Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 6, n. 2, 2018.

Trabalho apresentado no XXXVIII CNMAC, Campinas - SP, 2018.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Resilience and Structure of Metabolic Networks

Victor Hugo de Mello Pessoa[1]
Instituto de Física de São Carlos, Universidade de São Paulo
Cynthia de Oliveira Lage Ferreira[2]
Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo

Abstract. In this article we study the behavior of two types of metabolic networks, reaction and compound networks, under the effect of random and intentional attacks. We studied these phenomena in the networks of four species: worm *C. elegans*, fly *D. melanogaster*, yeast *S. cerevisiae* and *H. sapiens*. We also analyze the relationship between robustness and the entropy of the degree distribution. We calculate the correlation between six topological measurements of the networks and this entropy.

Keywords. metabolic networks, resilience, entropy of degree distribution, topological measurements

1 Introduction

A metabolic network [4] consists of metabolites that are converted into others through biochemical reactions. For simplicity, it can be represented as a graph with vertices representing metabolites and edges corresponding to reactions converting one metabolite into the other (compound network). Also, we can consider the reverse, with vertices representing reactions and edges corresponding to metabolites (reaction network). The drawback of these representations is that if the currency metabolites are not removed, biologically false pathways may be introduced in the network. These metabolites are molecules that participate in a large number of chemical reactions with secondary roles. An important question to be answered by the researchers of biological systems concerns the relationship between the structure and function of such networks, particularly with regard to the flow of information and regulation of cellular signals. The study of the relationship between the network structure and function is subject of several works in the field of biological complex systems [2, 7]. Here, we studied the relationship between resilience and network organization, making a comparative analysis of four species, considering the compound and the reaction metabolic networks and calculating their entropy of the degree distribution. By calculating the correlation between this measure of entropy and a set of network topological measurements, we can see which network structural property is most important for network resilience. The correlation between entropy measurements and topological network

[1]victor.pessoa@usp.br
[2]cynthia@icmc.usp.br

measurements were investigated to study the resilience of protein networks [5]. Elsewhere, the relationship between structure and robustness in the metabolic networks was studied using a cascading failure model based on a topological flux balance criterion [6]. The property studied in this work, resilience, plays an important role in the development of computational models, especially with regard to the natural evolution of organisms and the consequent appearance of more sophisticated architecture of the metabolic networks. This work is organized as follows: in the next section we describe the database used. In the following section we define the entropy-based measure and give a brief discussion with respect to the topological measures studied in the metabolic networks considered. The results are presented subsequently.

2 Metabolic Networks

2.1 Database

We consider the database of the Reactome (v56 - www.reactome.org). We compare the metabolic process in cells networks of four different species: *S. cerevisiae* (1450 nodes), *C. elegans* (2019 nodes), *D. melanogaster* (2221 nodes) and *H. sapiens* (4688 nodes). The Reactome database considers as nodes both metabolites and reactions. The reversibility information of the reactions is not included in this database. For all species, only the largest component was taken into account while performing the resilience analysis. We modified these networks in order to have only metabolites as nodes (compound network) and also to have the representation that considers only reaction as nodes (reaction network).

2.2 Currency Metabolites

The currency metabolites are molecules that participate in a large number of chemical reactions with secondary roles. Water, for example, can sometimes participate in a reaction to donate a proton to an amino acid chain and make it reactive, favoring the progress of a reaction. This recurrence generates a false conclusion on the analysis of the topological parameters of the network. In fact, as these molecules tend to appear in a large number of reactions, the metabolic network of the organism tends to have false paths between metabolites, because in this case biochemical pathways have this molecule in common. A recent work [3] characterized currency metabolites as nodes with high centrality. In this sense, our analysis considered two types of networks, as stated above, in which the currency metabolites were not considered: the compound networks and the reaction networks.

3 Network Analysis

The structure of a complex network can be represented as a graph and it can be characterized by its adjacency matrix A. The elements $a_{ij} = 1$ whenever there is an edge connecting the vertices i and j, and $a_{ij} = 0$ otherwise. When the graph is undirected, the adjacency matrix is symmetric. Metabolic networks can be characterized according

DOI: 10.5540/03.2018.006.02.0318

to their structure and topology through network measurements that characterize network connectivity, presence of cycles and distances between nodes. In this work, the measurements considered are the *average neighbor connectivity*, the *average shortest path lenght*, the *diameter*, the *clustering coefficient* and the *assortative coefficient*. Further details of mathematical formulation can be found in [5].

