Conservation of Edge Essentiality Profiles in Metabolic Networks Across Species

Tomasz Arodź

Institute of Computer Science, AGH University of Science and Technology,
al. Mickiewicza 30, 30-059 Kraków, Poland
arodz@agh.edu.pl

Abstract. Reactions involved in cellular metabolism form a complex network susceptible to targeted attacks. Recent experiments show that several descriptors of edge essentiality correlate well with lethality of silencing corresponding genes in a model organism, opening path to identifying targets for antimicrobial drugs that would disrupt network functioning in bacteria. However, correlation of high essentiality with experiment is necessary but not sufficient for a descriptor to be useful. Also, the essentialities of corresponding edges have to differ markedly between pathogens and hosts, to yield minimal effect on the latter. Here, we analyse similarity of profiles of several edge essentiality measures across multiple species. We show that local measures, based on degrees of a substrate and a product linked by the edge, or on the alternative paths connecting the two, are evolutionarily conserved within bacteria, archaea and eukaryotes, but also differ between these groups, leading to isolated clusters of species. Furthermore, comparison with a global topological measure, the relative decrease in network efficiency upon edge removal, shows that metabolic networks are more conserved locally than globally.

Keywords: Complex networks, metabolic networks, edge essentiality, clustering of organisms.

1 Introduction

Complex networks gained recognition in recent years as a universal framework for modelling many natural phenomena. In molecular biology, as the knowledge on individual constituents of the living cell grows, focus shifts to improving our understanding of cellular processes by exploring functional, structural or causal interactions between entities in cellular networks. One biological network that has been the subject of research is the metabolic network, an integrated view of known pathways of cellular metabolism. The network is composed of metabolites linked through enzymatic reactions transforming them.

Metabolic networks were shown to belong to a class of complex networks with heavy-tailed distribution of vertex connectivity [1]. Complex networks of this type are abundant in many natural, social and artificial systems [2]. Such networks are susceptible to planned attack on selected components, but are resistant to small random changes [3]. In the context of metabolic networks, this

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 1865–1876, 2009.

property suggests that random changes in the topology of the network may accumulate through evolution, with edges disappearing as a result of deletion or silencing of underlying genes. New edges and, consequently, sometimes also new nodes may appear as new genes are acquired through horizontal gene transfer or evolutionary innovation. Hence, networks of more evolutionarily distant species may diverge in a similar way as their individual genes do. Indeed, several aggregated measures of network topology were shown to differ on average between large group of organisms. These include mean shortest path length [4], mean betweenness centrality or mean cluster coefficient [5]. However, these measures lack direct correspondence to observed biological traits of the organisms.

In this paper, we analyse network descriptors that are designed to capture essentiality of edges, and thus of underlying reactions and of genes encoding enzymes catalysing them, for the functioning of a metabolic network. Several indicators of edge importance have been shown to correlate with experimental results on lethality of edge removal. The existence of alternative pathways linking the substrates and products of the removed edge was shown to be indicative of essentiality of a given edge for survival in *S. cerevisiae*, a species of eukaryota [6]. Edge essentiality was analysed also in the context of connectivity of the metabolites that the edge connects, in a study on lethality of gene deletion in *S. cerevisiae* and in *E. coli* [7]. More global view of edge essentiality, based on shortest paths connecting all vertices of the network, was also analysed, but without experimental validation in metabolic networks. Increase in the average shortest path distances between metabolites after edge deletion was proposed as a measure of edge vulnerability [8]. A related measure, better suited for disconnected networks, based on changes in global efficiency, was suggested for analysing the effect of multi-target drugs on regulatory networks [9,10].

Accurately capturing edge essentiality is crucial for isolating a group of edges corresponding to gene products that can be used as targets for drugs designed to act on metabolic pathways of human pathogens. Antimicrobial drugs inhibiting reactions that correspond to those genes could disrupt functioning of bacterial metabolic network and lead to eradication of the infection. However, in addition to being correlated with underlying physiology of the organisms' metabolism, a useful essentiality measure must be discriminative between different groups of species. In particular, the pattern of values of the descriptors should be capable of clustering networks corresponding to bacteria into a separate group than networks of their eukaryotic hosts, for example human. Otherwise, the lethality of silencing the identified genes would not differ between the organisms, leading to toxicity of the drug.

