CENTRALITY METRICS FOR IDENTIFYING NETWORK FRAGILITY IN PROTEIN-PROTEIN INTERACTION NETWORKS AND SYNTHESIZED NK SYSTEMS

Ken A. Hawick
Department of Computer Science, University of Hull, Cottingham Road, Hull HU6 7RX, UK.
Email: k.a.hawick@hull.ac.uk, Tel: +44 01482 465181 Fax: +44 01482 466666

ABSTRACT

Computational graph analysis metrics can be used to make quantitative comparisons between biological and regulatory networks obtained from real specimens with simulated synthetic networks that have well parameterized properties such as a scale-free structure. We compute the betweenness centrality metric for five public domain protein-protein network data sets and compare them with synthesized NK networks. We employ a node-culling procedure to progressively remove the highest connected nodes in these networks and assess the wholistic system changes as revealed by the resulting Floyd all-pairs distance and the number of component clusters in the networks as they fail and break up. We discuss the potential for this method in assigning characteristic signatures or categories to networks of this sort as well as for identifying network components that are most vulnerable to biological attack.

KEY WORDS

Betweenness centrality; node vulnerability; regulatory networks; computational biology; node culling; simulation.

1 Introduction

The concept of a centrality metric [26] for characterizing the nodes in a network is not new and various centrality measures [27] have been devised over the last forty years. The betweenness centrality metric [46] is relatively recent and being of particular interest is discussed in Section 3 below. This metric helps identify the most central or important node in a given network - and therefore the node whose removal or failure will have the greatest consequences for the system as a whole.

Recent work has explored this metric [44] for applications including social [22] networks, communications networks [68], power systems [37, 38], and water distribution networks [39]. It does not appear to have been uptaken widely in the biological and medical communities yet however.

Biologists and medical engineers are finding the use of computational graph analysis algorithms of growing importance for the classification and understanding of bio-networks [11]. A number of biological problems [6, 7, 23] can be posed in terms of graph and combinatorics [4]. While some of these can be analyzed using closed form mathematics, many interesting biological data sets are available in the form of simple or weighted [5, 69] graph or network structures and are too large and complex to be analyzed except using numerical or enumerative graph methods [20]. Some examples include genetic transcription networks and other experimentally obtained bio-net data such as: interaction networks [9]; conformation space networks for protein folding [59, 60, 78]; peptide folding applications [10, 65]; and metabolic graphs [61, 77].

Small-sized networks can be studied using visual rendering techniques [35, 74] and an intuitive understanding is often easily grasped. Larger networks are too hard to understand from a simple rendering and may not in fact be easily rendered in a simple space at all. There is usually far too much complexity to be able to interpret the highly connected core of nodes visually. In such cases it is important to have a battery of graph metrics [36] and to use them to categorize and identify the major properties of an unknown network by comparing its metrics signature to those of simpler or known network instances. A divide and conquer approach is also useful, whereby a large and complex network might be broken down into sub-graphs or components [43] with a well-defined or known set of properties.

A number of network analysis tools and techniques for determining properties from such large scale bio-networks have emerged in recent years and have been reported in the literature [51, 57] with ongoing efforts to formulate new approaches [50] still appearing. Properties such as numbers and connectivity distributions of vertices; arcs; input and output ratios; and maxima and minima of such connectivities are easily computed. An important finding from such analyses is that typically not all vertices in a graph are equally important [41] and some are often many times more important. Identifying these hubs or pivot nodes without recourse to visualization can aid the application analysis considerably.

More sophisticated properties such as the number of disparate clusters of nodes in graphs that are not fully connected, or the statistical pathway properties [18, 25] within fully connected clusters both involve more computationally demanding calculations but are still quite feasible even for large networks. Other attractive quantities such as the number of loops or circuits are impractical to compute for all but small networks. The methods are usually formulated in systematic and explicit searches and enumerative counting.

DOI: 10.2316/J.2014.216.818-0057

Recent work has considered the community structure of a network [12, 29, 56]. Communities in networks [40] are closely related to the notions of motif patterns [21] and cliques [62] and are closely-coupled nodes that can be detected even in networks which are completely connected. Localized properties such as the clustering coefficient can be computed for a cluster or a whole graph using explicit techniques.

Some work has considered the spectral properties of networks [33] where the quantities of interest are formed using matrices. Linear algebra and other manipulations on matrix derivations from a network's adjacency matrix are used to compute approximate groupings into communities or modules [52] of highly interconnected vertices. Such methods [34] are limited by the size of matrix that can practically be manipulated. In most cases there are practical memory limitations imposed that are more likely to be more severe than compute time limits using typical present generation processing systems.

For graphs and networks of a particular sort of structure including "small-world" networks, scale-free properties are often well characterized by metrics like the clustering coefficient [55]. The clustering coefficient [63] and various alternatives to it [1] have found use in the study of biological networks such as metabolic networks [53]. The Newman clustering coefficient is effectively the fraction of transitive triples or triangles present in the network and is considered in Section 5.

A number of authors have reported recent progress in analyzing the complexity [13] and robustness [15] of biological network data. A particular area of interest is in establishing good network analyses and properties to aid in the matching [79], querying [24, 45] and data mining [64] of biological network data-sets. Other important work has focused on discovery-specific linkages within bio nets [66] through techniques such as data slicing [76] and also on categorizing the typically occurring network architectures [72].

