# Topological Properties of Protein-Protein and Metabolic Interaction Networks of *Drosophila melanogaster*

Thanigaimani Rajarathinam and Yen-Han Lin*

*Department of Chemical Engineering, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5A9, Canada.*

The underlying principle governing the natural phenomena of life is one of the critical issues receiving due importance in recent years. A key feature of the scale-free architecture is the vitality of the most connected nodes (hubs). The major objective of this article was to analyze the protein-protein and metabolic interaction networks of *Drosophila melanogaster* by considering the architectural patterns and the consequence of removal of hubs on the topological parameter of the two interaction systems. Analysis showed that both interaction networks follow a scale-free model, establishing the fact that most real world networks, from varied situations, conform to the small world pattern. The average path length showed a two-fold and a three-fold increase (changing from 9.42 to 20.93 and from 5.29 to 17.75, respectively) for the protein-protein and metabolic interaction networks, respectively, due to the deletion of hubs. On the contrary, the arbitrary elimination of nodes did not show any remarkable disparity in the topological parameter of the protein-protein and metabolic interaction networks (average path length: 9.42±0.02 and 5.27±0.01, respectively). This aberrant behavior for the two cases underscores the significance of the most linked nodes to the natural topology of the networks.

Key words: topology, *Drosophila melanogaster*, scale-free network

## Introduction

The concept of networking and its omnipresent nature has been an issue of considerable curiosity, captivating scientists worldwide over the past few years. Simple networking can be witnessed in the physical connectivity of computer terminals or routers. On broader perspective, many networks, such as social ties—familial and professional, World Wide Web, network of scientific papers connected by citations, electrical power grids, transportation systems, and biological networks, are all illustrations of real-world complex systems (1). The exploration of this intrinsic blueprint would provide an in-depth examination of the major factors that cause the system to act in their distinguishing modes, and identify the nodes that play a pivotal role in the topology of the system.

The random network theory was the base for all computations about systemic pathways and distribution modeling, until the scale-free network model was proposed in 1998 (2-4). The scale-free network model falls between the classes of regular (p=0) and random

($p$=1) networks, where $p$ is the probability of finding a random distribution (5, 6). This modification results in a transitional ground with a probability $p$ between 0 and 1. In this type of network, a few nodes that have a large number of links, known as hubs, tend to dominate the rest of the nodes having smaller number of links. A notable feature of a scale-free network is that a node would require only a few steps to reach another node. The distribution of nodes in such a network declines with an increase in the number of links and is found to decay as a power law as given by the relation, $p(k) \sim k^{-\gamma}$, where $p(k)$ signifies probability, $k$ represents a specific number of links, and $\gamma$ is the exponent of the power law (7). The plot of the probability distribution $p(k)$ on a log-log scale gives a linear correlation, establishing the fact that the distribution follows a power law.

Although scale-free property is a ubiquitous phenomenon, it can not be truly called a universal one. The random network theory of Erdős and Rényi (2-4) still holds true for some structures like the power grid system of Western United States (8-10) and the graph of company directors (11) that appear to have

*Corresponding author.
E-mail: yenhan.lin@usask.ca

degree distributions with a wholly exponential tail.

Jeong et al (12) investigated the metabolic network structure of 43 organisms and reported that the average path length of the network for all studied species had a similar value. Removing the current metabolites, Ma and Zeng (13) analyzed 80 sequenced genomes including parasites, prokaryotes, and eukaryotes, which revealed that the average path length of these species could be statistically grouped into three values. They ascertained that the more complex the organism, the longer the average path length. From a biological standpoint, cell biology with the affiliation between its different components has become a focus to the scientific community as it seeks to unravel the organizing principles governing the pattern and growth of the network mechanism in cellular and molecular networks.

Although all organisms have a similar fundamental structure, the quantitative diversity in their metabolic network architecture can be differentiated by the topological parameter of the metabolic network. This diversity echoes the various evolutionary cycles that each organism has undergone over time. The differing topological parameter for the three domains indicates the mixed compactness and centrality of the metabolic pathways (14). A better depiction can be derived by exploring the reactions and pathways of an individual organism to gain a concrete understanding of the biological significance of the basic structural variations that subsist in each organism.

The assignment of gene function to newly sequenced genes as part of genome projects would not be feasible without the trappings to recognize the similarity in amino acid sequences. The ability to appreciate the nature of protein evolution allows the biotechnologist to develop novel and technologically useful proteins in vitro. The presence of some proteins with a large number of interactions may be due to a specific structural composition that is different from other less connected proteins (15). This perception is central to the design of long-lasting immunizations and associated drug treatments for human diseases.

