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 as a baseline against which we compare the effectiveness of other algorithms. We use the Girvan-Newman algorithm and a novel approach using higher order structures for network clustering as a comparison. We use the algorithms 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 [6]. 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 [10], systems biology [1] and neuroscience [11].

Metabolic networks [7] 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 [9].

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.

2 Related works

Paper [7] offers a general overview of the organization and structure of metabolic networks. They 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.

In [5] they present a method to decompose metabolic networks into subnetworks based on the global network structure. They use a modified Girvan-Newman algorithm [3] 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].

3 Methods

We apply a few different community detection methods and compare how well they detect the subsystems of a metabolic network.

We haven't decided on the algorithms yet, but we try some of the following:

- Some of the algorithms from the lectures.
- Community detection based on motifs [2].
- Removing nodes with the highest betweenness centrality (or some other measure) and observing how the network falls apart [5].

4 Results

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 [4].

We used a whole-cell metabolic network of the Chinese hamster ovary (CHO) cell that was taken from the

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

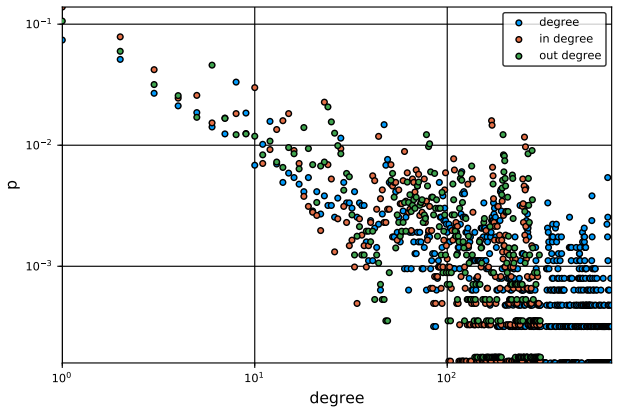

Fig. 1. The in-degree, out-degree and degree distributions of the network.

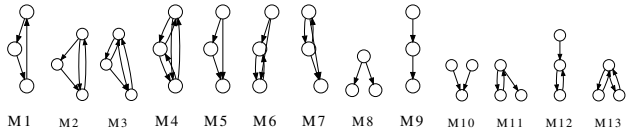

Fig. 2. All 13 directed three-node motifs.

BiGG database [8,4]. The original network contains 4,456 metabolites that take part in 6,663 reactions. The reactions and metabolites are annotated with additional metadata, such as name, subsystem, BiGG ID etc.

We simplify the network 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 546208 edges.

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. It has a very low clustering coefficient, around 0.069.

5 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. 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.
4. 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.
5. Petter Holme, Mikael Huss, and Hawoong Jeong. Subnetwork hierarchies of biochemical pathways. *Bioinformatics*, 19(4):532–538, 2003.
6. B. Ingalls. *Mathematical Modelling in Systems Biology: An Introduction*. Applied Mathematics, University of Waterloo, 2012.
7. 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.
8. Zachary A. King, Justin Lu, Andreas Drger, 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.
9. M. Newman. *Networks: An Introduction*. OUP Oxford, 2010.
10. 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.
11. Olaf Sporns. Contributions and challenges for network models in cognitive neuroscience. *Nature neuroscience*, 17(5):652–660, 2014.