In this paper, we explore essentiality profiles, describing patterns of values of essentiality measures for edges in each network. We analyse whether they allow for reliable clustering of related organisms into isolated groups. To this end, we analyse metabolic networks of 89 bacteria, archaea and eukaryotes, including human. We evaluate three measures of essentiality, a local one based on degree of the metabolites incident with an edge, a local neighbourhood-oriented measure estimating the existence and length of alternative paths linking substrate and

product of the deleted edge, and the measure of the effect of edge removal on global efficiency of transport in the network. For each measure, we analyse its general properties for metabolic networks, and evaluate the extent to which it is capable of partitioning networks into clusters that separate bacteria from archaea and from eukaryotes.

2 Methods for Analysing Edge Importance in Metabolic Networks

The importance of some edges for the functioning of a network can be analysed on the varying level of localness. Edge deletion has a direct effect on the ability to transform a substrate into a product the edge connects. For a directed graph G, with the length of the shortest path from metabolite v to w denoted as $d_{vw}(G)$, the *edge range* of the edge (v, w) [11,8] is

$$E^{rng}_{(v,w)}(G) = d_{vw}(G \setminus (v, w)) \ . \tag{1}$$

Edge range captures the length of the shortest alternative path bypassing the removed direct link between v and w, and reaches infinity in cases where such a path no longer exists in the entire graph. In this respect, despite its local focus on a single substrate-product transformation, E^{rng} is a measure that extends beyond the local neighbourhood of the edge.

Removing an edge may also have a more global effect on the network, in particular if the edge is on the connecting paths of many metabolites. In such cases, the new network may have lower efficiency of transport. We measure the efficiency of a metabolic network represented by a graph with V vertices using the *global efficiency* measure [12], E^{eff}, defined as the normalised sum over reciprocals of shortest paths lengths between all vertices pairs (v, w) within a graph,

$$E^{eff}(G) = \frac{1}{V(V-1)} \sum_{v \neq w} \frac{1}{d_{vw}(G)} \ . \tag{2}$$

Global efficiency has been proposed as a measure that captures parallel nature of the transfer through paths in the network [13]. In this aspect, it complements graph diameter, a measure that accentuates serial transfer, which has been already applied to analyse metabolic networks and was shown not to discriminate between taxonomic groups of species [5]. Global efficiency was employed in biological context for analysing the effect of multi-target drugs on regulatory networks [9,10]. It relies on the same information as the average shortest paths descriptor, but is better suited for graphs that are not connected, such as metabolic networks. For such graphs, average shortest paths are limited only to a single connected component, or only to pairs of connected vertices. The latter approach may result in unreliable results, such as drop in the value of the average shortest path length when the network loses connectivity after deleting an edge.

Global efficiency of the network may decrease when some edges are eliminated from the network, leading to increase in some shortest paths. We measure the drop in efficiency caused by inhibition of a reaction R as the percentage of the decrease in global efficiency, ΔE^{eff}, defined as

$$\Delta E_R^{eff}(G) = \frac{E^{eff}(G) - E^{eff}(G \setminus \{(v,w) \in R\})}{E^{eff}(G)} \ . \tag{3}$$

Edge-related decrease in global efficiency can also be calculated, if a single edge is used in place of the set of edges corresponding to the reaction R. Changes in global efficiency reflect network-wide effects of removal of some edges, and can in principle be used to analyse global essentiality of reactions and edges they represent.

We have also included in our analysis a purely local measure of edge significance, *edge degree* [14]. It relates edge importance to the degrees of the vertices it connects. For edge linking a substrate v to a product w in some reaction, the edge degree, E^{deg}, is defined as

$$E_{(v,w)}^{deg}(G) = deg_{out}(v)\, deg_{in}(w) \ . \tag{4}$$

It is proportional to the number of products the substrate can be transformed into, $deg_{out}(v)$, and to the number of substrates the product may be directly resulting from, $deg_{in}(w)$. Thus, high values of the measure capture the hub-status of the edge's incident vertices.