Proteins may have either direct or indirect interactions with one another. In a direct or physical interaction, two protein chains bind to each other. Indirect association refers to proteins being a member of the same functional module (for example, transcription initiation complex and ribosome). A protein of this nature may not directly bind to another protein. Several of these interactions echo the dynamic state of the cell, and their existence depends on the partic-

ular environment or developmental status of the cell. However, the coupling of existing and potential interactions together defines the protein-protein interaction network within the genome of a given organism.

Several groups reported systems for analyzing interactions in the protein-protein network. Ito et al (16) provided a two-hybrid screening method that facilitated the publication of a comprehensive protein-protein interaction map for yeast. Flores et al (17) published a protein-protein interaction map of yeast RNA polymerase III using the two-hybrid technique. The topological analysis of the protein-protein interaction network of Saccharomyces cerevisiae showed its heterogeneous scale-free architecture (18). Jeong et al (19) considered proteins in the protein-protein interaction network as nodes and the interactions between them to represent the links. The computational elimination of the well-connected yeast proteins caused the diameter of the network to increase steadily. On the contrary, the exclusion of arbitrarily chosen yeast proteins did not affect the topological parameter of the network, which has been shown to be in agreement with the result from mutagenesis experiments (19).

The lethality of a protein was found to depend on the number of connections it shared in the protein-protein interaction network. The proteins with a large number of connections were found to be highly essential, and their exclusion disturbed the network topology, proving lethal to the network. Although there were several proteins with smaller number of links, their elimination did not produce such adverse consequences. Jeong et al (19) further concluded that the highly connected proteins that play a central role in the network design were three times more essential than the proteins that interact with only a few other neighbors.

Mering et al (20) evaluated large-scale protein-protein interaction datasets to determine their precision and to recognize their biases, strengths, and weaknesses. They compared the significant approach methods to one another and to a reference group of formerly identified interactions. Bu et al (21) examined the topological formation of yeast, that is, quasi-cliques and quasi-bipartites, using a spectral analysis method. They concluded that the unknown topological configuration consists of biologically significant functional groups.

However, all endeavors to study the network design of biological organisms have been restricted to simple species (18, 22). A valid reasoning behind this predisposition is that the smaller the number of genes

involved, the fewer are the complications to be encountered during the study. Research communities around the world furnish an abundance of datasets on a regular basis, but the specific analyses of these datasets are underdeveloped due to the compound environment of the biological interaction networks. The nature of research is multifaceted due to the fact that even the simplest unicellular organisms have more than a few hundreds of genes transforming them into a compound life form from the perspective of a data analyst. Nevertheless, the shift of focus to more complex beings would provide a better understanding of the structural properties prevailing in the higher echelons of nature's offspring.

This study seeks to extend this widespread trend to the multi-cellular organism *Drosophila melanogaster*, which has been lending itself well to behavioral studies for almost a century due to its similarity with the human proteins (*23*). Scientists have discovered an identifiable match between the genetic code of fruit flies and over 60% of known human disease genes. Moreover, about 50% of fly protein sequences are believed to have mammalian analogues. *D. melanogaster* is being used as a genetic model for several human diseases including Parkinson's and Huntington's diseases, and the study of these proteins presents knowledge about the possible development of remedial measures for human diseases such as heart disease, cancer, or diabetes mellitus.

# Results

## Analysis of the protein-protein interaction network

A scale-free network is distinguished from the random network design by a power law distribution of connectivity, instead of an exponential curve. The graphical plot involving the probability of allocation of nodes and their definite number of links should produce a decreasing function for a network that is scale-free.

The probability distribution plot of the number of proteins with a distinct number ($k$) of interactions is an excellent indicator of the type of network design inherent in the protein-protein interaction network. As shown in Figure 1, using the high confidence interaction dataset (see Materials and Methods), a plot of probability distribution $p(k)$ versus the interactions $k$ on a log-log scale produced a linear correlation, indicating that the probability distribution follows a power law.

