EPJ manuscript No.
(will be inserted by the editor)

Network analysis of metabolic subsystems

Rok Novosel[a] and Matija Čufar[b]

Faculty of Computer and Information Science

Abstract. Subsystems are parts of a metabolism that perform different important tasks in a cell. In this article, we explore these subsystems from a network science point of view. We attempt to find ways of detecting subsystems in a metabolic network using community detection algorithms. We use the Louvain modularity optimization algorithm and the Clauset-Newman-Moore algorithms as a baseline against which we compare the effectiveness of other algorithms. As a comparison, we use the Girvan-Newman algorithm and a motif-based community detection approach. We present the results on a metabolic network of the Chinese hamster ovary cell, a mammalian cell that is commonly used in biomedical research and in biotechnology.

1 Introduction

Since the turn of the century, life sciences have been evolving rapidly. Advances in data acquisition, storage and analysis technology have allowed scientist to gather immense amounts of data and build complex models from it [9]. These complicated models have brought people of various backgrounds, such as physics, mathematics and computer science into the field of biology.

One such field is network science, which is often used to analyze different kinds of networks that appear in the various subfields of modern biology, including ecology [15], systems biology [1] and neuroscience [16].

Metabolic networks [10] are used to model the metabolisms of various organisms. They are usually represented with a bipartite graph composed of two types of vertices: reactions and chemicals produced and consumed by the reactions. The edges in such a network connect chemicals to reactions. Furthermore, the edges are directed indicating whether the chemical was produced or consumed. A third kind of vertex can be added to represent enzymes that catalyze the reaction, but do not directly partake in it. Other commonly used representations are simplified representations, where one of the types of vertices is omitted [14].

In this article, we analyze the subsystems in a metabolic network of the Chinese hamster ovary cell. We explore different methods of detecting the subsystems and compare their structures, focusing on the Girvan-Newman method [5] and an approach based on network motifs [2]. We expect these algorithms to outperform other commoly used algorithms such as Louvain optimization or the Clauset-Newman-Moore algorithm.

[a] e-mail: rn0450@student.uni-lj.si
[b] e-mail: matijacufar@gmail.com

2 Related works

In [10] Jeong et. al offer a general overview of the organization and structure of metabolic networks. The authors compare the metabolic networks of 43 different organisms and find that metabolic networks belong to the class of scale-free networks. Moreover, they find that the network diameter is consistent across all of the analyzed networks, irrespective of the number of substrates found in the given species.

In [8], Holme et. al present a method to decompose metabolic networks into subnetworks based on the global network structure. They use a modified Girvan-Newman algorithm [5] to construct a hierarchical clustering tree. They argue that instead of looking at a particular partitioning of a network, we should look at the hierarchical clustering tree as whole, since well-defined subnetworks appear at different heights.

A framework for clustering networks based on higher order structures (e.g. motifs) is introduced in [2]. The authors of the article present an algorithm that performs community detection by using three node motifs. It works by cutting the network into communities in a way that minimizes the ratio of the number of motifs cut to the number of nodes in instances of the motif.

3 Methods

In this article, we analyse a metabolic network of the Chinese hamster ovary (CHO) cell. The CHO cell is frequently used in biological and medical research and in the production of biopharmaceuticals [7].

We use a whole-cell metabolic network of the Chinese hamster ovary (CHO) cell that was taken from the BiGG database [11, 7]. The original network contains 4,456

C_d	C_u	E_{90}	ρ
0.012	0.069	15	0.015

Table 1. The directed and undirected clustering coefficients C_d and C_u, effective diameter E_{90} and the density ρ of the network.

metabolites that take part in 6,663 reactions. The reactions and metabolites are annotated with additional metadata, such as name, subsystem, BiGG ID etc.

The network was simplified to a simple directed graph, where reactions are represented with nodes. If one reaction produces a metabolite that is used by another reaction, they are connected by an arc. This network has 6,663 nodes and 656,609 arcs. If we treat as an undirected network, it has 546,208 edges.

We use two commonly used network community detection algorithms, namely the Louvain modularity optimization algorithm [3] and the Clauset-Newman-Moore algorithm [4], as the baseline and compare them to motif-based clustering methods [2].