With the measures above, we analysed publicly available metabolic networks of 89 organisms [4]. Each network is composed of vertices, representing metabolites, and enzymatic reactions transforming the metabolites. Reactions are represented by directed edges between vertices, with a single reaction corresponding to one or more edges, depending on the number of substrates and products, and their transformations. The edges represent only the transformations between main substrates and products of the reaction. Changes in cofactors and current metabolites during the reaction are not reflected in the metabolic network. This makes the paths through the network more realistic, by eliminating spurious connection through common reactants, such as ATP. In case the reaction is reversible, it is represented in the directed graph by two edges.

We evaluated edge range, edge degree and decrease in efficiency for each edge in every analysed metabolic network. The resulting values form, for each organism, a profile vector, with non-zero entries for edges present in the given network. Furthermore, we calculated relative decrease in global efficiency of the network for each metabolic reaction present in the organism. We normalized the vectors to unit lengths, and then calculated similarity between organisms as dot products between corresponding profile vectors. Thus, the similarity between organisms is equivalent to the cosine measure of their edge- or reaction-importance profiles. In this way, the agreement between organisms in elements of high essentiality is more valued than conformity between elements of low values of the importance measures. The resulting similarity matrix was visualised with the classical metric multidimensional scaling (MDS). This results in two-dimensional representation

of the vectors as points, with the Euclidean distances between points on the plane approximating their original dissimilarities between species.

3 Local Edge Importance in Metabolic Networks

The edge range may be either finite, indicating the existence of an alternative path leading from substrate to product, or it can be infinite when no alternative paths are present. The fraction of edges with no alternative paths is high in metabolic networks, with average and standard deviation values of $78 \pm 4\%$ for eukaryotes, $77 \pm 5\%$ for bacteria and $87 \pm 5\%$ for archaea. As shown in top left panel of Fig. 1, the fraction for larger networks is lower than for smaller networks, although the effect is not consistent through all species. For networks of the same size, the fraction is usually higher for archaea and eukaryotes than for bacteria. For the substrate-product pairs that have alternative paths, the length of the shortest alternative path is typically longer for eukaryotes and archaea than for bacteria of the same network size (top right panel of Fig. 1). In effect, for networks of similar size, edges of eukaryotes and archaea, compared to bacteria, have alternative paths less frequently, and when such paths do exist, they are longer.

High fraction of edges of infinite range stems from the existence of a single giant component in metabolic networks [15]. Networks of metabolites are composed of a single large connected component, and a set of metabolites that are either only substrates or products of the component, or are not connected to it. The fraction of edges with no alternatives is negatively correlated with the relative size of the giant component within the network, as depicted in bottom left panel of Fig. 1. As the size of the component does not extend 30% percent of the network, most edges have at least one end in the sparsely connected parts of the network, with reduced redundancy. Indeed, we have inspected each edge in all the networks, and for edges with both ends outside the giant component, $94 \pm 2\%$ on average across all species had infinite edge range, while for edges belonging to the component, only $47 \pm 10\%$ had no alternative paths. For edges with only the product in the component, the figure was $81 \pm 17\%$, and for ones with only the substrate in the component, it was $63 \pm 15\%$.

Edge range is local in the sense that it involves the transition from a single substrate to a single product. However, it captures some information about the neighbourhood of the edge, or even the whole graph in case of no alternative paths, but even then from the perspective of the influence of graph topology on this single substrate-product pair. We analysed also a purely local edge-related measure, the edge degree E^{deg}, which involves only the counts of edges incident with the edge. As shown in middle left panel of Fig. 1, mean edge degree is lower for eukaryotes and archaea than for bacteria of similar network size. Similar behaviour is observed with maximal edge degree (middle right panel of Fig. 1). Typical edge degree varying between 4 and 9 is in the range to be expected from the distribution of the number of incoming and outgoing links reported previously [1], with vertex degree distribution following a power law