This result is closely in accordance with the plot shown by Giot *et al* (*24*) for the protein interaction map. As the degree of the network increased, the probability of finding a protein with that specified number of interactions began to fall to a low value. This confirmed the salient feature of scale-free networks that there are several nodes with a low degree and a few dominant nodes (hubs) with a high degree. This result verified that the protein-protein interaction network of *D. melanogaster* follows a scale-free architecture.
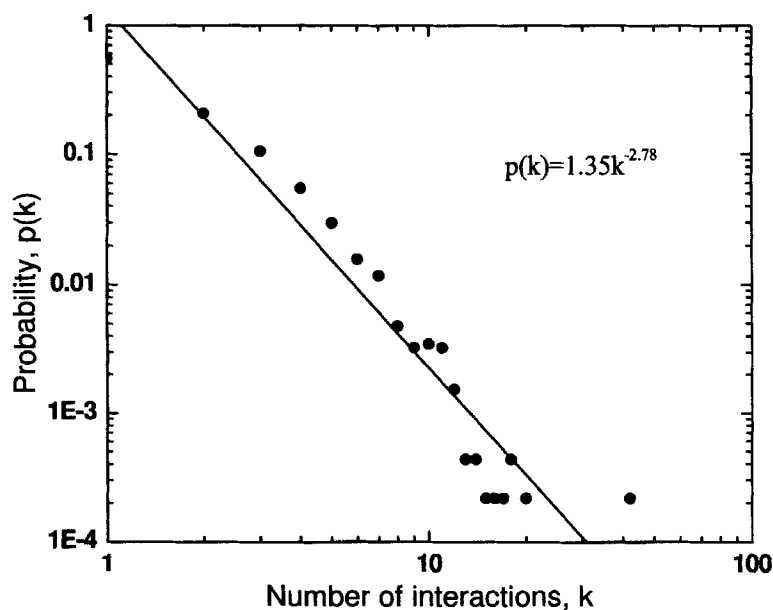


$$p(k) = 1.35k^{-2.78}$$

**Fig. 1** The probability distribution plot for the protein-protein interaction network.

There were 2,549 proteins having a single interaction, constituting more than half (~55%) of the total proteins (4,591) in the high confidence dataset. Only 1% of the total proteins had 10 or more interactions, and these proteins became the hubs of this protein-protein interaction network. The exponent in the power law relation was found to be 2.78. This value was in concurrence with those obtained from other networks (2.0–3.0) like the World Wide Web, citations of scientific articles, and the Internet that follow a scale-free architecture (7).

The Boost Graph Library was utilized to determine the shortest path length between every pair of accessible proteins in the interaction network (25). The average path length, that is, the number of pathways it would require on average for one protein to reach another accessible protein in the interaction network, was calculated as 9.42 for the high confidence interaction dataset (24).

The relatively short path length typifys a small world network. As the network architecture is of a scale-free nature, it has to be tested for its resistance against random failures and vulnerability to coordinated attacks. The proteins were ranked based on their degrees in decreasing order for this study. The exclusion of a particular node by chance or by deliberate measure can give rise to contrasting consequences. The severity of these consequences depends entirely on the node that is removed. With the rationale of examining the network for its weakness against sequential attacks and tolerance to random failures, a milieu simulating a targeted assault where the well-connected proteins (top 3%) were removed in sequential order of their degrees was replicated. The second part of the simulation created a random node failure by removing proteins in an arbitrary fashion. The upshot of those simulations is summarized below.

The consequences of the removal of the highly connected and random proteins are illustrated as a plot of the average path length versus the number of removed proteins (Figure 2). The plot accentuates the critical nature of the highly connected proteins, when 3% (~138) of the well-connected proteins were removed sequentially in decreasing order of their degrees. For the 138 proteins eradicated, the average path length of the network was determined to be 20.93. In comparison with the original topological parameter (average path length = 9.42) of the arrangement, it can be seen that the topological parameter of the network has doubled. This offers evidence of the vital quality of the highly connected proteins.

The elimination of hubs altered the topology of the network, corroborating their direct relation to the topological parameter that proves lethal to the system. The boost in the average path length signifies that the shortest path length required by a particular protein to get to another protein has increased. This implies that it takes more number of steps to get from a specific protein to its target protein, thereby disturbing the efficient, innate path that was utilized before the assault on the hubs of the network. This loss resulted in the formation of small and secluded
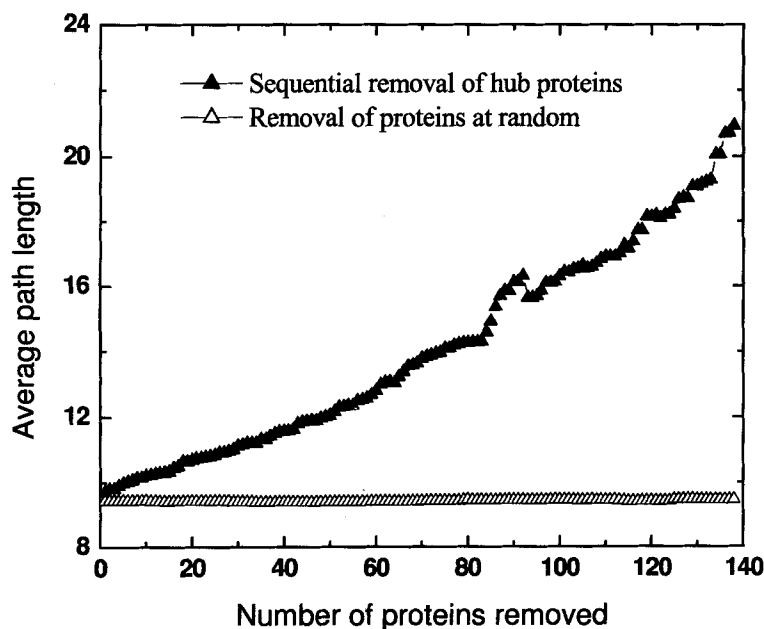