3.1 Entropy Measurement

Entropy related measures can quantify the resilience of metabolic networks. It has important physical implications related to the amount of disorder and information in a system. The *entropy of the degree distribution* is effective to reveal network's resilience under perturbations [8]. It measures the diversity of the link distribution, providing an average measure of network's heterogeneity. In this way, the entropy of the degree distribution $P(k)$ is defined as follows $H = -\sum_k P(k) log[P(k)]$.

Topological Measurements								
Species	S	cc	sp	d	r	s	γ	H
S.cerevisiae (C)	943	0.76	5.72	17	5.32	-0.07	-1.86	0.96
C.elegans (C)	1302	0.75	6.58	21	5.21	-0.06	-1.98	0.95
D.melanogaster (C)	1447	0.76	6.49	20	5.20	-0.06	-2.02	0.94
H.sapiens (C)	2926	0.70	6.98	20	5.29	-0.04	-2.05	0.99
S.cerevisiae (R)	507	0.62	4.99	16	14.79	0.71	-1.29	1.47
C.elegans (R)	717	0.63	5.77	20	12.94	0.73	-1.57	1.45
D.melanogaster (R)	774	0.62	5.65	19	12.28	0.72	-1.55	1.43
H.sapiens (R)	1762	0.57	6.34	19	13.65	0.76	-1.62	1.48

Table 1: Compound (C) and reaction (R) networks: set of measurements obtained for the four species; size of the largest component (S), average clustering coefficient (cc), average shortest path length (sp), network diameter (d), average neighbor connectivity (r), assortative coefficient (s), the scaling exponent of the power law (γ) and entropy of the degree distribution (H).

3.2 Discussion

Our study was based on the analysis of two types of metabolic networks, as described above, namely compound and reaction networks. This section presents a brief discussion regarding the topological measurements of these networks, according to data presented in the table 1. The cumulative degree distribution for all metabolic networks follows a power law $P_{cum}(k) = \alpha k^{\gamma}$ and the values of γ were obtained by using the last squares method. The clustering coefficient is relativity big, specially for the compound networks, indicating that these networks have a modular structure. A recent work [1] discusses the relationship between robustess and modularity, analytically and empirically. The structure of metabolic networks, in particular, play an important role in its functionality, as studied

DOI: 10.5540/03.2018.006.02.0318

in [7]. The cluster coefficient distribution of the compound networks follows a power law indicating that these networks have a hierarchical structure. Also, the small values of the average shortest path length indicate that these networks are small-world, which can be a consequence of the presence of *hubs* and of the modular structure. Finally, the reaction networks are assortative, with big coefficient s, indicating that nodes with similar degree tend to connect and conversely the compound networks are disassortative, with smaller coefficient s.

4 Metodology

The methodology used in this work is the same proposal in [5]. We investigated the effect of random and intentional removals of the nodes in both types of networks of the four species studied. With regard to random removals, all nodes had the same probability to be removed in a network. We calculate the variation of the size of the largest component (S), diameter (d) and average shortest path length (sp) for each network. Observe that each network measurement was normalized with respect to its original value, that is, the measurement of the network without perturbations. Networks present different numbers of nodes and edges and the studied measurements depend on the networks size and density of connections. Thus, with this approach, we study only the variation of the measurement avoiding the effects imposed by the size of the network and its connections density. The resilience of the compound and reaction metabolic networks was also investigated with respect to intentional attacks. The most connected nodes were removed sequentially, according to their degree. We did such analysis removing up to 5 percent of the total number of vertices and obtained the relative variation of the same measurements considered in the random attack as described above. As in the case of random fails, the relative variation was obtained by the ratio between the measurement calculated when a fraction f of nodes was removed and the original measurement.

5 Results

In figure 1 we can observe the results of the attacks in the networks of the four species studied and the variations observed in the measurements considered for the analysis of the resilience of these networks. Concerning random attacks, we can observe that both networks of the all species presente similar behavior. The main difference occurs with the diameter of the *H.sapiens*. While it decreases in the other species, the diameter increases for the both networks of this specie. Futhermore, the average shortest path length of all networks present very small variations. Therefore, the networks of all species have similar resilience against random fails. For all species, we can observe that the greatest variation of the measurements of the networks occurred for the compound networks when subjected to intentional attacks. The greater sensitivity of compound networks to this type of attack can be explained by the scale-free structure. In these networks a few *hubs* concentrate a greater number of links. The compound network of the *H.sapiens* has lower decrease of the size of the largest connected component (S) in relation to other species. It is reduced