In a recent work Hawick explored distances, component labelling and other static metrics that could be applied to protein-protein networks [73]. In this present paper we further examine some public domain biological network data sets and use the betweenness centrality metric to compare them with synthetic NK networks.

Although metrics such as the betweenness centrality metrics are well known and have been applied to rank individual nodes in biological data sets, to our knowledge none of the existing network analysis packages [31, 54, 67] support the systemic node-culling procedure we describe here. Our main contribution in this present work is to show how this technique can reveal the fragility and systemic break points in a biological network.

Our article is structured as follows: In Section 2 we summarize some of the protein-protein interaction network data we have used in this study along with references to the sources. We summarize the graph network terminology we employ in Section 3 and describe the betweenness centrality metric. We give a overview of the NK synthetic network model that has been often posited as a potential theoretical model for biological networks such as regulatory nets in Section 4. In Section 5 we present some quantitative results showing static metrics, but also the dynamic network break-up behaviour as nodes are attacked or "culled" in rank order. We present a discussion of this approach in Section 6 and offer some conclusions and areas for further study in Section 7.

2 Protein-Protein Networks

As described in [36] there are a number of public domain protein-protein network data sets available. We consider five of these as labelled and summarized in Table 1. These data sets are available from the Web site of the European COSIN project [14] amongst others. We chose these particular five as they range in size and features and make interesting comparisons with the synthetic NK networks we describe in Section 4.

Data Set	$\|V\|$	$\|A\|$	$\|C\|$	Form	Cite
elegans	375	437	20	gss	[70]
yeast	4,142	8,099	99	gss	*ibid.*
heliobacter	732	1,465	16	gss	*ibid.*
ecoli	270	716	20	gijw	[9]
beta3s reduced	1,287	33,813	1	gijw	[59]

Table 1. The numbers of vertices, arcs and (enumerated) clusters for the five publicly available biological data sets.

The elegans, yeast and heliobacter protein-protein interaction data-sets are due to the DIP Project [70] and they represent networks for *Caenorhabditis elegans*; *Saccharomyces Cerevisiae*(yeast); and *Heliobacter Pylori* respectively. They consist of string-labelled node pairs and are encoded as a single string-string graph format (gss). The ecoli protein-protein interaction data-set was reportedly confirmed by reciprocal tagging and purification or by repeat analysis and this set is due to Butland and coworkers [9]. It consists of 270 vertices and is split into 20 separate component clusters. It is encoded as un-weighted integer vertex index pairs (gijw format).

The Beta3s protein folding data-set represents the conformation space of a 20 residue anti-parallel beta-sheet peptide that has been sampled using molecular dynamics simulations and is due to Rao, Caflisch and co-workers [59]. The full data set is the largest one considered in this present work and has $132,168$ vertices. It consists of a weighted set of integer vertex index pairs with the trajectory transitions contributing to the weights. A reduced version of this set with only $1,287$ vertices is also available on the COSIN web site, and this set was apparently constructed for conformations that were visited at least 20 times. This reduced network data-set has a single fully connected component and can have its betweenness centrality more practically computed.

Table 1 shows the range of sizes of network, different ratios of arcs to vertices and also the number of disconnected clusters each network spits into when appropriately enumerated.

3 Betweenness Centrality Metric

We develop graph terminology for describing the betweenness centrality metric. Consider a Graph G with a set V vertices or nodes, of which there are N_V nodes and a set of E edges of which there are N_E individual connecting power lines. In this present paper we do not weight the arcs, and imagine them all part of a single integrated high voltage trunk system connecting exit points to downtransformers and the consumer network structure.

Centrality metrics attempt to rank the nodes in some order specifying which is the most connected or important to the network as a whole. A simple centrality measure is simply the in or out degree of a node. That is, the number of other nodes that it connects to or from. Another centrality metric is the so-called "betweenness". This is defined in terms of the node through which the most number of pathways connecting any two **other** nodes pass. Computing the betweenness involves computing the shortest path distance between each pair of nodes $(s,t); s \in V, t \in V$. We then compute the fraction of the shortest paths that pass through each vertex v and sum this fraction over all possible pairs of vertices (s,t). This can be written as:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{1}$$

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

We can optionally normalize by dividing by number of node pairs not including v. This factor is $(n-1)((n-2)$ although for the work we report here with a fixed and known number of nodes in the graph it is instructive to plot the betweenness centrality un-normalized so we can see an actual number of pathways in the context of the whole network, and other distances and metrics. Computing shortest path data for a network is a long standing problem and although there are several algorithms available [19, 30, 58] in practice the choice (for networks that are not too large) is dominated by the ease of integration with the data structures and the other software apparatus used.