**Fig. 2** Effect of sequential and random removal of proteins on the average path length of the protein-protein network.

clusters of proteins, very different from the original compact system design.

On the contrary, the random elimination of proteins did not show any significant variation in the topological parameter value. The simulation of a random node failure only showed a minor variation in the topological parameter of the network, when 3% of the proteins were removed. As the plot (Figure 2) indicates, it did not affect the innate topology of the system.

The top 3% (~138) of the highly connected proteins were examined to study their functions and their respective roles in biological pathways. Information retrieved from FlyBase provided the molecular function of these proteins (26). It was found that 62 of these proteins have not been annotated yet. The study of the remaining 76 proteins with an annotation using KEGG Orthology (27, 28) revealed that eight proteins could not be classified to a specific pathway. Analysis showed that 21 proteins play a part in metabolism, 10 proteins are involved in cellular processes, 20 proteins are responsible for genetic information processing, and 17 proteins serve in environmental information processing. Among the proteins involved in metabolism, 13 proteins are components of the central metabolism pathways. Hence, it can be seen that the molecular functions of the annotated hubs are distributed over a wide array of processes essential for the functioning and sustenance of the organism.

## Analysis of the metabolic network

The frequency distribution of the compounds involved in the reaction mechanism, classified based on their links (Figure 3), proves that the metabolic interaction network of *D. melanogaster* follows a scale-free topology.

The plot shows that there were a high number (~70) of metabolites (substrates and products) that took part in five or fewer reactions. As the number of links increased, the compounds involved in the reactions began to decrease steadily. To make the graphs clear, the number of links greater than 30 was included in one bar. When the plot was stretched out for all values of links, the number of metabolites involved declined progressively to reach a value of one and remained constant. This emphasized the fact that at higher degrees, there was only a solitary metabolite contributing to such many numbers of reactions. From the computations, it was established that there were 597 compounds that acted as substrates and 611 compounds that occurred as products. Of this mix, 282 compounds took part as both substrates and products.

The shortest pathway was computed for all pairs of metabolites that could be reached. The removal of current metabolites yielded 926 compounds with 3,326 connections between them. From the probability distribution plot for the reformed dataset shown in Figure 4, it can be deduced that the distribution
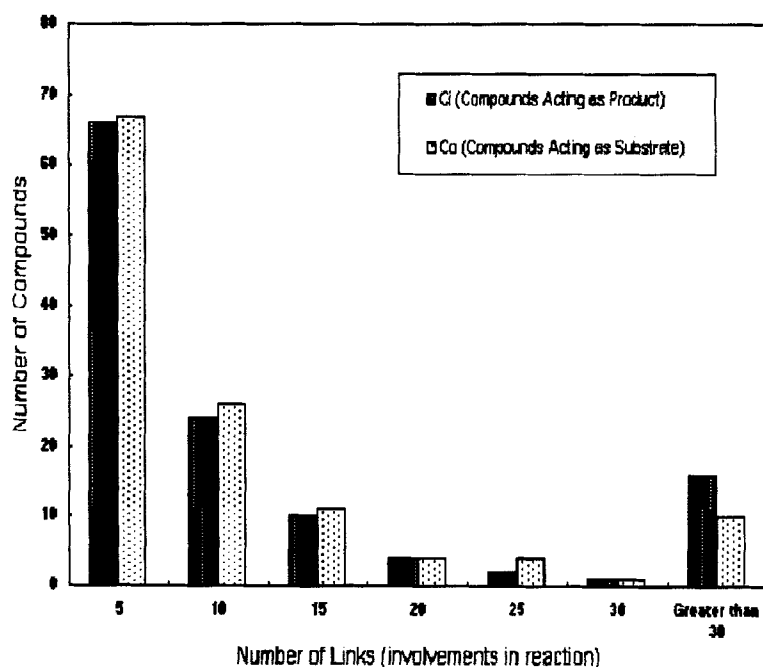