The motif-based clustering method, motif counting and the Clauset-Newman-Moore algorithm were taken from the SNAP library [12] and the Louvain modularity optimization algorithm was taken from the Networkx Python library [6].

4 Results

The network has a very large connected component of 6,036 nodes, while the other components are very small, as they are composed of at most 4 nodes. The largest connected component contains a strongly connected component of 5,307 nodes, while the other nodes are isolated. These probably represent sources and sinks of the metabolism.

The network appears to have a scale-free structure. Its in-degree, out-degree and degree distributions are plotted in figure 1. Some commonly used metrics are presented in table 1.

First, we count the appearances of all 13 directed three-node motifs in the network. The significances [13] of each motif are presented in the second row of table 2. The significances were calculated by comparing the motif occurances with 10 instances of Erdős-Rény random graphs of the same size and density.

The normalized mutual information scores of the motif based clustering approach are listed in the last row of table 2. We have calculated these scores by comparing the algorithm's results and the actual subsystems listed in the network annotation. The results are not the best, but still much better than the Louvain and the Clauset-Newman-Moore method, listed in table 3. We partly attribute the quality of the results to the fact that most biological networks tend to be noisy.

NOTE: we plan on also using Infomap and the Girvan-Newman algorithms, but it seems like they will take a few more days to compute.

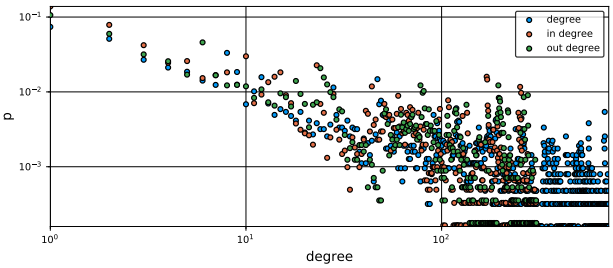

Fig. 1. The in-degree, out-degree and degree distributions of the network, plotted on a log-log scale.

5 Discussion

6 Conclusion

7 Authors contributions

All the authors were involved in the preparation of the manuscript. All the authors have read and approved the final manuscript.

References

1. Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.
2. Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.
3. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
4. Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
5. Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.
6. Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.
7. Hooman Hefzi, Kok Siong Ang, Michael Hanscho, Aarash Bordbar, David Ruckerbauer, Meiyappan Lakshmanan, Camila A Orellana, Deniz Baycin-Hizal, Yingxiang Huang, Daniel Ley, et al. A consensus genome-scale reconstruction of chinese hamster ovary cell metabolism. *Cell Systems*, 3(5):434–443, 2016.
8. Petter Holme, Mikael Huss, and Hawoong Jeong. Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, 19(4):532–538, 2003.
9. B. Ingalls. *Mathematical Modelling in Systems Biology: An Introduction*. Applied Mathematics, University of Waterloo, 2012.
10. Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

													
motif	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
Z	-379.0	496.4	6,523	1,171,385	1,055	3,566	4,604	1,411	-867.2	2,599	1,293	1,387	40,286
NMI	0.44	0.40	0.48	0.64	0.23	0.43	0.46	0.11	0.09	0.09	0.20	0.23	0.42

Table 2. Motif significance Z compared to a random network and normalized mutual information score for motif clustering using each of the 13 motifs.

Algorithm	Louvain Modularity	Clauset-Newman-Moore
NMI	0.1	0.27

Table 3. The normalized mutual information scores of the other algorithms we used.

11. Zachary A. King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A. Lerman, Ali Ebrahim, Bernhard O. Palsson, and Nathan E. Lewis. Bigg models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):D515, 2016.
12. Jure Leskovec and Rok Sosič. Snap: A general-purpose network analysis and graph-mining library. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 8(1):1, 2016.
13. Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
14. M. Newman. *Networks: An Introduction*. OUP Oxford, 2010.
15. Stephen R Proulx, Daniel EL Promislow, and Patrick C Phillips. Network thinking in ecology and evolution. *Trends in Ecology & Evolution*, 20(6):345–353, 2005.
16. Olaf Sporns. Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience*, 17(5):652–660, 2014.