Computationally, the complexity of obtaining the shortest paths is $\mathcal{O}(N_V^3)$ using the Floyd-Warshall algorithm [25]. There are other and newer algorithms such as Brandes' algorithm, which takes $\mathcal{O}(N_V N_E)$ [8]. In this present work the calculations were done repeatedly as networks were progressively allowed to fail and it was sufficient and easiest to use the Floyd-Warshall algorithm

4 NK-Network Models

Random Boolean Network (RBN) models are effectively a generalization of the 1-dimensional Cellular Automata model [71] and are often simulated on a network substrate such as an NK graph. Kauffman's NK-Model [48] of an N-node network with K-inputs to a boolean function residing on each node has found an important role in the study of complex network properties. RBNs have found important applications in biological gene regulatory networks [49] but also in more diverse areas such as quantum gravity through their relationship with ϕ^3-networks [2, 3]. RBNs have many interesting properties [28] and have been amenable to various analyses [47] including mean-field theory. They continue to be an important and interesting tool in studying biological and artificial life problems.

A key property of RBNs is the now well established existence of a frozen phase and a chaotic phase [16, 17] and the critical phase transition lies at the integer value of connectivity $K_c = \frac{1}{2p(1-p)} = 2$ for unbiased networks with a mean boolean function output value of $p = 0.5$. It has therefore been of most interest to study RBNs at or around this critical value. In this present work however we see that biological data generally indicates that protein-protein nets have considerably higher average connectivities.

The Random Boolean Network or graph G is expressed as a four-tuple $\mathbf{G} = (V, E, F, x)$ and has $N = |V| = |F| = |x|$ nodes or vertices, and $N_e = |E|$ directed edges or arcs, which express the K_i inputs for node i. The Kauffman NK-Network is constructed with fixed $K = 1, 2, 3, ..$ and a boolean function f_i of K_i inputs is assigned to each node. All the nodes of the network carry boolean variables x_i which may be initialized randomly and which are updated (usually, but not necessarily) synchronously so that:

$$x_i(t) \leftarrow f_i\left(x_j(t-1)\right), j = 1, 2, ..., K_i \tag{2}$$

In generating network instances of the NK-network model we assign the K_i inputs for node i randomly and with uniform probability across all nodes. Even for a large network there is still a non-zero probability of assigning a node as one of its own inputs. In the case of $K_i > 1$ there is also a possibility of assigning a node j as an input of i more than once. These self-edges or multiple edges can have a subtle effect on the behaviour of the NK-network model [42].

Work has been carried out on a number of different update mechanisms for boolean networks including asynchronous algorithms [32]. In this paper we focus solely on the static structure of NK graphs and we use graph robustness methods and betweenness centrality measures to investigate the role that connectivity makes in comparisons with protein-protein networks. As well as making some calculations of the static network properties we employ a procedure developed by Hawick [37, 38] for analyzing network robustness. We compute the node with the highest betweenness centrality and remove it to see how the whole network properties are changed. This process can be repeated to study the systemic failure or dynamic change pattern of the network as individual components are "attacked" and removed.

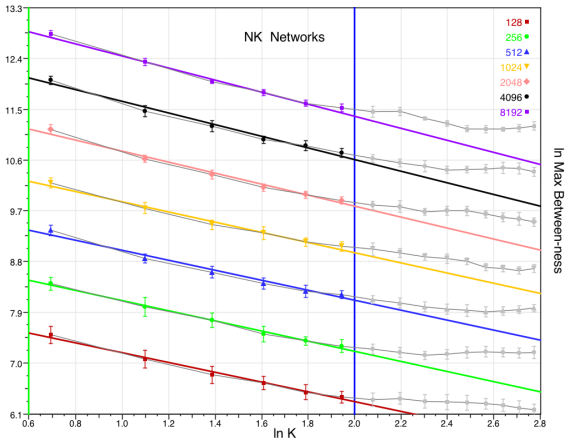

Figure 1. Max betweenness vs K.

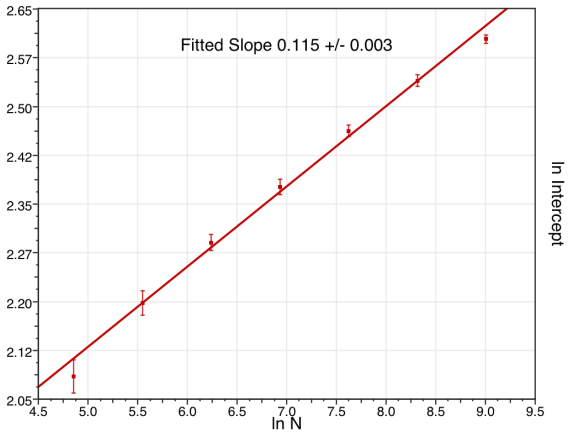

Figure 2. Fitted slope to the computed intercepts of the straight-line segments of the plots of max betweenness vs K.

5 Selected Results

We present a range of graph analyses made on the five protein interaction biological data sets but also on a range of synthetic NK networks. It is first useful to consider the betweenness centrality metric in our synthetic networks. We can adjust both the size N and the connectivity K of our synthetic data. We expect a crucial variation with connectivity, but intuitively we might hope that once we use a big enough network, then it will be a good representation and with average properties similar to other large networks and so we choose $N = 2048$ for the work reported here.

Figure 1 shows maximum values of the betweenness plotted against connectivity K, for various network sizes N. The plots are on a log-log scale and indicate that at small connectivity K the maximum betweenness follows straight lines on the plot. This implies that the maximum betweenness $B_{\max}$ follows a power law and from the systemic slopes we see these themselves do change with network size N but only very weakly.