**Fig. 3** Frequency of interaction of compounds involved in the metabolic pathways.
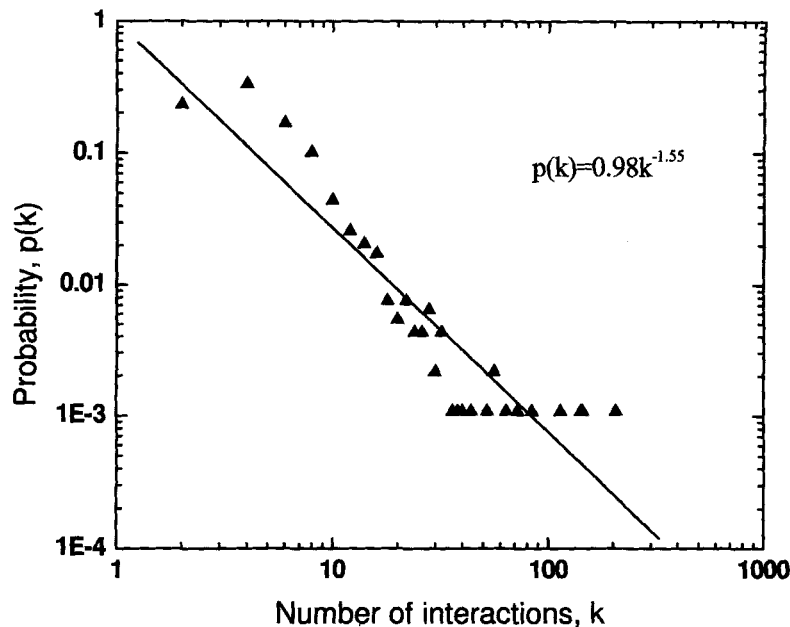
$$p(k)=0.98k^{-1.55}$$

**Fig. 4** The probability distribution plot for the metabolic interaction network.

still decays as a power law, indicating that the network follows a scale-free pattern.

The results indicated that over 56% of the metabolites were involved in just one or two interactions, while only about 2% of the metabolites participated in 30 or more interactions—a classic trait of scale-freeness. The exponent of the power law relation was found to be 1.55. This value was smaller than those obtained from other networks like the World Wide Web and the Internet that follow a scale-free architecture (1). Nevertheless, the intrinsic topol-. ogy of the complex network remained unaltered. The average path length for this network was calculated as 5.29. Hence, on average, each metabolite could be converted to any other metabolite (that can be reached or converted) in five steps (13). This showed that the metabolic network is highly compact as indicated by the small number of steps required to get from any one metabolite to another. This small world nature reinforces the fact that the metabolic interaction network of D. melanogaster pursues a scale-free architecture.

The simulation of coordinated and random attacks proved that the eradication of well-connected metabolites increases the average path length, altering the inherent architecture of the network. After the exclusion of 5% of the hubs from the network, the average path length was enhanced to 17.75 (Figure 5). It can be seen that the topological factor suffered a more than three-fold augmentation after the 5% removal of hub metabolites. The abnormal increase can be at-

tributed to the fact that the organism lost its competent shortest pathway, and a new set of pathways was organized. This created a drastic change in the network design, affecting the ability of the organism to produce a particular metabolite in a relatively small number of reactions.

Conversely, when the metabolites were eradicated in a random fashion (Figure 5), there was scarcely any change in the topological parameter. The elimination of any compound resulted in the search for a new pathway as the original pathway can not be traversed. For a random compound, the organism was able to choose an alternate pathway without affecting the topology of the network. The novel pathway would lead to the target metabolite in more or less the same number of steps as the original pathway, proven by the steadfast nature of the topological parameter. Therefore, the critical nature of the hubs can be witnessed by the radical alterations that occur due to their simulated elimination.

