Towards a Principled, Computational, and Risk-Based Perspective in Dark Networks

Joseph A.E. Shaheen

ORISE Intelligence Community Postdoc Fellow

in residence at the

Department of Computational and Data Sciences

George Mason University

Summary

- towards a re-definition of Dark Networks
- framing the issues & challenges
- the foundations of a new perspective
- a critique

The Labyrinth: A Legend

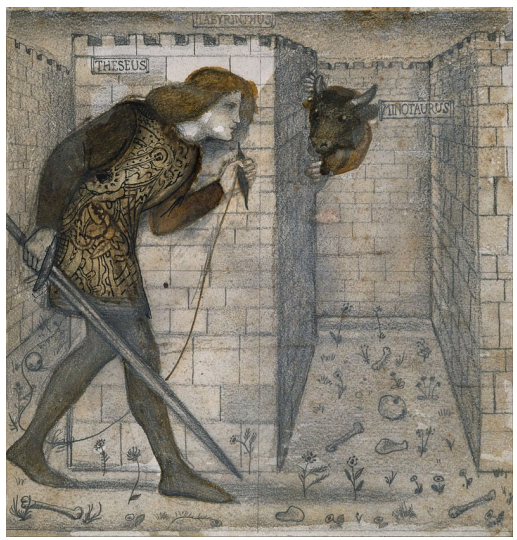
- King Minos of Crete was gifted a bull by Poseidon intended for sacrifice. In his hubris, he broke his covenant and decided to keep the bull
- As retribution Poseidon caused Pasiphae, Minos' wife to desire the bull and consequently she asked Daedalus a brilliant but mad immigrant scholar to fashion her a wooden cow with which she can mate with the bull.

The Labyrinth: A Legend

- The result was the birth of Minotaur,a being with a human body and a bull's head. Shocked, Minos requested Daedalus construct a Labyrinth so puzzling that Minotaur could never escape it, and so he did. Such a a construction he created that he himself could not understand or know its design.

The Labyrinth: A Legend

- When Theseus–a great hero-prince–was sent to Crete to be sacrificed to Minotaur, a daughter of Minos fell in love with him and implored Daedalus to show Theseus the secret of his Labyrinth. Too complex for any direct solution or a map of its inner designs, Daedalus provided a heuristic–a method–by which Theseus can escape the Labyrinth.
- Daedalus provided Theseus with a flaxen thread and instructed him to unspool it on his entrance. This thread will help him identify/find a way out so long as he follows it back.



A Lesson

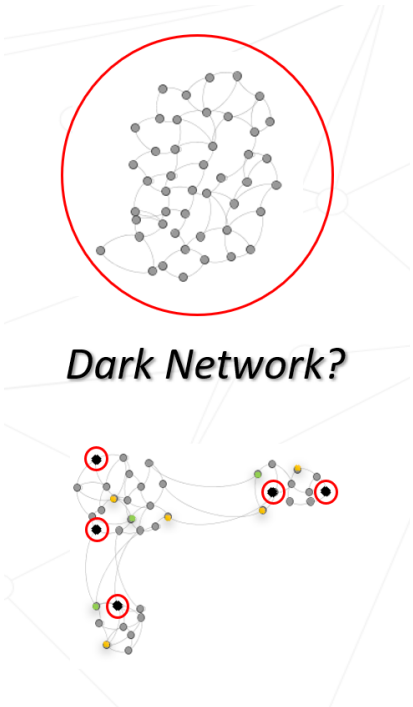
The No Model, Model: Daedalus, as brilliant as he was, created a system even **he** could not map–could not model to help Theseus. However, given the complexity of the system he was able to produce a method–a tool–to help Theseus identify **a** path.

- Public Policy $\supseteq$ Security Policy $\supseteq$ Targeting Policy, at their best, should not be concerned with the singular correct answer but with providing the correct set of policy (tactical or strategic) options. The "system"–even one that is pre-designed–is often too complex, and consequently, a single correct answer may not exist or be practical to identify.
- To provide said options, any collection of tools, frameworks, and/or methods must possess 3 attributes

They must be: principled flexible provide a measure of confidence

problems with the so-called Dark Networks framework*

- Ambiguous definition [18, 15, 16, 9, 7] : are Dark Networks networks of bad actors or just networks with bad actors embedded.
- Structural viewpoint insufficient [23, 14, 1]: Little evidence that Dark Networks are structurally discernible
- Lacking (or incorrect) confidence reports [13, 19, 12, 3, 5, 8]
- Relies on explanatory models without inference capability [10, 4, 21, 22]



*citations are critiqued works, not supporting evidence

Centrality Measures

- Degree Centrality $C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$ for undirected graphs with $C_D^*(i) = \dfrac{\sum_j A_{ji}}{n-1}$ *normalized* $\in [0, 1]$
- Closeness Centrality $C_D^*(i) = \dfrac{n-1}{\sum_j d(i,j)}$ *normalized* $\in [0, 1]$
- Betweenness Centrality *normalized* $\in [0, 1]$

$$C_B^*(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

- Eigenvector Centrality given by $\upsilon_i = \frac{1}{\lambda} \sum_j A_{ij} \upsilon_j$ such that $A\upsilon = \lambda\upsilon$, & choosing the eigenvector associated with the largest eigenvalue. Consequently, normalized by maximum value $\in [0, 1]$ by the factor $\sum_i \upsilon_{max} - \upsilon_i$