Figure 2 shows the fitted intercepts plotted vs network

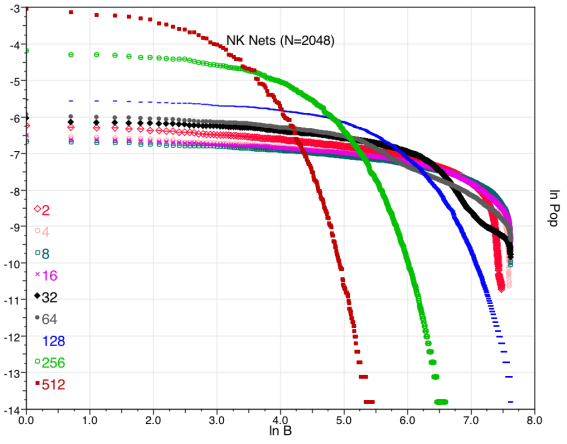

Figure 3. Population distribution of different betweenness values in synthetic NK networks, for N=2048.

size N on a log-log scale. The weak dependence from a least-squares fitted "slope of the slopes" appears to confirm we are justified in studying a fixed network size of $N = 2048$ for this study.

In addition to the maximum betweenness found in a particular network we can use the whole distribution as a potential signature e spectrum to characterize individual networks. We perform this both for the synthetic NK graphs as well as for the biological data sets under consideration.

Figure 3 shows the population distribution of different betweenness values present in the synthetic NK networks, for different connectivities K but for a fixed network size N. They have been normalized (by their total sum) and are plotted on a log-log scale. The curves all follow a characteristic shape with a relatively rapid cut-off on important nodes with a high betweenness. Two curves stand out as having additional features - that for $K = 2$ which as described in Section 4 is known to be the critical K value for these networks, but also for $K = 323$. These plots are averaged over twenty individually randomly generated network instances, and from the standard deviation of which distribution the error bars are generated. We hypothesize that $K = 32$ is closely related to the connectivity limits possible for a network size of $N/2 = 1024$. This is of peripheral interest for our present work but is worth further systemic study in the context of NK and other scale free network generation models.

Figure 4 shows the population distribution of different betweenness values present in the protein-protein networks. They have also been normalized and are plotted on a log-log scale. These are each of their own characteristic size but they appear to exhibit the same characteristic tail off of connectivity but with differing degrees of sharpness. The yeast and beta graphs are large but appear to have smaller connectivities compared to the other three which compare more closely with the synthetically generated high-K graphs. The yeast data set also exhibits a kink

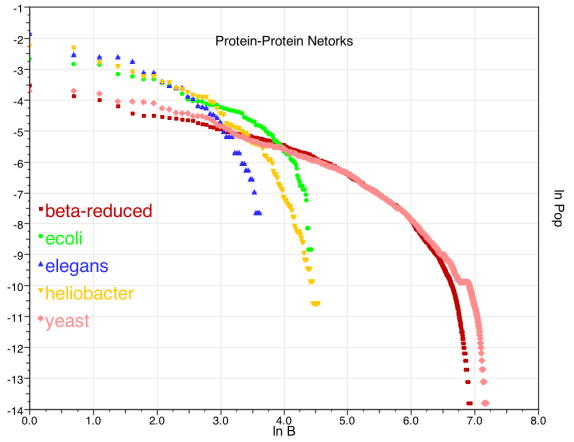

Figure 4. Population distribution of different betweenness values in protein-protein networks.

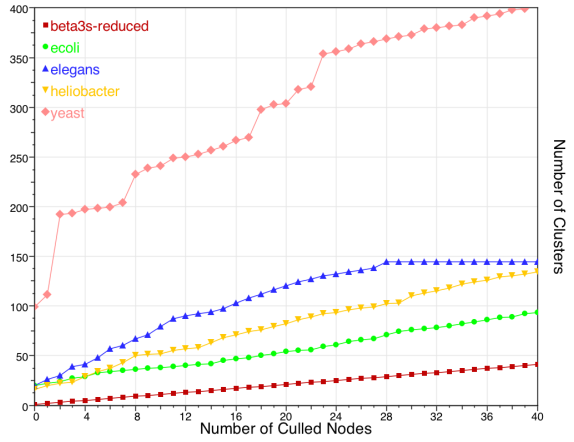

Figure 5. Number of Clusters with culled nodes in order of ranked betweenness centrality.

that is qualitatively similar to that of the $N-2048, K=32$ set.

The node culling procedure described in [37] can be applied to these data sets as well. The original network has its nodes ranked by betweenness centrality and the most crucial node is removed. The process can be repeated an arbitrary number of times. In the plots shown, forty of the nodes are removed in turn. Useful properties to examine during this process are the number of connected clusters of nodes – which obviously rises as the network is attacked – and the Floyd all-pairs distance which can have more complex behaviour as nodes are attacked.

