Modular Communities in Infectious Disease Networks

S. Chatterjee†, B.S. Sanjeev†

†Department of Applied Sciences, Indian Institute of Information Technology, Allahabad

Preliminaries

In graph theory and network analysis, we make use of several indicators of centrality to identify the most important vertices within a graph [1, 2]. Herein, for our undirected graph $\mathcal{G}$, such that $\mathcal{G}_p = (\mathcal{V}_p, \mathcal{E}_p)$,

- $\mathcal{V}_p$ is the set of nodes depicting $\mathcal{H}_p$ (Human genes) and $\mathcal{P}$ (Pathogens).
- $\mathcal{E}_p$ is the set of corresponding edges (interactions) between them wherein $\mathcal{E}_p \subseteq [\mathcal{V}_p]^2$.

Centrality Indicies

As Figure 1 indicates, the betweenness centrality [3] of a node v is given by the expression,

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (1)$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

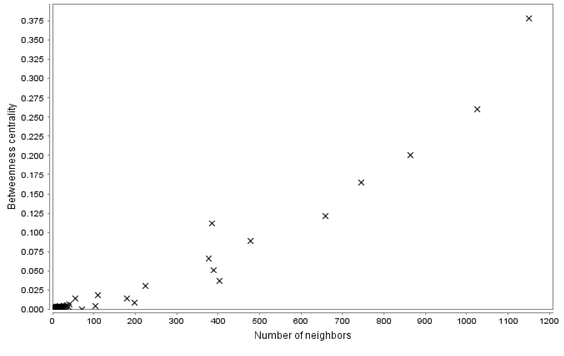

Figure 1: Nodes with higher betweenness centrality index are often important controllers of power of information, because more information tends to pass through that node.

- Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v.
- Nodes that tend to have high betwenness centrality are expected to lie in between many shortest paths and exhibit that even low-degree nodes with high betweenness may reveal a modular network structure [4].

Network Overview

After analyzing the pathogen-host interactome [5], we could pick out the important interacting proteins (IIPs) in the human host that acted as hubs in the pathogen-host interactive graph and subsequently vulnerable to the pathogenic proteins. It may be imperative to identify the communities in networks since the communities may have quite different properties such as node degree, clustering coefficient, betweenness, centrality. etc., from that of the average network.

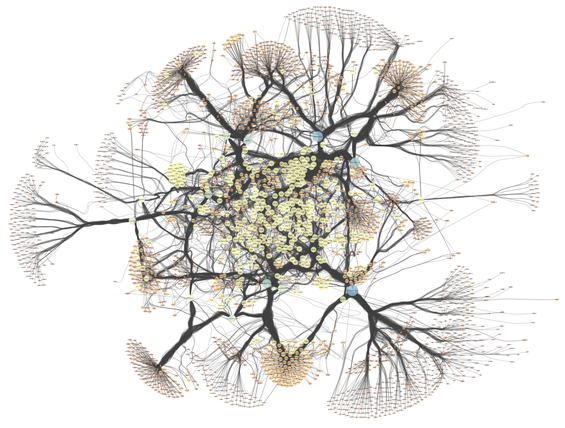

Figure 2: The Pathogen-Human Gene Network.

The above Figure 2. is a Bipartite Graph $\mathcal{G}$ and the nodes can be divided into two disjoint sets $\mathcal{U}$ and $\mathcal{V}$.

$$P_k \sim k^{-\gamma} \qquad (2)$$

- Each link connects a $\mathcal{U}$-node to a $\mathcal{V}$-node.
- For most scale-free networks the degree exponent γ is between 2 and 3.
- The lack of an internal scale, denotes the fact that nodes with widely different degrees coexist in the same network.
- It is important to detect communities and also to study how they affect the spreading processes in various settings.
- Modularity is one such measure, which when maximized, leads to the appearance of communities in a given network.

Network Modularity

A module is defined as a sub-graph in a network that is more internally connected within itself than the other nodes in the network.

Entities	#
Total no. of pathogens	17
Total no. of associated diseases	307
Total no. of human genes (H_p)	2,536
Total no. of interactions	7,019

Table 1: Overview of Pathogen-Host Network

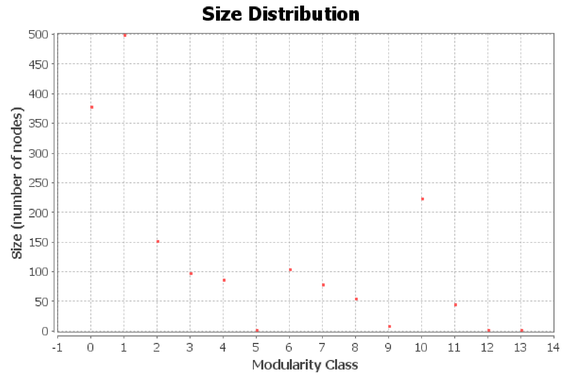

Figure 3: The distribution of a number of communities (clusters) of various size (number of nodes) along with their modularity.

- Modular communities can be mathematically represented and topologically studied to reveal some unexpected structural features.
- As pertinent from Figure 1, the size of the hubs grows polynomially with network size, in scale-free [6] networks, in contrast to a random network wherein, the size of the largest node grows logarithmically.
- A closely connected community in a biological network as depicted in Figure 2 will imply a faster rate of transmission of information or critical instruction among them than a loosely connected community.
- We observed the Network Modularity $\mathcal{Q} = 0.537$ of the observed 14 communities in our pathogen-host network.

Conclusions and Outlook

We have analyzed 17 pathogenic species mainly belonging to bacterial and viral pathogenic classes, with the aim to identify the interacting human proteins which are targeted by both bacteria and virus specifically. Our study reveals that the TLRs play a crucial role between the pathogen-associated molecular patterns (PAMPs) and the damage associated molecular patterns (DAMPs). PAMPs include bacterial lipopolysaccharides (LPs), endotoxins, bacterial flagellin, lipoteichoic acid, peptidoglycan in bacterial organisms and nucleic acid variants associated with viral organisms, whereas DAMPs are associated with host biomolecules that perpetuate non-infectious inflammatory responses. We found out the presence of anti-oxidizing agents helps in eliminating oxidative stress by preventing damage from reactive oxygen species. Hence, such strategies can be used as new therapies for anti-inflammatory diseases with significant clinical outcomes. Interestingly, such kind of approaches can be exploited to study common origin of many diseases and linkages between them.

References

[1] P. Erdös and A. Rényi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., **5**, 17-61, (1960).

[2] D. J. Watts, S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature, **393**, 409–10, (1998).

[3] U. Brandes. A Faster Algorithm for Betweenness Centrality. Journal of Mathematical Sociology, 25(2):163-177 (2001).

[4] M. Newman, A.-L. Barabási, and D. J. Watts. The Structure and Dynamics of Networks. Princeton University Press, (2006).

[5] S. Chatterjee, B. S. Sanjeev, IEEE DOI:10.1109/BSB.2016.7552164 pp. 1-4, (2016).

[6] A.-L. Barabási, H. Jeong, R. Albert. Mean-field theory for scale free random networks. Physica A, **272** 173-187, (1999).