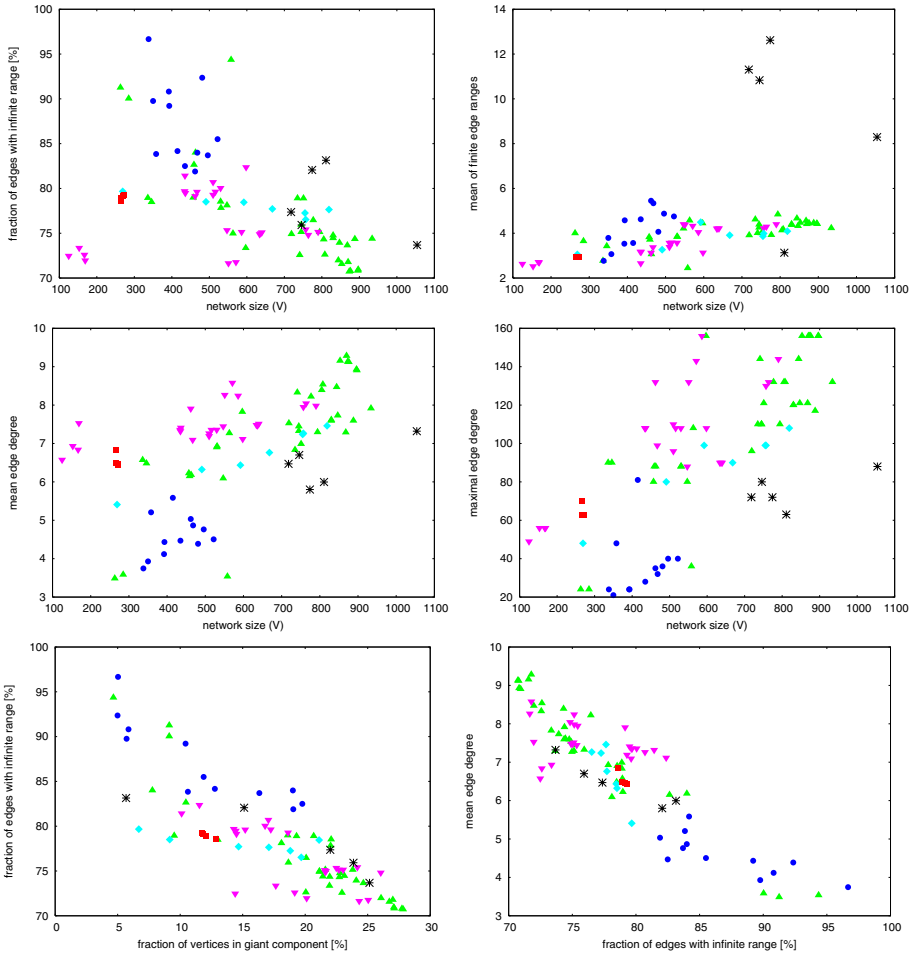

Fig. 1. Properties of local measures of edge essentiality. Points represent eukaryotes (black asterisks), archaea (blue circles), and four phyla of bacteria - proteobacteria (green upward triangles), firmicutes (magenta downward triangles), actinobacteria (cyan diamonds) and chlamydia (red squares).

and low connected metabolites being the most abundant. Furthermore, as shown in bottom right panel of Fig. 1, metabolic networks with higher fraction of edges with no alternative paths have lower mean degrees. This would suggest that the edges with high range value have low degrees, that is, are typically linking substrates that can be transformed into small number of products and products with few source substrates.

We analysed whether there is indeed a relation between edge range and edge degree for individual edges. To this end, we calculated the reciprocal of edge range, arriving with finite values even for infinite edge range, and estimated the coefficient of its correlation with edge degree. For most species, as shown in the bottom panel in Fig. 2, the correlation is larger than 0.5. The exception is a

group of several archaea and bacteria that have very high fraction of infinite range edges, exceeding 85% or even 90%, which form pattern vectors containing mostly zeros when edge range reciprocals are used.

Differences in averages or other aggregated statistics of edge importance measures are informative regarding the general properties of topological organisation of metabolic networks, and patterns of its variability across species. They do not indicate, however, the extent to which these measures are conserved locally, in the corresponding neighbourhoods within the networks. We have analysed this by constructing profiles based on the values of edge range and edge degree, and comparing the profiles edge-wise. The results presented with the use of classical metric MDS in Fig. 2 show that organisms from the same taxonomic groups form clusters. Eukaryotes are isolated from both bacteria and archaea, and the latter also form well-defined groups. The only exception is a group of four Mycoplasmatales, which are not aligned with the rest of bacteria, but form an isolated cluster. However, these organisms are indeed significantly different from other bacteria in the study, being parasites having small genomes, small networks, and lacking cell wall.