Figure 5 shows how the number of clusters present in the biological data sets increases as nodes are culled in order of maximum betweenness centrality. We see the number of disconnected or separate clusters grows smoothly for the beta protein-protein network suggesting it is a scale free structured system very like the NK sets. The NK data are not shown but they superpose almost exactly on top of the

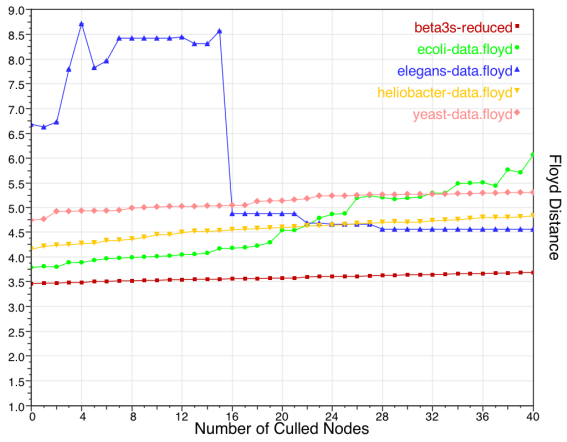

Figure 6. Floyd distance computed as nodes are culled in betweenness rank from the biological data sets.

beta set curve shown, The other sets exhibit a qualitatively similar growth of component clusters but with less monotonic smoothness indicating a greater degree of inhomogeneity in cluster sizes present. An analysis of the modular communities present in these data sets would reveal more details on this inhomogeneity.

Figure 6 shows how the simple hop-based Floyd all-pairs distance metric increases in the biological data sets as nodes are culled in order of maximum betweenness centrality. This is computed for the largest single component present – it is infinite for disconnected nodes. Comparing this to the number of clusters we see the elegans data shows some pathological and counter intuitive changes. The Floyd distance appears to fall but this is due to the changes in the cluster size distribution and emphasizes that this data set degrades in a complex manner and not just with a simple breaking off of small clusters. The ecoli data set also shown some discontinuities but is mostly monotonically rising like the remaining three sets. The smooth behaviour of these is also closely followed by the synthetic NK data sets.

Figure 7 shows how the Floyd distance metric (again in terms of simple graph hops) increases as nodes are culled from the synthetic networks. The curves are almost flat in the number of culled nodes. This indicates a slow and gradual break up of the networks as highly central nodes are removed. Note that the smooth synthetic structure has given rise to a much more gradual break up. This will likely be reflected in the simpler community structure of these synthetic systems with no definite modules dominating breakup and pathological discontinuities.

6 Discussion

It is useful to try to summarize our systems by plotting the Newman clustering coefficient [63] for both the synthetic as well as the biological data sets.

Figure 8 shows a log-log scale plot of the synthetic NK data with the biological sets placed at a guessed effec-

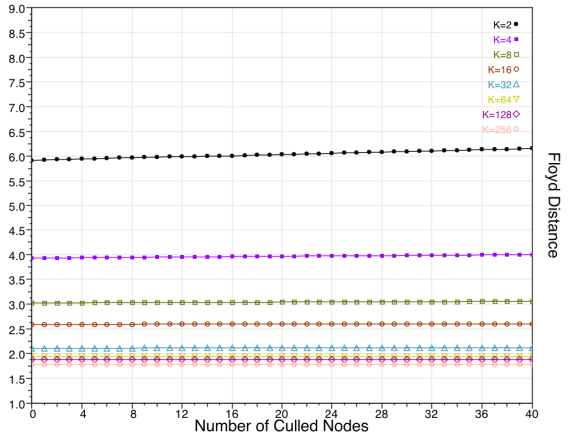

Figure 7. Floyd distance computed as nodes are culled in betweenness rank from the synthetic NK datas sets.

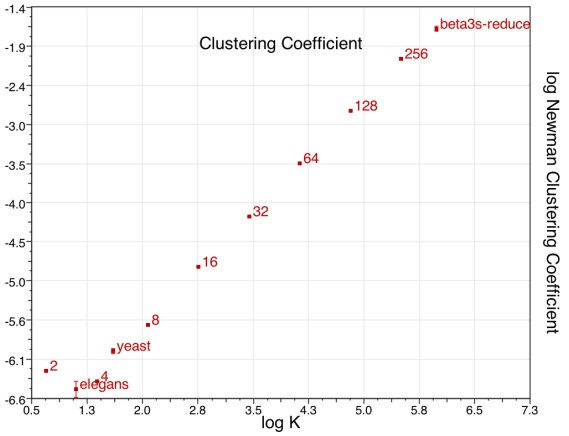

Figure 8. Mean Newman clustering coefficient for biological data sets and also synthetic NK networks.

tive K-connectivity to fit the trend. This analysis seems consistent with the average number of arcs and so forth present in the different data sets as a ratio with the number of nodes.

There are other properties such as the number of monomers, leaf-nodes, and associated ratios that can also be used to categorize networks during the culling process. The size of the largest cluster present is also shown to be useful generally in power and water diustribution networks [39], but the number of clusters and all-pairs distance seem to be most useful in this biological network context.

7 Conclusion

We have demonstrated the utility of the betweenness centrality metric from graph analysis to compare static network properties of both synthetic NK networks of different connectivities and also biological protein-protein interaction data sets obtained from real biological systems.

We have seen that the manner in which the networks fail under attack or removal of the most highly ranked nodes gives a dynamical signature that allows us to compare and categorize biological sample networks. We see for example that the beta network system is most likely a near scale-free system and with a high degree of connectivity. Other systems seem to have lower general connectivities.

