ABSTRACT. Fifth International Conference on Guided Self Organization (GS05).
Striking absence of modularity in large-scale, evolving, self-organized peer to peer networks
Somwrita Sarkar[1,2]
[1]Complex Systems Group, School of Physics, University of Sydney, Australia 2006
[2]Faculty of Engineering and Information Technology, University of Sydney, Australia 2006.

Introduction. Peer-to-peer or p2p networks are decentralized, self organized systems, in which individual computers connect to each other and communicate directly for the purposes of sharing information and resources, without dedicated or centralized servers [1]. Though these systems are guided by common goals (for example, of sharing CPU cycles and storage space), there is no central guiding authority. The resulting network topology and the dynamics of communication occurring on it are emergent; i.e., individual users interacting locally with other users determine the local decisions, but the large scale system behavior cannot be determined trivially from the local interactions alone. The highly decentralized self organized nature of these evolving networks ensures large fluctuations in network size and numbers of edges, as the size and resulting topology of the network are completely determined by how many individual users are joining and leaving the network. Since many self organized systems in nature and society show modular organization, we were interested in looking at the modularity properties of large scale evolving peer to peer networks, and to chart how modular organization of a guided self-organized system evolves dynamically over time.

Methods. To explore the modular organization of these networks, we explore the eigenvalue spectra of temporal snapshots of the peer-to-peer Gnutella file sharing network [data from (2)]. The data represents a sequence of 9 snapshots of the p2p Gnutella network, collected in August 2002 (from 4 Aug 2002 [6301 nodes, 20777 edges] to 31 Aug 2002 [62586 nodes, 147892 edges]), with the nodes representing hosts and edges representing connections between hosts. From spectral graph theory, we use the idea that the spectrum of a graph encodes important information on the structure of the graph, including its modular organization [3-6]. Specifically, we use the property that a non-modular random or scale free graph shows a distinctly different spectral fingerprint than their modular counterparts [7,8].

Results and Discussion. The results show a striking absence of large modules in all the 9 networks. The eigenvalue spectra show only one large eigenvalue well separated from the eigenvalue cloud, distinctly showing that there is no significant modularity present in the network. If there is any modularity present, we surmise that it is local; i.e., the size of the module is insignificant as compared to the size of the system, and that the modularity is very weak so as to be rendered undetectable by the spectral approach. As opposed to this signature, if there was any significant modularity present, the spectrum would have shown more than one eigenvalue well separated from the eigenvalue cloud. We also extracted smaller parts of the large scale networks to check if the spectra of these smaller networks show modularity. Remarkably, these smaller networks also showed no significant modularity. At no level of viewing the network were we able to see any significant presence of modularity. This was a very surprising result. Many self-organized systems in nature and society are known to be modular, given that modularity imparts several performance and evolutionary benefits over non-modular systems. Moreover, one principle driving modularity could be that users on the p2p network are likely to have specific file sharing or information needs and exercise freedom in connecting to other users. Thus, an expected trend could be that modularity emerges in the network, even with the possibility that it is transient. Thus, it is remarkable that the evolving Gnutella p2p network (at least over a month of observations) shows a distinct absence of modularity. This finding has implications for p2p system design and performance. We note that the non-scalability of existing p2p Gnutella architecture, its reported mismatch with the underlying Internet topology, and new strategies for designing scalable and robust p2p systems has been the topic of much research [1]. Our work shows that these can be related to the finding that existing self-organized p2p systems appear to be non-modular.

References.
[1] M. Ripeanu, I. Foster and A. Iamnitchi (2002) Mapping the Gnutella Network: Properties of Large Scale Peer to Peer systems and implications for system design, IEEE Internet Computing, 6(1), pp. 50-57.
[2] The Stanford Large Network Dataset Collection at http://snap.stanford.edu/data/index.html, last viewed 23 July 2012.
[3] Cvetkovic D, Doob M, and Sachs H (1995) Spectra of Graphs: Theory and Applications. Johann Ambrosius Barth Verlag.
[4] Cvetkovic D, Rowlinson P, and Simic S (1997) Eigenspaces of Graphs, Volume 66, Encyclopedia of Mathematics and Its Applications. Cambridge University Press.
[5] Chauhan S, Girvan M, and Otto E (2009) Spectral properties of networks with community structure, Phys Rev E 80: 056114.
[6] Farkas I, Derenyi I, Barabasi A, and Vicsek T (2001) Spectra of “real-world" graphs: Beyond the semicircle law. Phys Rev E 64: 026704.
[7] Goh KI, Kahng B, and Kim D (2001) Spectra and eigenvectors of scale-free networks. Phys Rev E 64: 051903.
[8] Sarkar S, Dong A (2011) Community detection in graphs using singular value decomposition, Phys. Rev. E 83: 046114; Sarkar, S., Henderson, J.A., and Robinson, P.A., 2012, Spectral fingerprints of modularity and hierarchy in networks, in prep. for PLOS One.