Profiles based on dot product of reciprocals of edge degree values (top right panel in Fig. 2) and on dot product of Boolean values indicating if the edge has alternative pathways, that is, whether its range is infinite (top left panel

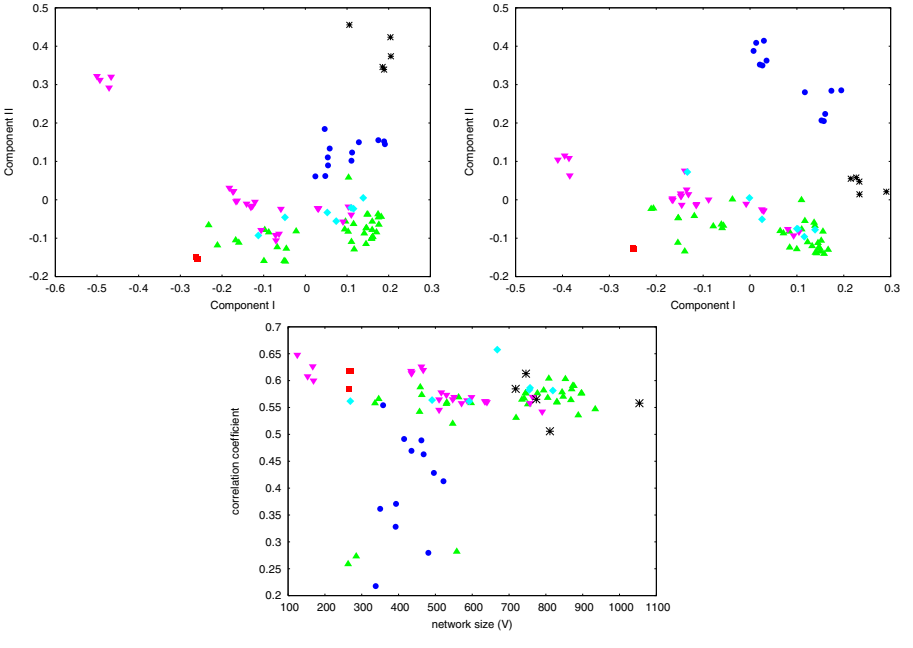

Fig. 2. Classical metric MDS visualisation of similarity of profiles based on infinity of edge range (top left) and reciprocal of edge degree (top right). Correlation coefficients between edge degree and reciprocal of edge range (bottom). Species denoted as in Fig. 1.

in Fig. 2), show similar behaviour. This indicates that both these properties are conserved across corresponding edges in closely related species. For edge range, the conservation in profile values is the more striking as it involves not only the immediate neighbourhood of the removed edge. Indeed, as shown in top right panel of Fig. 1, the alternative paths involve on average four steps for prokaryotes, or ten for eukaryotes. Similarity of results for edge degree and edge range is partially explained when correlation between the two measures across edges in the network are considered (bottom panel in Fig. 2). Smaller correlation for archaea is consistent with change of position of this group of organisms relative to eukaryotes and bacteria when comparing the results for the two essentiality measures.

The importance of conservation of edge degrees and of the existence of alternative pathways is put into perspective by studies relating the two quantities to experimental results on organism lethality. The connection between enzyme essentiality for the organism and the existence of alternative pathways linking substrates and products of the reaction catalysed by the enzyme has recently been explored [6] for *S. cerevisiae*. Reaction was deemed essential if knock-out of at least 75% of genes responsible for it was lethal. All analysed essential reactions were shown to correspond to edges in the graph with no alternative paths connecting one of the substrates with one of the products. Moreover, majority of essential reactions were outside the giant component of the network, showing the essentiality of connections at the periphery, not only at the centre, of metabolic networks. A study involving *E. coli* showed that within genes corresponding to edges without alternative pathways, the ones that lead to loss of the connection to larger group of metabolites from the rest of the network are more likely essential for cell growth [16]. The distribution of the sizes of disconnected parts follows a power law, with enzymes that, when deleted, leave only one metabolite unavailable being most frequent. As for edge degree, the importance of edges connected to vertices of low connectivity, not only those connected to hubs, for functioning of metabolic networks was shown using in silico models for predicting lethality of gene deletion in *E. coli* and *S. cerevisiae* [7].

