Structure of Metabolic Networks

Saumya Bhadani(201301100), Nidhi Pitroda(201301404), Jahnavi Suthar(201301414) and Deeksha Koul(201301435)

***Abstract*—Metabolic networks represent the biochemical reactions that take place in a cell or microorganism. In this paper, we have attempted to determine its structure. We have worked with metabolic networks of 43 organisms and they show properties of both modular and scale-free distributions. To resolve this contradiction we have proposed a 'hierarchical' distribution, which aptly explains all our observations.**

I. Introduction

The metabolism of an organism is formed by a series of biochemical reactions which are very important to sustain life, as they are responsible for the production of amino acids, sugar etc. A large number of metabolic substrates participate in these reactions and their interactions can be demonstrated using complex network. In this network, also referred as the metabolic network, the substrates can be nodes and the reaction itself can be the link[1]. Using the metabolic network, valuable insights can be gained regarding the metabolism of an organism and the functioning of cells.

The main objective of our project is to determine the structure of the metabolic network. We discovered that the metabolic networks show properties of both modular networks and highly integrated module free(scale free) networks. Thus, the structure is not immediately apparent as the modular and scale-free structures are quite contrary to each other. To explain these observations, we prove that the modules in the metabolic network actually follow a hierarchical organization[2][3].

We have studied the metabolic networks of 43 different organisms based on the data from Barabasi group[4]. We will first show that these networks demonstrate properties of both modular and scale-free networks and subsequently, we will explain how a hierarchical organization can elucidate all these properties.

II. Description of Metabolic Network

The metabolic network describes the biochemical reactions taking place in the cells of an organism[1]. The nodes of this network are substrates that participate in the reaction. There is a directed edge from a reactant substrate to its product. A product of one reaction can further act as a reactant, responsible for the formation of some other product. The physical entity of the edge is actually a temporary educt-educt complex, which catalyzes the reaction[1].

A. Determination of Structure of Metabolic Network

In this section, we will determine the probable structure of metabolic networks using the networks of 43 organisms.

1) Power Law Degree Distribution: The almost linear curve of degree distributions on the log-log scale in Fig.1 indicates that they follow power law. The power law degree distribution is a property of scale-free topology and thus the metabolic networks of all 43 organisms should also be scale-free. This degree distribution suggests the existence of few highly connected nodes called hubs, which participate in a large number of reactions.

The scale free degree distribution of out degree indicate that there are some small number of substrates which are responsible for large number of biochemical reactions. Similarly, due to scale free distribution of in-degrees, there must be some substrates that can be formed by many reactions.

2) Uniform Diameter: From Fig.2b, we can observe that diameter of metabolic networks of all 43 organisms is almost same, irrespective of number of nodes. This offers evidence that these networks are scale-free. In scale free networks, new nodes are preferentially attached to some preexisting hubs, thus the addition of edges does not increase the diameter of the graph.

Another evidence of scale free model is random and targeted removal of nodes. If hubs are removed, there would be a sharp increase in the diameter of the network. But the removal of any other node may not have any effect on the diameter or the connectivity. From Fig.2c, we observe that metabolic networks have a similar effect on targeted and random removal of nodes.

These two sections clearly indicate that the metabolic networks follow a scale free distribution.

3) Existence of Modularity: In the above two sections, we arrived at a conclusion that metabolic networks should be scale-free. A property of scale-free networks is the absence of highly connected components or communities. This is due to the presence of hubs. Hubs integrate

