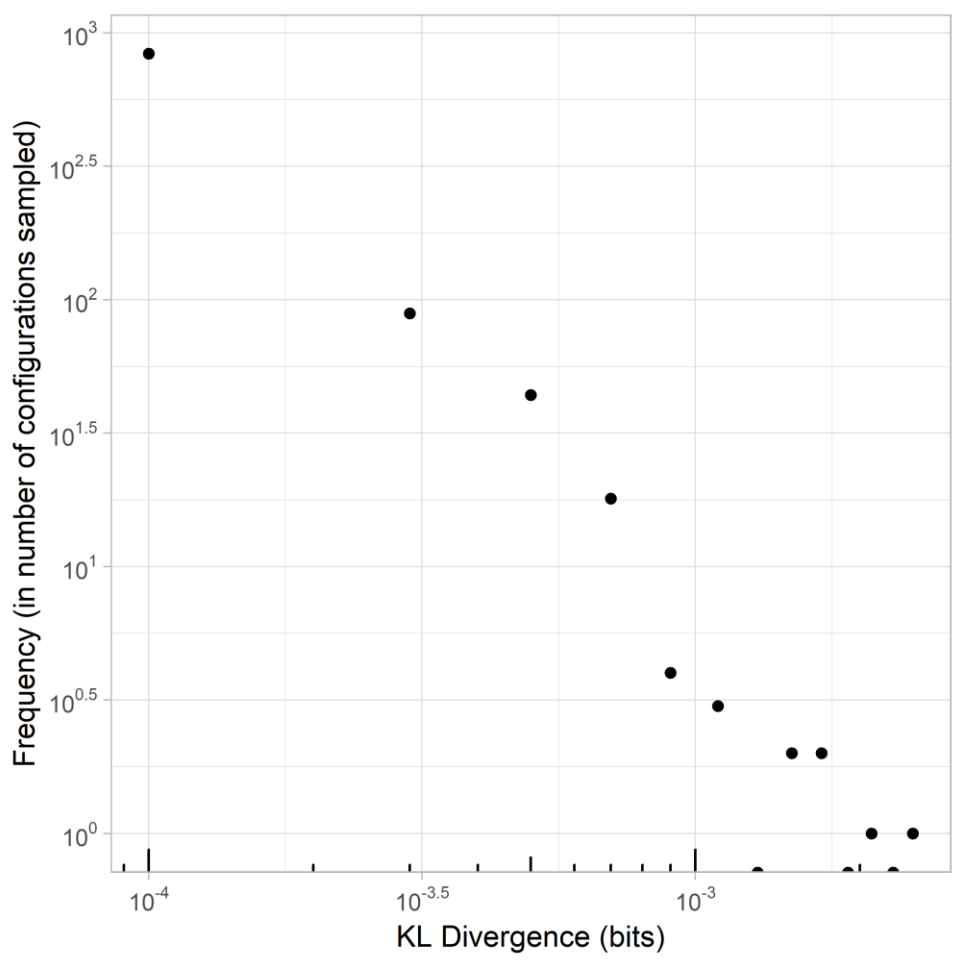


Zachary's Karate Club





Zachary's Karate Club



Is "Surprise" Power Law?

Asking for a friend!

Future Work—So much to do

- Expanding into joint entropy distributions (multinomial) allowing for multiple statistics to be considered
- Testing on a variety of real-world networks
- Validating predictive performance
- Incorporating temporal rules and behaviors into the permutation scheme (think empirical agent-based modeling!!!)

Contact me:

Cell: 703-340-0048

Email: jshaheen@gmu.edu

The Tweeeeter Site: @josephshaheen

Das Face of the Book: I don't know. Just search for a bald brown guy

Thank you to my mentors and sponsors. You guys are #awesome