4 Edge Influence on Global Efficiency of Metabolic Networks

We measured the global efficiency E^{eff} for metabolic network of each of the 89 organisms. First, we analysed if the efficiencies depend on the size of the networks. As shown in the top left panel of Fig. 3, while the values of E^{eff} vary over an order of magnitude, between 0.005 and 0.03, no consistent influence of network size is present within bacteria, archaea or eukaryotes. However, the efficiencies of the analysed eukaryotes are lower than of bacteria of similar network size, with average and standard deviation E^{eff} values of 0.0109 ± 0.0047 and 0.0172 ± 0.0041, respectively. For archaea, the mean is 0.0099 ± 0.0025. This is in agreement with shorter average path lengths of bacteria reported previously [4].

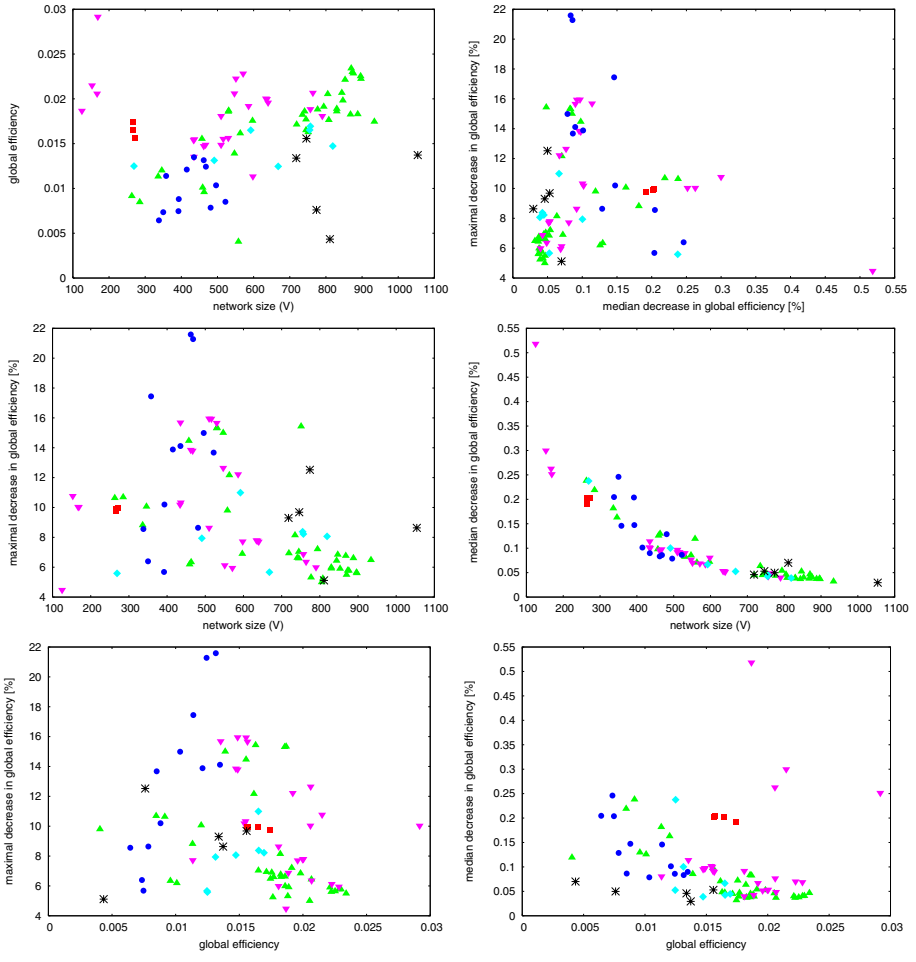

Fig. 3. Properties of global efficiency and its decrease after edge removal. Species denoted as in Fig. 1.