DOI: 10.5540/03.2018.006.02.0318

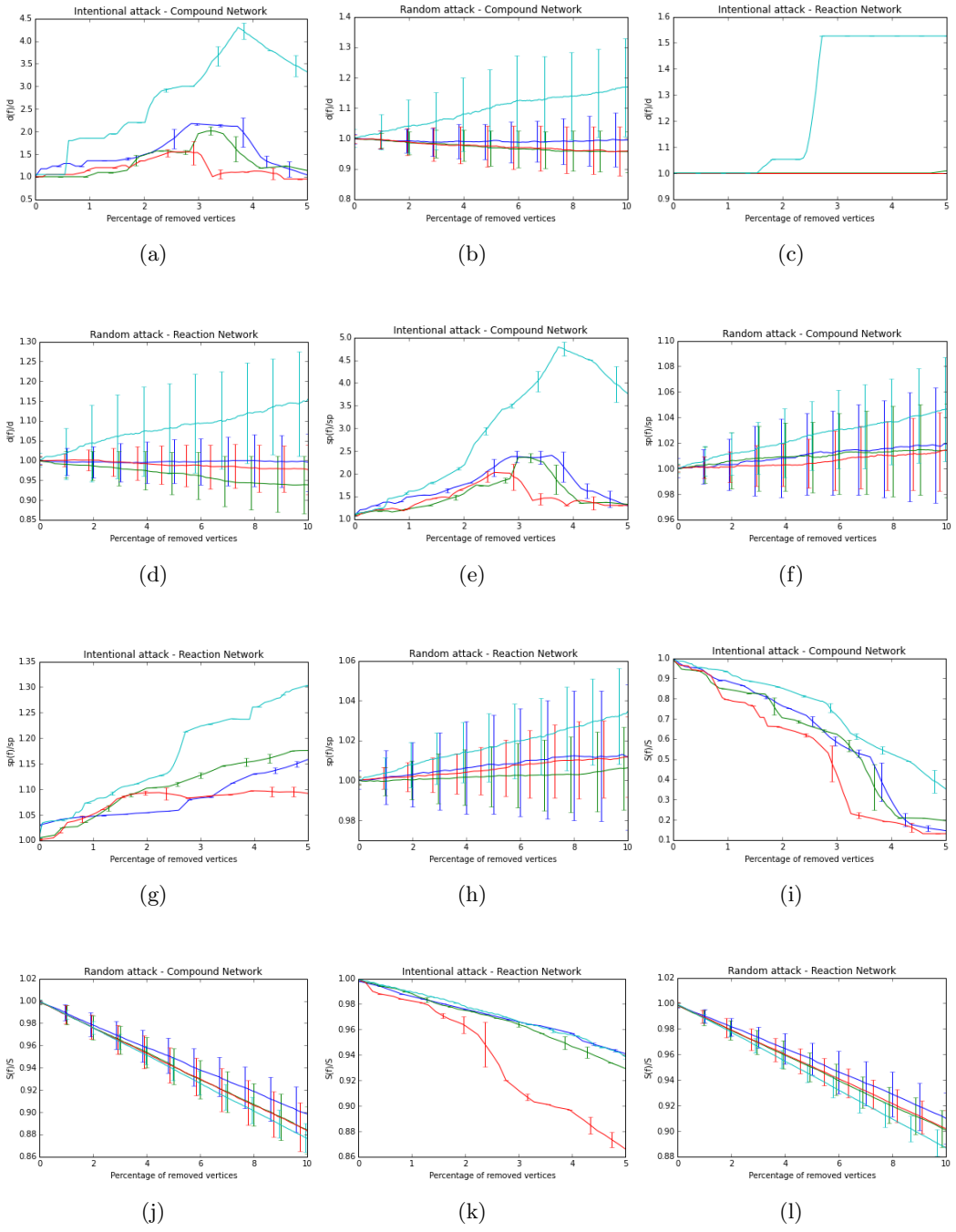

Figure 1: The line colors in the graphs represent *S. cerevisiae* (in blue), *D. melanogaster* (in red), *C. elegans* (in green) and *H. sapiens* (in cyan). These images illustrate the variation of several topologic parameters in each of the selected organisms when their metabolic networks are under attack. The graphs were obtained from the average of 200 simulations.