Centrality Measures (extended)

- Katz Centrality [11]

$$C_{katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (A^k)_{ji}$$

- PageRank [17]

$$PR(i) = \frac{1-d}{n} + d \sum_{j \in M(i)} \frac{PR(j)}{C_D^{out}(j)}$$

- Subgraph Centrality [6]

$$C_{sub}(i) = \sum_{j=1}^{N} (v_j^i)^2 e^{\lambda_j}$$

- Information Centrality [20]

$$C_I(i) = \frac{n}{\sum_{j=1}^{n} \frac{1}{I_{ij}}}$$

Case Study: Centrality without confidence & Key Player algorithm

Consider: [2] identifies a heuristic that can maximally fragment networks given that node "neutralization" occurs, dubbed **Key Player**.

- Is fragmentation (always) the correct answer? Case: AlQaeda in Iraq $\Rightarrow$ ISIS
- A heuristic measure that reports no level of confidence: What is the distribution of similar networks where the set of key players proposes results in the fragmentation of an observed network?

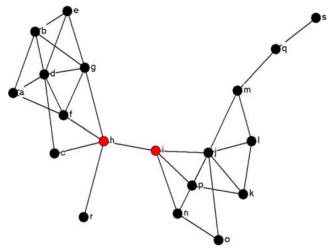


Fig. 3 Network in which removing the two most central nodes ("*i*" and "*j*") is not as disruptive as removing a different pair of nodes ("*i*" and "*j*")

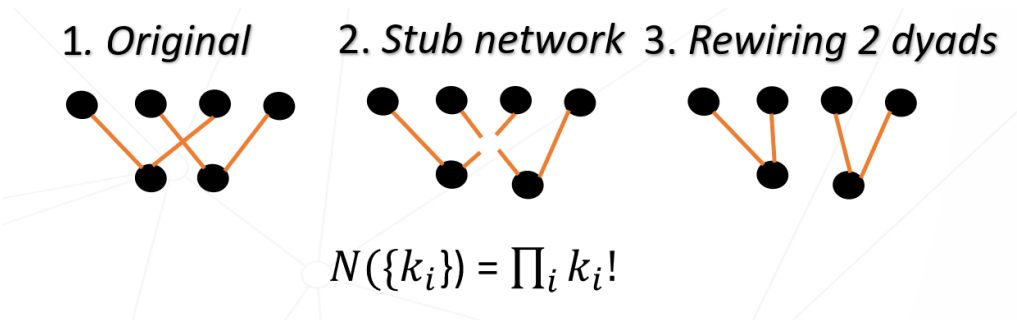
Misuse of Social Network Analysis

Observed networks are realizations from an ensemble of random networks each of which equi-probable. The best it can offer is a risk management perspective.

Generating an ensemble: degree-preserving rewiring

Many methods of generating ensembles of random networks given an observed network. The key is to first choose an **invariate quantity** and **computationally** generate it

- observe the invariate
- explore the probability space
- calculate quantities of interest
- calculate statistic of interest



Risk Perspective: Information & Mathematical Surprise (Entropy Divergence)

- Information (Shannon) Theory offers a natural mathematical framework for risk analysis of Dark Network centrality distributions
- Specifically Information Divergence (Kullback-Leibler Divergence)
- $$D_{kl}(P||Q) = \sum_{x \in \chi} P(x) log\left(\frac{P(x)}{Q(x)}\right)$$
- How much does one quanity diverge from another quantity (probability distribution)

Information Divergence

How much will I (the analyst) be surprised if this node of interest had a different centrality measure given its random network ensemble under the constraints of my rewiring/invariate?

Combining Centrality Measures: Letting the Analyst Choose

Because we can generate our network ensemble and thus compute the centrality measures of our ensemble for each object we can generate a vector of centrality information divergences for each node. Reducing that to a mean value for each centrality measure, a concurrent comparison can take place and now we can use multivariate analysis

- Our data is now a collection of independent observations (from a comparison to the ensemble of generated networks)
- Anything that can be computed with independent and identically distributed datasets can now be computed for our centralities' information divergences without additional assumptions

$$\begin{pmatrix} D_{kl}(C_1) \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ D_{kl}(C_p) \end{pmatrix}$$

The Multivariable Case: Regular Equivalence & Structural Equivalence

Defining New Quantities

Now we have a vector of centrality information divergences (suprise quantities), and we can being to explore simple mathematics to understand our new framework. For example, how would the notion of structural and regular equivlance be described in this framework?