We have seen how known properties of a synthetic network structure such as NK graphs can help assess our technique and its value in studying real world networks from measured data such as the bionets discussed. This approach is potentially useful in discussing network properties and in particular inhomogeneities and more complex community structures that exhibit more sophisticated behaviours than smoothly grown synthetic networks. The systemic culling technique allows a categorization of bio networks according to their fragility and progressive break-up properties.

Our focus was on a static analysis of NK systems here since that is the only means of comparison we can make with the five biological data sets available to us. However one could also look at graph structures in the trajectories of the associated RBN dynamic system and measure for example the graph structure of the time loops or attractors in these models.

We have primarily used the betweenness metric to compare signatures of biological protein-protein networks with NK networks, but there is scope to apply these methods more systematically to a broader parameter range of NK class networks themselves to assess their robustness [75] against node attack or failure.

In summary, this approach has potential for categorizing highly complex networks of constrained finite size such as protein-protein or regulatory networks, against the known and more easily studied behaviours of artificially generated networks such as NK systems, which of themselves still exhibit scale-free and other complex systems properties.

References

[1] Abdo, A.H., de Moura, A.P.S.: Clustering as a measure of the local topology of networks. Tech. rep., Universidade de São Paulo, University of Aberdeen (February 2006), http://arxiv.org/abs/physics/0605235v4

[2] Baillie, C.F., Johnston, D.: Damaging 2d quantum gravity. Physics Letters B 326, 51–56 (1994)

[3] Baillie, C., Hawick, K., Johnston, D.: Quenching 2d quantum gravity. Physics Letters B 328(3-4), 284–290 (June 1994)

[4] Barabasi, A.L., Oltvai, Z.N.: Network biology: Understanding the cell's functional organization. Nature Reviews - Genetiocs 5, 101–113 (2004)

[5] Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. PNAS 101(11), 3747–3752 (2004), http://www.pnas.org/cgi/content/abstract/101/11/3747

[6] Blomberg, C.: Physics of Life - The Physicist's Road to Biology. No. ISBN 978-0-444-52798-1, Elsevier (2007)

