Klimm, Florian; Deane, Charlotte M.; Reinert, Gesine
Hypergraphs for predicting essential genes using multiprotein complex data. (English)
J. Complex Netw. 9, No. 2, Article ID cnaa028, 16 p. (2021)

Summary: Protein-protein interactions are crucial in many biological pathways and facilitate cellular function. Investigating these interactions as a graph of pairwise interactions can help to gain a systemic understanding of cellular processes. It is known, however, that proteins interact with each other not exclusively in pairs but also in polyadic interactions and that they can form multiprotein complexes, which are stable interactions between multiple proteins. In this manuscript, we use hypergraphs to investigate multiprotein complex data. We investigate two random null models to test which hypergraph properties occur as a consequence of constraints, such as the size and the number of multiprotein complexes. We find that assortativity, the number of connected components, and clustering differ from the data to these null models. Our main finding is that projecting a hypergraph of polyadic interactions onto a graph of pairwise interactions leads to the identification of different proteins as hubs than the hypergraph. We find in our data set that the hypergraph degree is a more accurate predictor for gene essentiality than the degree in the pairwise graph. In our data set analysing a hypergraph as pairwise graph drastically changes the distribution of the local clustering coefficient. Furthermore, using a pairwise interaction representing multiprotein complex data may lead to a spurious hierarchical structure, which is not observed in the hypergraph. Hence, we illustrate that hypergraphs can be more suitable than pairwise graphs for the analysis of multiprotein complex data.

MSC:
05-XX Combinatorics

Keywords:
gene essentiality; protein interaction networks; hypergraphs; null models; hierarchical exponent; centrality; clustering coefficient

Full Text: DOI