## Discussion

The high confidence protein-protein interaction dataset was used to find the network architecture and to determine other topological considerations. The dataset revealed an inherent scale-free architecture, thereby underlining the fact that the majority of complex biological networks have a scale-free topology. The shortest path length of all the accessible proteins
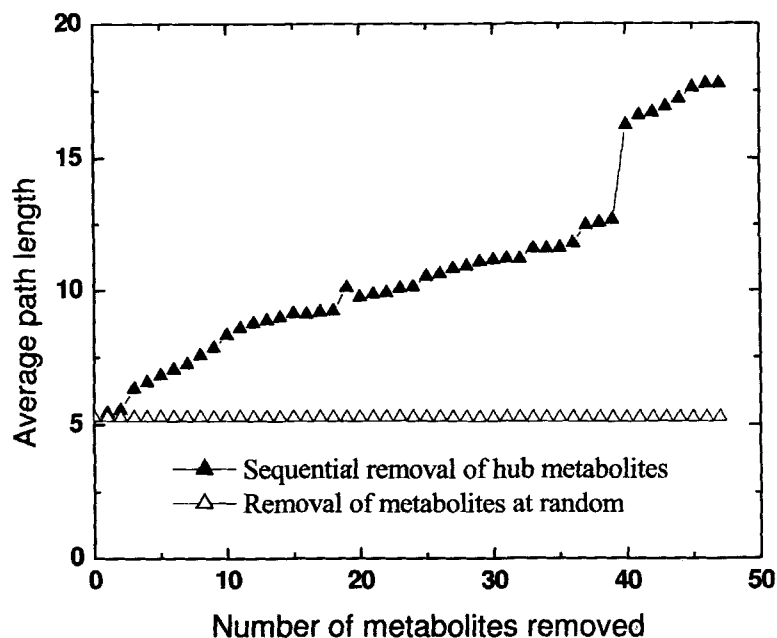
**Fig. 5** Effect of the sequential and random elimination of metabolites on the average path length of the metabolic network.

and the average path length of the network were also determined. The far-reaching alteration in the network topology (two-fold increase in the average path length) due to the exclusion of hubs of the network confirmed their essence to the network. The analysis of the molecular function of the well-connected metabolites proved that an extensive range of biochemical processes was managed by the hubs.

A striking feature that deserves investigation is the ability of the organism to transfer to an alternate connection mechanism most times, when a non-hub protein is removed, to form new links and entail the smooth progress of the formation of biochemical products. Despite the fact that the highly connected proteins were removed, the organism was able to sustain its routine tasks due to its ability to locate another protein with a function similar to the eradicated protein. This resilient nature of the organism needs further consideration. A physical denotation involving the strength of each interaction would provide a more comprehensive understanding of the network.

The study of the intrinsic doctrines that characterize the metabolic network of *D. melanogaster* provides a crucial understanding of the construction blocks of the organism. The evaluation of the metabolic network of *D. melanogaster* identified that 926 compounds were involved in the metabolic reactions as substrates, products, or both. The graphical plots showed that the probability distribution of these metabolites followed a scale-free design. The

frequency distribution of these metabolites also indicated that only a few hubs exist in the network that participated in 30 or more reactions.

The presence of current metabolites generated an artificially short pathway between any two metabolites and proved detrimental to the computation of the topological parameter. The elimination of these cofactor compounds helped to determine the realistic number of steps required for the conversion of one compound to another, generating pragmatic topological parameter values. The exclusion of the hub metabolites demonstrated the harmful effects on the system, since it disturbed the existing topology of the network. Their removal was lethal to the overall topology of the system, leading to the failure of the efficient innate pathways. Although the organism was able to find alternate pathways to form products, it required greater number of intermediate steps. The average path length suffered a three-fold increase due to the top 5% exclusion of the hubs. This shows that a hub metabolite can not be chosen as a target compound for therapeutic research; any interruption to the original pathway involving hubs may cause the system to be secluded into diminutive groups resulting in a larger number of steps to be utilized for the production of the same compound. In some cases, the mechanism may not be able to find any avenues for an alternate pathway due to the formation of small clusters of metabolites causing the loss of that specific product metabolite. A random metabolite with less

participation in reactions could have a better efficacy to serve as a target metabolite and resist intrinsic errors.

A weight-based interaction study can generate a better interpretation of the network. An understanding of the importance of each individual reaction can be offered because of such an analysis. The major hurdle in the assignment of weights to the interactions is the determination of the strength of each interaction. The influence of each interaction in the context of the web of interactions must also be considered. The binding forces that govern the interactive mechanism must be evaluated before a physical denotation can be provided to them. These factors require an advanced understanding about the interactive forces.

An alternate view to this problem would be to utilize the hubs of the network. The hubs play a critical role in linking several other nodes, thereby enabling them to have shorter path lengths. The adverse consequence observed due to the exclusion of well-linked nodes underscores their importance to the shorter path lengths. A higher weight could be provided to the edges of nodes not interacting with a well-connected node. For such nodes, the absence of any links to a hub produces a longer path length to a target node. On the other hand, the edges of nodes that link to a highly connected node can be assigned a lower weight. This approach will provide path lengths based on weights to a target node. Nodes that have a link to a hub will produce shorter weight-based path lengths to its target than when computed in their absence.