- One answer: Cosine Similarity

$$\cos(\theta) = \frac{A \cdot B}{||A||||B||}$$

- Here A and B are a vector of centrality information divergences for node a and node b.
- $\cos(\theta)$ is a measure between 0 and 1 describing their similarity - hence the notion of regular equivalence

The Multivariable Case: Outliers & Mahalanobis Distance

Defining New Quantities

We can also compute interesting measures that were not easily computed prior. In some cases identification of node outliers may be of interest. In this case we can use distance-grouping measures such as the Mahalanobis Distance which identified outliers based on data clustering

$$D_M(\vec{x}) = (\vec{x} - \vec{y})' \mathbf{C}^{-1} (\vec{x} - \vec{y}))^{\frac{1}{2}}$$

What else?

...many other measures

we said

They must be: principled flexible provide a measure of confidence

but...

Is it useful?

Public and Security Policy

- public policy should not be concerned with the correct answer, nor was it ever intended to
- the golden age of public and security policy occurred before most significant advances in technology
- public and thus security policy should and is concerned with the correct ensemble of options; method should reflect that
- public policy scholars should be concerned with the study of an ensemble of choices

Bibliography

[1] John Bohannon. Investigating Networks: The Dark Side. *Science*, 325(July):410–411, 2009.

[2] Stephen P. Borgatti. Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, 12(1):21–34, apr 2006.

[3] David A. Bright, Caitlin E. Hughes, and Jenny Chalmers. Illuminating dark networks: A social network analysis of an Australian drug trafficking syndicate. *Crime, Law and Social Change*, 57(2):151–176, 2012.

[4] Christopher Couch, William P. Fox, and Sean F. Everton. Mathematical modeling and analysis of a dark money network. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 13(3):343–354, jul 2016.

[5] Fatih Demiroz and Naim Kapucu. Anatomy of a dark network: The case of the Turkish Ergenekon terrorist organization. *Trends in Organized Crime*, 15(4):271–295, 2012.

[6] Ernesto Estrada and Juan A. Rodríguez-Velázquez. Subgraph centrality in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(5):1–9, 2005.

Bibliography

[7] Sean F. Everton. *Tracking, Destabilizing and Disrupting Dark Networks with Social Networks Analysis*. Cambridge University Press, 2012.

[8] Sean F. Everton and Dan Cunningham. Detecting significant changes in dark networks. *Behavioral Sciences of Terrorism and Political Aggression*, 5(2):94–114, 2013.

[9] Audrey Heffron, Jarrett Broder, and Brad Skillman. Organizational De-Evolution ; the Small Group or Single Actor Terrorist. *World Academy of Science, Engineering and Technology International Journal of Information and Communication Engineering Vol:6,*, 6(4):33–36, 2012.

[10] Rouslan Karimov and Luke J. Matthews. A simulation assessment of methods to infer cultural transmission on dark networks. *Journal of Defense Modeling and Simulation*, 14(1):7–16, jan 2017.

[11] Leo Katz. A New Status Index Derived From Sociometric Analysis. *Psychmetrika*, 18(1):39–43, 1953.

[12] Brian Keegan, Muhammad Aurangzeb Ahmad, Dmitri Williams, Jaideep Srivastava, and Noshir S Contractor. Dark Gold: Statistical Properties of Clandestine Networks in Massively Multiplayer Online Games. *IEEE International Conference on Social Computing*, pages 201–208, 2010.

Bibliography

[13] Jared P. Keller, Kevin C. Desouza, and Yuan Lin. Dismantling terrorist networks: Evaluating strategic options using agent-based modeling. *Technological Forecasting and Social Change*, 77(7):1014–1036, 2010.

[14] Yoshiharu Maeno. Node discovery problem for a social network. *arXiv preprint*, pages 1–20, 2009.

[15] Brint Milward. Dark Networks as Organizational Problems. 2006.

[16] H. Brinton Milward and Jörg Raab. Dark networks as organizational problems: Elements of a theory. *International Public Management Journal*, 9(3):333–360, 2006.

[17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank Citation Ranking: Bringing Order to the Web. *World Wide Web Internet And Web Information Systems*, 54(1999-66):1–17, 1998.

[18] Jorg Raab and H Brinton Milward. Dark Networks as Problems. *Journal of Public Administration Research and Theory*, 13(4):413–439, 2003.

[19] Christian Robert and George Casella. *Introducing Monte Carlo Methods with R*. Number 2007. 2010.

Bibliography

[20] Karen Stephenson and Marvin Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37, 1989.

[21] Maksim Tsvetovat and Kathleen M. Carley. Generation of Realistic Social Network Datasets For Testing of Analysis and Simulation Tools. *SSRN*, 2016.

[22] Maksim Tsvetovat, Jana Diesner, and Kathleen M. Carley. Netintel: A Database for Manipulation of Rich Social Network Data. *Ssrn*, 2016.

[23] Jennifer Xu and Hsinchun Chen. The topology of dark networks. *Communications of the ACM*, 51(10):58, 2008.

Thank You

Thank you

Special thanks to ORISE, ODNI, NCTC, GMU

contact at **jshaheen@gmu.edu**