We analysed the impact of edge deletion on relative decrease in global efficiency. We measured the maximal, as well as median drop in efficiencies associated with removing a single connection between a substrate and a product. The results, presented in middle left and right panels of Fig. 3, respectively, indicate that susceptibility of the network to this type of perturbation is not consistent within species groups. However, a trend of decreasing effect of the perturbation, in particular of median decrease in efficiency, with the increasing size of the network is present. This reflects the fact that within smaller networks, lower number of edges is present, and thus each edge makes a larger contribution to E^{eff}, as global efficiency does not increase significantly with network size. With the exception of four Mycoplasmatales, which have very small genomes, with network sizes below 200 nodes, a trend of lower median ΔE^{eff} for more efficient

networks is observed within all three groups of species, indicating that high efficiency involves more redundancy against random failures in the network (Fig. 3, bottom right). However, no such trend can be observed when maximal relative drop in efficiency is measured against network global efficiency (Fig. 3, bottom left). In fact, no clear correlation is observed between maximal and median drops in efficiencies across species (Fig. 3, top right).

We analysed if the profile of ΔE^{eff} for particular reactions is preserved across related organisms. We used dot product between the profiles to measure similarity. The results, visualised with classical metric MDS in the left panel of Fig. 4, show lack of cluster structure that would group evolutionarily related species based on profiles of efficiency decreases for edges within each organism. This is the more striking when compared with similar results of profiles resulting from reaction content of organisms in the right panel of Fig. 4, where the profiles are binary values indicating the presence or absence of a reaction in the metabolic network of given organisms. For such profiles, as it has been shown previously [17,18], evolutionarily close organisms tend to form visible clusters. That is, when substituting ΔE^{eff} values for all edges present in a network with a constant to obtain binary edge presence profiles, the clustering improves significantly, showing that ΔE^{eff} distorts rather than promotes clustering.

The results indicate that removing corresponding substrate-product connections in metabolic networks of organisms from the same taxonomic group results in varying values of ΔE^{eff}, the relative decreases in global efficiencies of their metabolic networks. This finding shows that while global efficiency is similar in related organisms, values of ΔE^{eff} are not good candidates for surrogate measures in identifying the most vulnerable enzymes across related species, for example for identifying potential targets for new drugs that target metabolism, such as some antimicrobial agents. Global efficiency based only on the topology of the network may not be accurate in this respect because it relies in transport on only the shortest paths. Also, the shortest paths are measured simply as the number of reactions that are needed to transform one metabolite into another, which does not take into account the kinetics of the reaction nor the

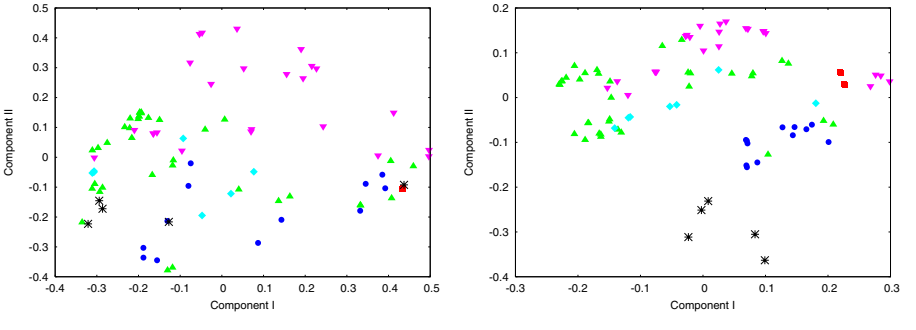

Fig. 4. Classical metric MDS visualisation of similarity of profiles based on relative decrease in global efficiency (left) and on reaction content of metabolic networks (right). Species denoted as in Fig. 1.

fluxes through them. Still, even the fluxes may be different in different conditions for the same organism. Moreover, in *E. coli*, lethality is not correlated with the estimated level of flux through the reaction [19]. In local measures of edge importance, simplifications in modelling may not influence the results as heavily. Furthermore, in the analysis involving edge range, the purely topological feature of infinite range is definitive and translates into no flux between two metabolites.