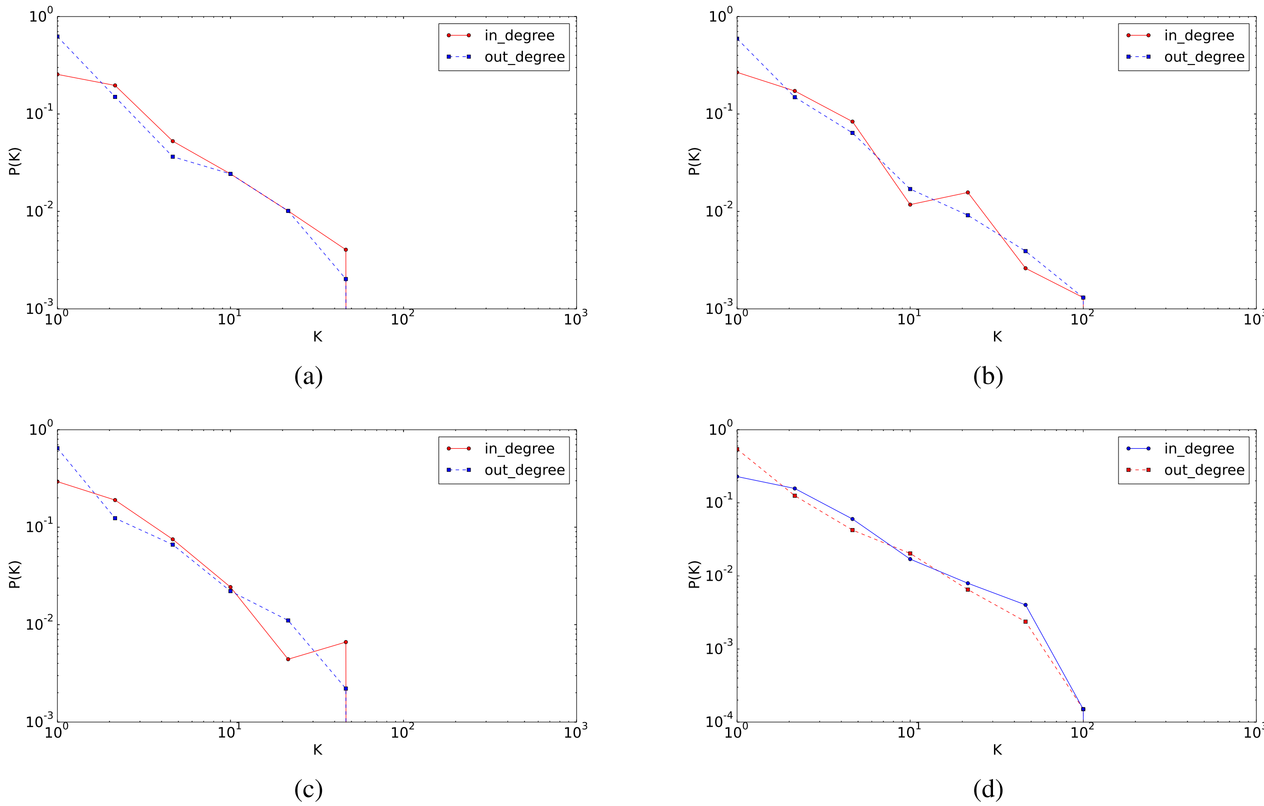


Fig. 1: Degree Distribution for a)Archaeoglobus fulgidus (archae);b) E. coli (bacterium); c) Caenorhabditis elegans (eukaryote), shown on a log log plot, counting separately the incoming (In) and outgoing links (Out) for each substrate.d)Degree Distribution averaged out over all 43 organisms, The figures indicate Power Law Distribution of degrees, Very few nodes with very high degree known as hubs are present, showing characteristic of scale free network