DOI: 10.5540/03.2018.006.02.0318

to approximately 35% of the original size while other species were reduced in size to between 10% and 20%. However, the diameter and the average shortest path length of the *H.sapiens* increase. Concerning reaction networks, we can highlight the sudden change in diameter of the *H.sapiens* network, indicating that few vertices may have influence with respect to this measure. In the next section we will study the topological measurements and their relationship with the robustness of the metabolic networks.

6 About Network Structure and Resilience

The entropy of the degree distribution has been taken into account for quantification of network resilience. Here, we adopted this measurement in order to identify which topological properties of the metabolic networks are directly related to their resilience. In the table 1 we can observe the entropy values for each metabolic network. We can note that more resilient networks tend to have higher structural entropies. Continuing this work, we calculated the Pearson correlation coefficient (ρ, table 2) between the entropy and a set of six topological measurements. In this way, we indentificated the relationships between each given structural features and the robustness. By looking the correlation values, we can observed that the average neighbor degree r and the assortative coefficient s are the measurements that present the highest correlation with the entropy. In fact, large values of average neighbor degree of nodes indicate alternative connections between neighbors of such nodes. Consequently, when considering *hubs* removals, the high density connection between nodes favor network resilience. Likewise, high values of the assortative coefficient s indicate that nodes with a similar degree tend to connect. Again, when we consider the removal of *hubs*, the fact that they are connected to highly connected nodes also favors network resilience. Thus, we can observe that, in fact, reaction networks have much higher values for the measurements cited than the compound networks, which may explain a lower resilience of the latter to intentional attacks. It is also interesting to note that the clustering coefficient cc is strongly correlated, negatively, with entropy. This indicates that small values of the clustering coefficient favor resilience. In fact, metabolic networks have a high clustering coefficient due to their high modular structure. And as studied in [1], this modular structure can make the network less robust, especially when this modular structure has importance in the network functionality.

Pearson Correlation Coefficient ρ						
Topological Measures	cc	d	γ	r	s	sp
Compound Networks	-0.87	-0.16	-0.2	0.74	0.65	0.35
Reaction Networks	-0.65	-0.38	0.15	0.79	0.56	0.19

Table 2: The correlations between the entropy and each measurement is indicated by ρ.

DOI: 10.5540/03.2018.006.02.0318

7 Conclusion

In this article we study the behavior of two types of metabolic networks of four species under the effect of random and intentional attacks. Computational studies showed that the measurements considered of all networks presented similar variations when subjected to random attacks. However, when considered intentional attacks, such networks presented topological variations. We also analyzed the relationship between robustness and the entropy of the degree distribution by using the Pearson correlation. We noted that the average neighbor degree and the assortative coefficient contributed significantly to resilience. We also observed, in particular, that the clustering coefficient contributes, negatively, to the resilience. This research could be repeated considering other topological measures, entropy measures or node removal dynamics. In addition, in the case of metabolic networks, to consider this study in directed networks may mean a promising study.

8 Acknowledgements

Cynthia and Victor thanks FAPESP (2013/13330-9) and CNPq, respectively.

References

[1] J. P. Bagrow, S. Lehmann and Y. Y. Ahn, *Robustness and modular structure in networks*, Network Science, 3 (4): 509-525, 2015.

[2] E. Bullmore and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems*, Nat Rev Neurosci, 10 (3): 186-198, 2009.

[3] P. Holme and M. Huss, *Currency Metabolites and Network Representations of Metabolism*, arXiv:0806.2763v1 [q-bio.MN], 2008.

[4] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A. L. Barabasi, *The large-scale organization of metabolic networks*, Nature, 407 (6804): 651-654, 2002.

[5] F. A. Rodrigues, L. F. Costa and A. L. Barbieri, *Resilience of protein-protein interaction networks as determined by their large-scale topological features*, Mol. BioSyst., 7: 1263-1269, 2011.

[6] A. G. Smart, L. A. N. Amaral and J. M. Ottino, *Cascading failure and robustness in metabolic networks*, Proc. Natl. Acad. Sci. U.S.A., 105: 13223-13228, 2008.

[7] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster and E. D. Gilles, *Metabolic network structure determines key aspects of functionality and regulation*, Nature, 420 (6912): 190-193, 2002.

[8] B. Wang, H. Tang, C. Guo and Z. Xiu, *Entropy optimization of scale-free networks robustness to random failures*, Physica A, 363 (2): 591-596, 2006.

