Inter-connectivity leads to vulnerability in biological networks

S. Chatterjee†, B.S. Sanjeev†

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

The Big Question ?

Is diversity prompting us to call the scale-free property a *universal* network characteristic ?

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 all nodes (proteins) of $\mathcal{H}_p$ (Human) and $\mathcal{P}_p$ (Pathogenic) proteins.
- $\mathcal{E}_p$ is the set of corresponding edges (interactions) wherein $\mathcal{E}_p \subseteq [\mathcal{V}_p]^2$.

Degree Distribution

As Figure 1 indicates, the Poisson form offers a poor fit for the real world networks. Instead on a log-log scale the data points form an approximate straight line, suggesting that the degree distribution [5] is well approximated with,

$$P_k \sim k^{-\gamma} \tag{1}$$

Equation 1 is called a power law distribution and the exponent γ is its degree exponent. If we take a logarithm of Eq. 1, we obtain,

$$\log P_k \sim -\gamma \log k \tag{2}$$

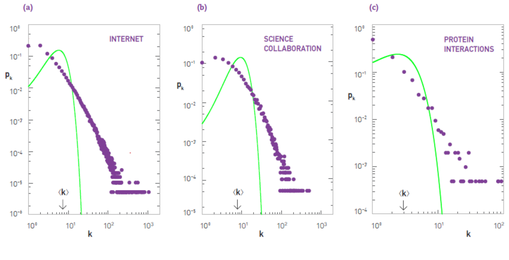

Figure 1: The degree distribution of the (a) Internet, (b) Science collaboration network, and (c) Protein interaction network. The green line corresponds to the Poisson prediction, obtained by measuring $< k >$ for the real network.

Network Analysis

After analyzing the pathogen-host interactome dataset [6], 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.

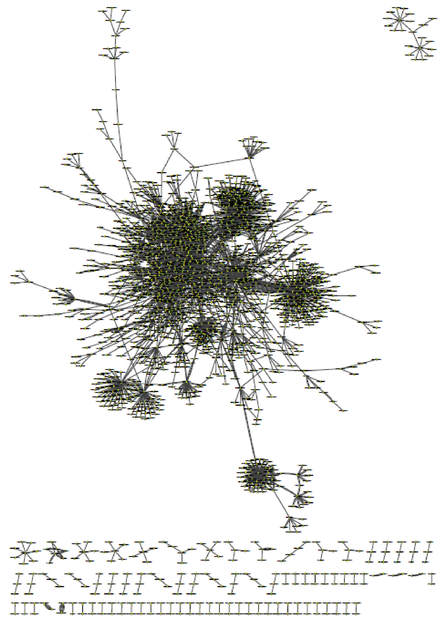

Figure 2: The Pathogen-Host Interacting Network.

Is there a Universality ?

Indeed, the WWW is a man-made network with a history of little more than two decades, while the protein interaction network is the product of four billion years of evolution. The diversity of the systems that share the scale-free property whether they are biological entities, or computers is remarkable.

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

- 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.
- Most nodes have comparable degrees in a random network [3] and hence *hubs* are forbidden.
- As clear from the Figure 3, the size of the hubs grows polynomially with network size, in scale-free [4] networks, in contrast to a random network wherein, the size of the largest node grows logarithmically.

Network Overview

We could list out 9 highly interacting hubs present in the human host that were involved with 560 pathogenic proteins.

Entities	#
Total no. of pathogens	182
No. of human proteins (H_p)	668
No. of pathogen proteins (P_p)	1,188
Total no. of proteins (H_p) + (P_p)	1,856
Total no. of interactions	3,905

Table 1: Overview of PHI Networks

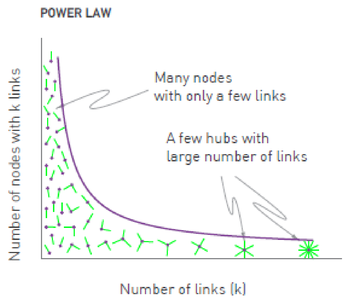

Figure 3: In a network with a power-law degree distribution most nodes have only a few links. These numerous small nodes are held together by a few highly connected hubs.

Conclusions and Outlook

Herein, we have emphasized only on those human protein / protein coding genes that were predominantly interacting with 5 or more viral pathogenic proteins and also subsequently with 2 or more bacterial proteins. We have identified the proteins / protein coding genes that were more susceptible to both bacterial and viral proteins. These proteins may well be targeted and act as potential drug targets. Our analysis of the pathogen-host interactions (PHIs) to identify the important interacting proteins (IIPs) can be practically applied over to other infectious pathogens and protein-protein interactive networks (PPINs) as well. The predominant proteins / protein coding genes are namely P53 (Tumor protein p53), NFκB1 (Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), GBLP (Guanine nucleotide-binding protein subunit beta-2-like-1), TOX4 (TOX high mobility group box family member 4), PDIA1 (Protein disulfide-isomerase precursor), MHY9 (Myosin 9), RAC1 (Ras-related C3 botulinum toxin substrate 1), CCAR2 (Cell cycle and apoptosis regulator protein 2) and ILF3 (Interleukin enhancer binding factor 3).

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] B. Bollobás. Random Graphs. Cambridge University Press, (2001).

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

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

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