The datasets for the two interaction networks that were investigated are incomplete and prone to update sooner than later, when experimental results become available. There has been no revision at the time of drafting this paper, but it would be interesting to analyze the topology including the new data after it becomes accessible. Presently, we are devoted to remodeling the networks using a weight-based approach. An enhanced review should be available in the not so distant future.

## Materials and Methods

### Protein-protein interaction network dataset

The protein-protein interaction dataset was retrieved from Curagen Corporation (www.curagen.com). This dataset was experimentally obtained by Giot et al (24) using the yeast two-hybrid screening technique, resulting in 41,068 protein-protein interaction pairs. The obtained dataset was subjected to further classification into high confidence and low confidence subsets by studying the Gene Ontology (29) and using R statistical package. For this investigation, the high confidence (confidence scores higher than 0.5) protein-protein interaction pairs were selected, since they have a greater support for occurrence. A set of 4,591 proteins involving 9,334 high confidence interactions, exclusive of self-interactions, was used for the protein-protein interaction network. This dataset contained both unidirectional and bi-directional links. A careful study of this dataset revealed that two specific proteins stand apart. Among the proteins that had a one-directional interaction, two proteins, namely CG4039 and CG12918, only had incoming edges. There were several other proteins with unidirectional links, but all those proteins had outward-bound edges. The strength of each protein-protein interaction was unknown, and the usage of confidence scores for the weight of the interaction links did not supply any material denotation to them. Therefore, it was decided that the weight of the edges would be assumed as one when an interaction existed, and in the absence of an interaction, the influence was understood to be zero to produce a binary network.

To probe the network architecture and characteristics, proteins in the fruit fly protein interaction network were ranked based on their degrees in descending order. Top 3% of the highly connected proteins were removed sequentially to feign a targeted attack, and the effect on the average path length of the interaction network was studied. On the other hand, 3% of the proteins were removed at random, creating an unintended failure of nodes, and the effect on the topological parameter was calculated again. A graphical plot was used to demonstrate the effect of removal of proteins on the network topology. The molecular function of the hub proteins, retrieved from Flybase (26), was used to identify any possible association to the distinguishing behavior exhibited due to the targeted elimination of those proteins.

The Boost Graph Library, which contains the breadth first search algorithm, was used to calculate the shortest path length between every pair of accessible proteins (25). The shortest path length of the proteins was then employed to compute the average path length of the protein-protein interaction network.

## Construction of the metabolite-metabolite interaction network

The metabolic pathways and the associated metabolites for *D. melanogaster* were retrieved from the KEGG database (*28*). The substrates and the products involved in a reaction were then isolated. The outgoing links, represented by $C_{out}$, indicated the substrate metabolites. In contrast, the incoming links signifying the metabolites that occur as products in the reaction were designated by $C_{in}$. A frequency distribution histogram was used to verify the nature of the network architecture and to determine the impact of the compounds that play a part in the metabolism of the organism.

To explore the metabolite-metabolite interaction that affects the topological properties of the metabolic network, metabolites involved in the same reaction were treated as follows. For example, in a reaction involving four metabolites, C05125+C00011<=>C00068+C00022, the connections (interactions) were built as C05125–C00068, C05125–C00022, C00011–C00068, C00011–C00022, C00068–C05125, C00068–C00011, C00022–C05125, and C00022–C00011. On a similar basis, a list of connections was organized for all the reactions. If any reaction contained more than a single mole of compound, for example, 2C00001, the reaction was then modified by a substitution of the compound index as C00001+C00001. The strength of each association was homogeneously maintained as one to generate a binary network.

The dataset generated above could not be used *per se*, as the list of prepared connections contained a large number of current metabolites (*13*). Current metabolites are normally used as carriers for transport of electrons and other functional groups to facilitate the catalysis of a reaction (*30*). The current metabolites may be explained as being analogous to an external metabolite that takes part in more than a few reactions but does not occur in the pseudo steady state in a sub-network (*31*). The fast-paced nature and high yield of metabolites during reactions has resulted in a pseudo steady state assumption that, on longer time scales, the concentration of metabolites and the rate of reactions are stable. This condition guarantees that none of the metabolites are produced or consumed in the overall stoichiometry (*32, 33*). The calculation of the shortest path length, or in other words, the least number of reactions required to get from one compound to another using this set of data, would provide an inaccurate interpretation of the topology of the network (*13*).