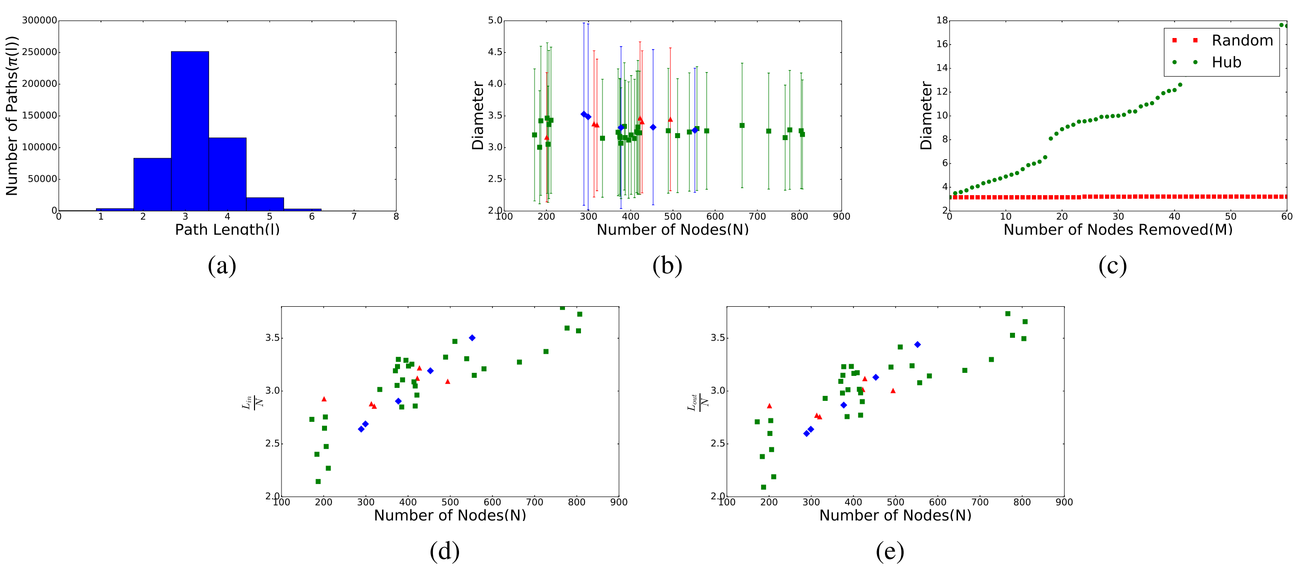


Fig. 2: a)Histogram of shortest path lengths in E-coli; b)Diameter(Average shortest path length)s of all 43 organisms remains almost constant with increase in number of nodes,the error bars show the standard deviation of shortest path length; c)Diameter of the E-coli after removal of M nodes,it remains constant for random nodes removed, increases rapidly with removal of hubs; d/e)Average Incoming/Outgoing Links, increases with increase in number of nodes, Color code denotes, green=Bacterium,red=Archae,blue=Eukaryote

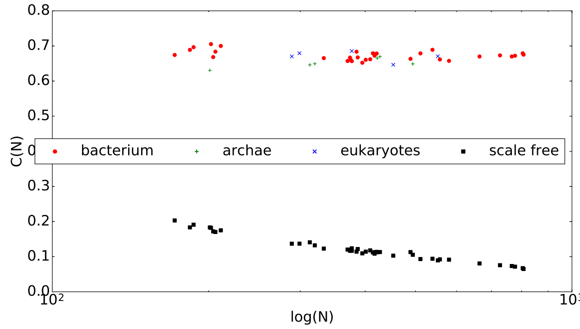
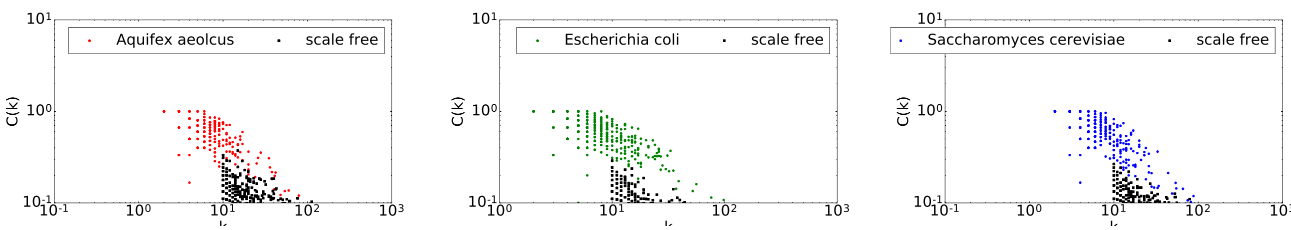


Fig. 3: The average clustering coefficient for 43 organisms as a function of nodes N. Species belonging to bacteria(red circles) archae(green cross) and eukaryotes(blue X) are shown. The blue squares indicate average clustering coefficient for scale-free networks for similar N. The clustering coefficients for the species is quite high than corresponding scale-free models and they remain almost constant as N increases.



(a) Archae

(b) Bacteria

(c) Eukaryotes

Fig. 4: The dependence of clustering coefficient on degree is shown for three class of organisms. The clustering coefficient decreases as degree of nodes increase in these organisms.

all the nodes in almost a single web, thus the presence of communities or modules is highly improbable. But, from Fig. 3, we can see that clustering coefficient of all 43 metabolic networks is quite high compared to corresponding scale-free networks. High clustering coefficient is a property of modular organization. Also, from Fig 3, clustering coefficient of metabolic networks remain almost constant when N is increased, while for scale-free distribution, clustering coefficient decreases linearly, with increase in N. Thus, through observations of clustering coefficient, metabolic networks should have modular organization, but a clear-cut modular network cannot have scale-free degree distribution.