[7] Blossey, R.: Computational Biology A Statistical Mechanics Perspective. No. ISBN 1-58488-556-4, Chapman and Hall (2006)
[8] Brandes, U.: A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology 25(2), 163–177 (2010)
[9] Butland, G., Peregrin-Alvarez, J.M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., Emili, A.: Interaction network containing conserved and essential protein complexes in escherichia coli. Nature 433, 531–537 (February 2005)
[10] Caflisch, A.: Protein folding: Simple models for a complex process. Structure 12, 1750–1752 (2004)
[11] Choen, J.: Bioinformatics - an introduction for computer scientists. ACM Computing Surveys 36(2), 122–158 (June 2004)
[12] Clauset, A.: Finding local community structure in networks. Phys. Rev. E 72, 026132–1–6 (2005)
[13] Cordero, F., Manini, D., Gribaudo, M., Balbo, G.: A new framework for modeling complex biological systems: user?s manual. Ricercatore - Universita degli Studi di Torino, Dipartimento di Informatica, Corso Svizzera 185 - 10149 Torino (Italy) (2011), http://www.di.unito.it/?manini/OBAM-Tool-Draw/
[14] COSIN: Complexity in networks - european project coevolution and self-organization in dynamical networks (cosin). http://www.cosinproject.org/ (August 2011)
[15] Darabos, C., Tomassini, M., Cunto, F.D., Provero, P., Moore, J.H., Giacobini, M.: Toward robust network based complex systems: from evolutionary cellular automata to biological models. Intelligenza Artificiale 5, 37–47 (2011)
[16] Derrida, B., Pomeau, Y.: Random networks of automata: A simple annealed approximation. Europhys. Lett. 1(2), 45–49 (1986)
[17] Derrida, B., Stauffer, D.: Phase transitions in two-dimensional Kauffman cell automata. Europhys. Lett. 2, 739 (1986)
[18] Dijkstra, E.W.: A note on two problems in connection with graphs. Numerische Mathematik 1, 269–271 (1959)
[19] Dijkstra, E.: A Note on Two Problems in Connexion with Graphs. Numerische Mathematik 1(1), 269–271 (1959)
[20] Ebden, J.H., Phillips, C.A., Rogers, G.L., Langston, M.A.: The maximum clique enumeration problem: Algorithms, applications and implementations. In: Proc. ISBRA 2011. pp. 306–319. No. 6674 in LNBI, Springer (2011)
[21] ElisabethWong, Baur, B., Quader, S., Huang, C.H.: Biological network motif detection: principles and practice. Briefings in Bioinformatics June, 1–14 (2011)
[22] Everett, M., Borgatti, S.P.: Ego network betweenness. Social Networks 27, 31–38 (2005)
[23] Fenyo, D. (ed.): Computational Biology. No. ISBN 978-160-7618-416, Springer, NY, USA. (2010)
[24] Ferraro, N., Palopoli, L., Panni, S., Rombo, S.E.: Asymmetric comparison and querying of biological networks. IEEE/ACM Trans. Computational Biology and Bioinformatics 8, 876–889 (2011)
[25] Floyd, R.W.: Algorithm 97: Shortest Path. Communications of the ACM 5(6), 345 (1962)
[26] Freeman, L.C.: Centrality in social networks conceptual clarification. Social Networks 1, 215–239 (1978)
[27] Freeman, L.C., Borgatti, S.P., White, D.R.: Centrality in valued graphs: A measure of betweenness based on network flow. Social Networks 13, 141–154 (1991)
[28] Gershenson, C.: Introduction to random boolean networks. Tech. Rep. arXiv.org:nlin/0408006v3, Vrije Universiteit Brussel (August 2004)
[29] Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Nat. Acad. Sci. 99, 7821–7826 (2002)
[30] Goldberg, A.V.: A simple shortest path algorithm with linear average time. In: Proc. 9th Annual European Symposium on Algorithms. pp. 230–241 (2001)
[31] Hagburg, A., Schult, D., Swart, P.: NetworkX Reference (4 August 2013), https://networkx.github.io/, release 1.8.1
[32] Harvey, I., Bossomaier, T.: Time out of joint: Attractors in asynchronous random boolean networks. In: Husbands, P., Harvey, I. (eds.) Proc Fourth European Conference on Artificial Life (ECAL97). pp. 67–75. MIT Press (1997)
[33] Hawick, K.A.: Eigenvalue spectra measurements of complex networks. In: H.Arabnia (ed.) Proc. Int. Conf on Scientific Computing (CSC'08). pp. 174–179. CSREA, Las Vegas (14-17 July 2008)
[34] Hawick, K.A.: Detecting and labelling wireless community network structures from eigen-spectra. In: Proc. International Conference on Wireless Networks (ICWN'10). pp. 398–404. WorldComp, Las Vegas, USA (12-15 July 2010), iCW5189
[35] Hawick, K.A.: Interactive Graph Algorithm Visualization and the GraViz Prototype. Tech. Rep. CSTN-061, Computer Science, Massey University, Albany, North Shore 102-904, Auckland, New Zealand (October 2010), www.massey.ac.nz/~kahawick/cstn/061/cstn-061.pdf
[36] Hawick, K.A.: Applying enumerative, spectral and hybrid graph analyses to biological network data. In: Int. Conf. on Computational Intelligence and Bioinformatics (CIB 2011). pp. 89–96. IASTED, Pittsburgh, USA (7-9 November 2011)
[37] Hawick, K.A.: Betweenness centrality metrics for assessing electrical power network robustness against fragmentation and node failure. In: Proc. International Conference on Power and Energy Systems (EuroPES 2012). pp. 186–193. IASTED, Napoli, Italy. (25-27 June 2012)
[38] Hawick, K.A.: Node-failure and islanding in national grid scale electricity distribution networks. In: Proc. Int. Conf. Power and Energy Systems and Applications. pp. 52–58. IASTED, Las Vegas (12-14 November 2012)
[39] Hawick, K.A.: Water distribution network robustness and fragmentation using graph metrics. In: Proc. Int. Conf. on Water Resource Management (AfricaWRM 2012). pp. 304–310. No. 762-037, IASTED, Gabarone, Botswana (3-5 September 2012), cSTN-158
[40] Hawick, K.A., James, H.A.: Managing community membership information in a small-world grid. Research Letters in the Information and Mathematical Sciences 7(CSTN-002), 101–115 (March 2005), iSSN 1175-2777
[41] Hawick, K.A., James, H.A.: Node importance ranking and scaling properties of some complex road networks. Tech. Rep. CSTN-021, Computer Science, Massey University, Albany, North Shore 102-904, Auckland, New Zealand (August 2005)
[42] Hawick, K.A., James, H.A., Scogings, C.J.: Structural Circuits and Attractors in Kauffman Networks. In: Abbass, H.A., Randall, M. (eds.) Proc. Third Australian Conference

on Artificial Life. LNCS, vol. 4828, pp. 189–200. Springer, Gold Coast, Australia (4-6 December 2007), 978-3-540-76930-9

[43] Hawick, K.A., Leist, A., Playne, D.P.: Parallel Graph Component Labelling with GPUs and CUDA. Parallel Computing 36(12), 655–678 (December 2010), www.elsevier.com/locate/parco

[44] Hines, P., Blumsack, S.: A centrality measure for electrical networks. In: Proc. 41st Hawaii Int. Conf. on System Sciences. pp. 1–8. HICSS, Hawaii, Big Island (7-10 January 2008)

[45] Jamil, H.: Computing subgraph isomorphic queries using structural unification and minimum graph structures. In: Proc. SAC 2011, Taichung, Taiwan. pp. 1053–1058 (2011)

[46] Joy, M.P., Brock, A., Ingber, D.E., Huang, S.: High-betweenness proteins in the yeast protein interaction network. Journal of Biomedecine and Biotechnology 2, 96–103 (2005)

[47] Kadanoff, L., Coppersmith, S., Aldana, M.: Boolean dynamics with random couplings. In: Kaplan, E., Marsden, J., Sreenivasan, K. (eds.) Perspectives and Problems in Nonlinear Science. Springer (2003)

[48] Kauffman, S.A.: The Origins of Order. Oxford University Press (1993)

[49] Kauffman, S., Peterson, C., Samuelsson, B., Troein, C.: Random boolean network models and the yeast transcriptional network. Proc. Natl. Acad. Sci. USA 100, 14796 (2003)