The current metabolites have to be removed before the calculation of the topological parameter, since their inclusion would generate fallacious parameters. Nevertheless, the deletion of the current metabolites and their possible connections could not be done *per se*. Some of the current metabolites may be primary metabolites acting as either substrates or products. A primary metabolite is essential for regular growth and reproduction. Accordingly, those reactions in which the current metabolites acted as primary substrates or products were permitted to be part of the network. The remaining connections that did not involve the current metabolites as primary metabolites (either as substrates or products) were deleted and the links between substrates and products were reconstructed. This strategy and the exclusion of redundant connections caused the number of links to be reduced to 3,326. This restructured dataset was used for the calculation of the shortest path lengths of the network.

The metabolite interaction network was examined for its tolerance against random failures and weakness to sequential errors. The metabolites were sorted based on their degrees (inclusive of incoming and outgoing edges) in decreasing order. The well-connected metabolites were removed successively to simulate a premeditated attack, and metabolites were removed in an unsystematic fashion imitating an accidental failure of nodes. The outcome on the average path length was studied for both cases. A graphical plot was utilized to highlight the effects of exclusion of metabolites on the overall topology of the network.

## Authors' contributions

TH carried out the study, and YHL supervised the work. Both authors read and approved the final manuscript.

## Competing interests

The authors have declared that no competing interests exist.

## References

1. Albert, R. and Barabási, A-L. 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74: 47-97.

2. Erdős, P. and Rényi, A. 1959. On random graphs. *Publ. Math.* 6: 290-297.

3. Erdős, P. and Rényi, A. 1960. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5: 17-61.

4. Erdős, P. and Rényi, A. 1961. On the strength of connectedness of a random graph. *Acta Math. Acad. Sci. Hung.* 12: 261-267.

5. Watts, D.J. and Strogatz, S.H. 1998. Collective dynamics of "small-world" networks. *Nature* 393: 440-442.

6. Watts, D.J. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness.* Princeton University Press, Princeton, USA.

7. Strogatz, S.H. 2001. Exploring complex networks. *Nature* 410: 268-276.

8. Amaral, L.A., *et al.* 2000. Classes of small-world networks. *Proc. Natl. Acad. Sci. USA* 97: 11149-11152.

9. Newman, M.E. 2001. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* 98: 404-409.

10. Newman, M.E. 2001. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E* 64: 016131.

11. Newman, M.E., *et al.* 2001. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64: 026118.

12. Jeong, H., *et al.* 2000. The large-scale organization of metabolic networks. *Nature* 407: 651-654.

13. Ma, H.W. and Zeng, A.P. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19: 270-277.

14. Ma, H.W. and Zeng, A.P. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19: 1423-1430.

15. Hasty, J. and Collins, J.J. 2001. Protein interactions. Unspinning the web. *Nature* 411: 30-31.

16. Ito, T., *et al.* 2000. Toward a protein-protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci. USA* 97: 1143-1147.

17. Flores, A., *et al.* 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proc. Natl. Acad.*

*Sci. USA* 96: 7815-7820.

18. Schwikowski, B., *et al.* 2000. A network of protein-protein interactions in yeast. *Nat. Biotechnol.* 18: 1257-1261.

19. Jeong, H., *et al.* 2001. Lethality and centrality in protein networks. *Nature* 411: 41-42.

20. von Mering, C., *et al.* 2002. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417: 399-403.

21. Bu, D., *et al.* 2003. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Res.* 31: 2443-2450.

22. Rain, J.C., *et al.* 2001. The protein-protein interaction map of *Helicobacter pylori. Nature* 409: 211-215.

23. Patterson, J. and Stone, W. 1952. *Evolution in the Genus Drosophila.* Macmillan Publishers, New York, USA.

24. Giot, L., *et al.* 2003. A protein interaction map of *Drosophila melanogaster. Science* 302: 1727-1736.

25. Siek, J.G., *et al.* 2001. *The Boost Graph Library: User Guide and Reference Manual.* Addison-Wesley Professional, Boston, USA.

26. Drysdale, R.A., *et al.* 2005. FlyBase: genes and gene models. *Nucleic Acids Res.* 33: D390-395.

27. Ogata, H., *et al.* 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27: 29-34.

28. Kanehisa, M., *et al.* 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30: 42-46.

29. Ashburner, M., *et al.* 2000. Gene ontology: tool for the unification of biology. *Nat. Genet.* 25: 25-29.

30. Neidhardt, F.C., *et al.* 1990. *Physiology of the Bacterial Cell: A Molecular Approach.* Sinauer Associates, Sunderland, USA.

31. Schuster, S., *et al.* 2002. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae. Bioinformatics* 18: 351-361.

32. Schilling, C.H., *et al.* 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203: 229-248.

33. Schuster, S., *et al.* 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18: 326-332.