Thus, apparent contradiction arises due to above observations. The high, size independent clustering coefficient supports modular structure while the power law degree distribution supports scale-free structure. We need a model which encompasses both modularity and power law degree distribution. One such model is the Hierarchical Network[1].

B. The Hierarchical model for metabolic networks

Consider a simple hierarchical network[1]. The starting point in this network is four densely connected nodes. Generate three replications of this module and

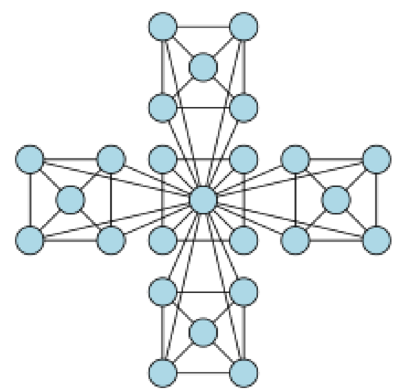


Fig. 5: A simple hierarchical model. This network has four highly connected modules connected in a hierarchical way. The node in the center has the highest degree as it is also connected to peripheral nodes in the outer four modules. The clustering coefficient of all the nodes is almost 0.6

connect three external nodes(one from each replicated module) to the centre of the previously created module, obtaining a 16-node module[1]. Again create three replication of the 16-node module and connect the peripheral nodes to the center of the existing module[1]. If we continue this, we obtain a hierarchical model 5.

The degree distribution of this model would clearly follow the power law. The central node would act as a hub 5 because as we increase the modules, we connect the peripheral nodes to the central node only. The degrees of almost all other nodes would be small. The degree exponent of this network would be 2.26(1+ln(4)/ln(3)) and the degree exponent obtained from metabolic networks is almost 2.2.

Also, this model in manifestly modular with clustering coefficient almost equal to 0.6, again in agreement with metabolic networks. The clustering coefficient would not change in this model as number of nodes are increased, as all the nodes would be part of some highly connected module.

Another property of the metabolic networks observed from the networks of 43 organisms, is the dependence of clustering coefficient on degree as observed from Fig.4. This again is an evidence of hierarchical organization as the clustering coefficient of all nodes is almost uniform, but degree is not uniform. There would be an inverse relationship between clustering coefficient and degree for hierarchical networks and that is clearly being depicted in Fig 4 for metabolic networks as well. Modular and scale-free models predict clustering coefficient of nodes to be independent of their degrees.

C. Results and Discussions

All the properties observed using metabolic networks of 43 organisms, power law degree distribution, probability of few hubs, high and size independent clustering coefficient and inverse relationship between clustering coefficient and degree of nodes, is explained by the hierarchical organization.

From a biological perspective, we can say that cellular metabolism is divided into different functional modules, but those modules are not independent of each other, but are related in a hierarchical manner, with one functional module building up on other. The hierarchical model is also consistent with the idea of evolution of cells. Some complex functional modules may be created through simple modules due to hierarchy in the metabolism of the cell.

D. Conclusion

In this project, we have studied a very important class of networks called metabolic networks. These networks represent the biochemical reactions that take place in a cell or a microorganism, and which are responsible for creating energy, keeping organisms and microorganisms alive. Thus, a deep understanding of this network can help us uncover truths regarding life itself.

Through this project, we have tried to determine the structure of the metabolic network. We studied the metabolic networks of 43 different organisms belonging to bacterium, archae and eukaryote classes and reached a contradiction. These networks showed properties of both modular and scale-free topology. We then concluded that the metabolic network is actually modular but the modules are connected in a hierarchical order. This hierarchical model of the metabolic network was able to explain all our observations.

We also tried to acknowledge the significance of the hierarchical model from the biological perspective.

Although we have looked only at metabolic networks in this study, there can be many other networks(both biological and non-biological) which can combine both scale-free and community structure, and this hierarchical model can be considered for those networks as well.

References

[1] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A.-L. Barabasi, "The large-scale organization of metabolic networks."

[2] E. Ravasz, A. Somera, D. Mongru, Z. Oltvai, and A.-L. Barabasi, "Hierarchical organization of modularity in metabolic networks."

[3] A.-L. Barabasi, "Network science."

[4] "Database of metabolic networks of 43 organisms," http://www3.nd.edu/~networks/resources/metabolic/index.html.