5 Conclusion

Effects of inhibition or knock-out of enzymes that catalyse reactions in metabolic networks are widely analysed, with implications ranging from determining gene function to identifying targets for drug design. With the latter application in mind, we analysed here the similarity across species of three measures of edge importance in metabolic networks. We show that the two analysed local measures, which were proven previously to correlate with experimental data on lethality, are conserved across related species. The results indicate that these biologically meaningful measures are suitable for identifying groups of enzymes that would share high essentiality within a group of related bacteria, but would have different, lower essentiality in their eukaryotic hosts. On the other hand, we show that a measure that goes beyond local neighbourhood in analysing the effects of edge removal does not lead to profiles more consistent among closely related species than among distant ones. Thus, of the previously suggested measures of edge essentiality, only the local ones are discriminative in addition to being informative in the context of metabolic networks.

Acknowledgements

The author acknowledges Polish Ministry of Science and Higher Education grant 3T11F01030 and AGH grant 11.11.120.777, and thanks Dr. W. Dzwinel and Mr. Marcin Kurdziel for comments on the manuscript. The author also acknowledges the creators of the Matlab BGL library, which was used in computations.

References

1. Jeong, H., Tombor, B., Albert, R., Oltvai, Z., Barabasi, A.: The large-scale organization of metabolic networks. Nature 407, 651–654 (2000)
2. Barabasi, A., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
3. Crucitti, P., Latora, V., Marchiori, M., Rapisarda, A.: Efficiency of scale-free networks: error and attack tolerance. Physica A: Statistical Mechanics and its Applications 320, 622–642 (2003)
4. Ma, H., Zeng, A.: Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19, 270–277 (2003)
5. Zhu, D., Qin, Z.: Structural comparison of metabolic networks in selected single cell organisms. BMC Bioinformatics 6, 8 (2005)

6. Palumbo, M., Colosimo, A., Giuliani, A., Farina, L.: Functional essentiality from topology features in metabolic networks: A case study in yeast. FEBS Letters 579, 4642–4646 (2005)
7. Mahadevan, R., Palsson, B.: Properties of metabolic networks: Structure versus function. Biophysical Journal 88, 7–9 (2005)
8. Kaiser, M., Hilgetag, C.: Edge vulnerability in neural and metabolic networks. Biological Cybernetics 90, 311–317 (2004)
9. Agoston, V., Csermely, P., Pongor, S.: Multiple weak hits confuse complex systems: A transcriptional regulatory network as an example. Physical Review E 71, 51909 (2005)
10. Csermely, P., Agoston, V., Pongor, S.: The efficiency of multi-target drugs: the network approach might help drug design. Trends in Pharmacological Sciences 26, 178–182 (2005)
11. Watts, D.: Small worlds. In: The dynamics of networks between order and randomness. Princeton Univesity Press, Princeton (1999)
12. Latora, V., Marchiori, M.: Efficient behavior of small-world networks. Physical Review Letters 87, 198701 (2001)
13. Vragović, I., Louis, E., Díaz-Guilera, A.: Efficiency of informational transfer in regular and complex networks. Physical Review E 71, 36122 (2005)
14. Holme, P., Kim, B., Yoon, C., Han, S.: Attack vulnerability of complex networks. Physical Review E 65, 056109 (2002)
15. Ma, H., Zeng, A.: The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 19, 1423–1430 (2003)
16. Lemke, N., Heredia, F., Barcellos, C., dos Reis, A., Mombach, J.: Essentiality and damage in metabolic networks. Bioinformatics 20, 115–119 (2004)
17. Ma, H., Zeng, A.: Phylogenetic comparison of metabolic capacities of organisms at genome level. Molecular Phylogenetics and Evolution 31, 204–213 (2004)
18. Podani, J., Oltvai, Z., Jeong, H., Tombor, B., Barabási, A., Szathmáry, E.: Comparable system-level organization of Archaea and Eukaryotes. Nature Genetics 29, 54–56 (2001)
19. Ghim, C., Goh, K., Kahng, B.: Lethality and synthetic lethality in the genome-wide metabolic network of escherichia coli. Journal of Theoretical Biology 237, 401–411 (2005)