[50] Mattiussi, C., Durr, P., Marbach, D., Floreano, D.: Beyond graphs: A new synthesis. Journal of Computational Science 2, 165–177 (2011)

[51] Mueller, L.A.J., Kugler, K.G., Dander, A., Graber, A., Dehmer, M.: Quacn: an r package for analyzing complex biological networks quantitatively. Bioinformatics 27, 140–141 (2011)

[52] Muff, S., Rao, F., Caflisch, A.: Local modularity measure for network clusterizations. Phys. Rev. E 72, 056107 (2005)

[53] Nacher, J.C., Ueda, N., Yamada, T., Kanehisa, M., Akutsu, T.: Study on the clustering coeffcients in metabolic network using a hierarchical framework. In: Fourth International Workshop on Bioinformatics and Systems Biology. pp. 34–35. Kyoto, Japan (31 May - 3 June 2004)

[54] Nepusz, T., Setany, P.P.: iGraph. Budapest, Hungary (4 February 2014), http://igraph.org/python/doc/python-igraph.pdf, version 0.7.0

[55] Newman, M.E.J.: Models of the small world: A review. Journal of Statistical Physics 101, 819–841 (2000)

[56] Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113 (2004)

[57] Pal, D., Rao, P.R.: A tool for fast indexing and querying of graphs. In: Proc. Int. World Wide Web Conf. (IW3C2), Hyderabad, India, 2011. pp. 241–244 (28 March - 1 April 2011)

[58] Prim, R.: Shortest connection networks and some generalizations. The Bell System Technical Journal November, 1389–1401 (1957)

[59] Rao, F., Caflisch, A.: The protein folding network. J. Mol. Biol. 342, 299–306 (2004)

[60] Rao, F.: Theoretical and Computational Studies of Protein Folding Energy Landscapes. Ph.D. thesis, University of Zurich (2005)

[61] Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabasi, A.L.: Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555 (2002)

[62] Samatova, N.F., Schmidt, M.C., Hendrix, W., Breimyer, P., Thomas, K., Park, B.H.: Coupling graph perturbation theory with scalable parallel algorithms for large-scale enumeration of maximal cliques in biological graphs. Journal of Physics: Conf. Series 125, 012053–1–7 (2008)

[63] Schank, T., Wagner, D.: Approximating clustering coeficient and transitivity. Journal of Graph Algorithms ad Applications 9(2), 265–275 (2005)

[64] Schietgat, L.: Graph-based data mining for biological applications. AI Communications 24, 95–96 (2011)

[65] Settanni, G., Fersht, A.R.: High temperature unfolding simulations of the trpz1 peptide. Biophys J. BioFAST 94(11), 4444–4453 (June 2008)

[66] Sevon, P., Eronen, L., Hintsanen, P., Kulovesi, K., Toivonen, H.: Link discovery in graphs derived from biological databases. In: Proc. 3rd Int. Workshop on Data Integration in the Life Sciences (DILS 2006). pp. 35–49 (2006)

[67] Team, J.D.: Java Universal Graph/Network Framework (JUNG2) (22 April 2009), http://sourceforge.net/apps/trac/jung/wiki/JUNGManual, version 2.0

[68] Tizghadam, A., Leon-Garcia, A.: Betweenness centrality and resistance distance in communication networks. IEEE Network Nov/Dec, 10–16 (2010)

[69] Toivonen, H., Zhou, F., Hartikainen, A., Hinkka, A.: Compression of weighted graphs. In: Proc. 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), San Diego, CA, USA. pp. 1–9 (August 2011)

[70] UCLA: Database of interacting proteins (dip). http://dip.doe-mbi.ucla.edu/dip/ (August 2011)

[71] Wolfram, S.: Theory and Applications of Cellular Automata. World Scientific (1986)

[72] Wuchty, S., Ravasz, E., Barabasi, A.L.: The architecture of biological networks. In: Proc. Complex Systems in Biomedecine. pp. 1–30. Kluwer (2003)

[73] Yang, M., Li, R., Chu, T.: Construction of a boolean model of gene and protein regulatory network with memory. Neural Networks 52, 18–24 (2014)

[74] yWorks: yed - java graph editor (2008), http://www.yworks.com/

[75] Zertuche, F.: On the robustness of nk-kauffman networks against changes in their connections and boolean functions. J. Math. Phys. 50, 043513 (2009)

[76] Zgeng, G., Jiang, M., He, X., Zhao, J., Guo, H., Chen, G., Zha, Q., Lu, A.: Discrete derivative: a data slicing algorithm for exploration of sharing biological networks between rheumatoid arthritis and coronary heart disease. BioData Mining 4, 1–21 (2011), www.biodatamining.org/content/4/1/18

[77] Zhou, W., Nakhleh, L.: Properties of metabolic graphs: biological organization or representation artifacts? BMC Bioinformatics 12, 1–12 (2011)

[78] Zhou, Z.: Using expansion algorithm to identify protein from biological networks. Journal of Convergence Information Technology 6, 63–69 (2011)

[79] Zhu, L., Ng, W.K., Cheng, J.: Structure and attribute index for approximate graph matching in large graphs. Information Systems 36, 958–972 (2011)