

ATria: A Novel Centrality Algorithm Applied To Biological Networks

Trevor Cickovski and Eli Peake

Department of Computer Science

Eckerd College

St. Petersburg, FL 33711, USA

Email: {cickovtm, eapeake}@eckerd.edu

Vanessa Aguiar-Pulido and Giri Narasimhan

Bioinformatics Research Group (BioRG)

& Biomolecular Sciences Institute

School of Computing & Information Sciences

Florida International University

Miami, FL 33199, USA

Email: {vaguarp, giri}@fiu.edu

Large-scale biological networks such as gene regulatory [1] and PPI [2] have become commonplace through systems biology. Figure 1 shows a *microbial* social network [3], which attempts to infer interactions between microbes within a community from metagenomics studies. In particular this is a co-occurrence network [4], where green/red edges respectively represent positive/negative *correlations*, or how strongly two bacterial taxa tend to co-infect samples.

We may be interested in bacterial taxa that drive community behavior, initial infectors of a community, or the effects of structural changes. Answers to these questions can come from discovering "important" nodes in these biological networks [5], which is the goal of social network *centrality*. Depending on the definition of *importance* used, many different centrality notions currently exist [6]. We suggest three notions that are potentially important to biological networks, and especially to microbial social networks:

- 1) For each *club* or high density subgraph, a dominant or *leader* node [7] that is responsible for connecting many individuals and driving club behavior.
- 2) A *villain* node or common enemy [8], defined as having many strong negative edges to a club.
- 3) A *bridge* node that connects two or more clubs, thus having the ability to link different social circles [8].

Current centrality approaches do not readily generalize to networks with *negative* edge weights, making it difficult to distinguish a villain node vs. a node with few connections. They also tend to produce a list of the most central nodes that is clustered in a specific region of the network, finding only limited amounts of leaders and bridges, if any. We address both shortcomings by proposing *ATria*, an iterative centrality algorithm that combines aspects of economic theory, social network theory, and path-based algorithms. We iteratively remove a node with the highest centrality along with some of its neighborhood edges before finding the node with the next highest centrality, using social network theory to determine the appropriate edges to remove. We verify that *ATria* finds leaders, villains and bridges within signed, weighted social networks. Our test networks include synthetic, scale-free [9], gene expression, PPI, and microbial social networks (Figure 1).

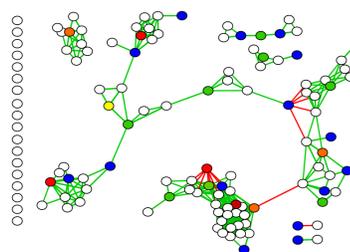


Figure 1. Results of *ATria* on a sample co-occurrence network. Red nodes are most central, blue the least.

The authors acknowledge support from the Florida Department of Health (FDOH 09KW-10) and the Alpha-One Foundation, the College of Engineering and Computing at Florida International University and a Faculty Development Grant from Eckerd College.

References

- [1] E. Davidson and M. Levin, "Gene regulatory networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 14, p. 4935, 2005.
- [2] B. Schwikowski, P. Uetz, and S. Fields, "A network of protein-protein interactions in yeast," *Nat Biotech*, vol. 18, no. 12, pp. 1257–1261, 2000.
- [3] J. Ackerman, "The ultimate social network," *Scientific American*, vol. 306, no. 6, pp. 36–43, 2012.
- [4] P.-J. Kim and N. D. Price, "Genetic Co-Occurrence Network across Sequenced Microbes," *PLoS Comput Biol*, vol. 7, no. 12, pp. e1002340+, Dec. 2011.
- [5] D. Koschutzki and F. Schreiber, "Centrality analysis methods for biological networks and their application to gene regulatory networks," *Gene Regul Syst Bio.*, vol. 2, pp. 193–201, 2008.
- [6] P. Bonacich, "Power and centrality: A family of measures," *American Journal of Sociology*, vol. 92, pp. 1170–1182, 1987.
- [7] Q. Wu, X. Qi, E. Fuller, and C.-Q. Zhang, "Follow the leader: A centrality guided clustering and its application to social network analysis," *Scientific World Journal*, vol. 2013, pp. 1–9, 2013.
- [8] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
- [9] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.