Journal of Research and Applied Biosciences

Journal home page: www.biolresearch.com

Research Article

Graph theory on protein-protein interaction analysis and genomic DNA analysis

R. Ajitha

Department of Mathematics, Udaya College of Arts and Science, Vellamodi, Ammandivilai 629 104, Kanyakumari District, Tamilnadu, India
E.mail: drajitha1990@gmail.com

Article info
Received: Nov 07, 2017, Received in revised form: Feb 21, 2018, Accepted: March 12, 2018

Abstract

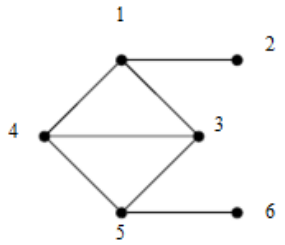
Graph theory is widely used to analyze the functional properties and interaction of proteins and DNA. This theory of complex networks plays critical role in molecular biology and population biology. Protein-protein interaction networks carry vital information on the organization of molecular interactions in cellular systems. The identification of protein-protein interactions is an important application in biological network analysis Graph Theory is used to apply genomic analysis and in genetic engineering technique. In this paper application of graph theory in protein –protein interaction and genomics were analyzed.

Key words: Graph Theory, Genomics, Protein-protein interaction, biological network

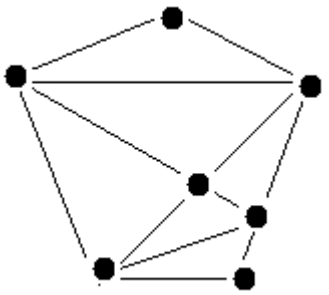
Intoduction

The *line graph L(G)* of a graph *G* is defined to have as its vertices the edges of *G*, with two being adjacent if the corresponding edges share a vertex in *G*. The term line graph was first coined by Harary and Norman (1960). However, line graphs were the subject of investigations in 1930s (Whitney, 1932). Whitney (1932) investigated edge isomorphism and reported that for connected graphs, edge-isomorphism implies isomorphism except for $K3$ and $K1,3$. Krausz (1943) first characterized partition into complete sub graphs. Line graph can view as a transformation $G \rightarrow L(G)$. Repeated uses of this transformation yield *iterated line graphs*, which have been studied by various researches.

Thus iterated line graphs are defined by $L1(G) = L(G)$ and $Ln+1(G) = L(Ln(G))$



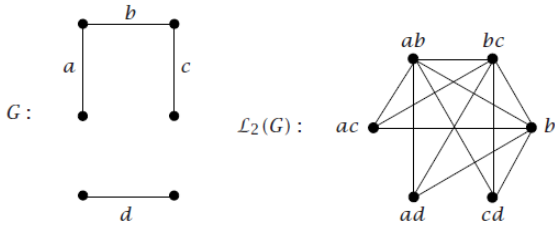
G



If *G* is the path *Pn* on *n* vertices, then *L(G) = Pn*−1. Hence *{Ln}* terminates for paths. If *G* is a cycle, then *L(G) = G*. Also *L(K*1,3*) = K*3. Hence *{Ln}* becomes a constant in these graphs. For all other connected graphs, *{Ln}* contains arbitrarily large graphs (van Rooij and Wilf, 1965).

The old generations of line-graphs are total graph *T(G)*, *middle graph* mid*(G)*, entire graph *e(G)*, *path graph* _*r(G)* and sub graphs.

Old and new generalizations of line graphs

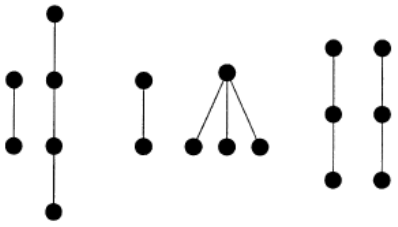


Super line graphs

Definition and properties of super line graph

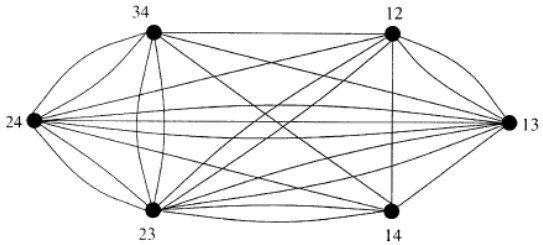
For a fixed integer *r* (with 1 ≤ *r* ≤ *q* = |*E(G)*|),

the *super line graph* _*r(G) of index r* has the sets of *r* edges in *G* as its vertices, and two vertices are adjacent if an edge in one set is adjacent (in *G*) to an edge in the other.



It follows that _1*(G)* is the usual line graph.

Many variations of the definition of a super line graph can be considered. One could form a multigraph by joining two vertices with as many edges as there
are adjacencies between the two sets of edges. This is called as super line multigraph.



Graph theory in biological experiments

Graph theory is used to analyze biological networks. This theory of complex networks plays critical role in molecular biology and population biology. Within the fields of medicine and biology, network analysis is used to identify drug targets, determination of protein gene and protein function, designing strategies for treating various diseases or giving early diagnosis of various disorders. Protein-protein interaction networks, transcriptional regulation networks, biochemical networks, metabolic networks or signal transduction are the highlighted network categories in systems biology often sharing characteristics and properties.

Protein-protein interaction (PPI) networks

Protein-protein interaction (PPI) networks mainly show how different proteins interact in coordination with others to enable the biological processes within the cell. The majority of proteins and their complete sequences are known, however, their molecular function is not yet fully elucidated. Predicting the functions of protein is still a bottleneck in computational biology and many computational techniques and biological assays developed in order to conclude protein function from interactions with other biomolecules. High-throughput and large-scale techniques can significantly detect proteins that interact within an organism. The important biological assays based on computational biology are pull down assays (Vikis and Guan, 2004), mass spectrometry Gavin et al., 2002), microarrays (Stoll et al., 2005), tandem affinity purification (TAP) (Puig et al., 2001), yeast two-hybrid (Y2H) (Ito et al., 2001), phage display (Willats, 2002). Some of the data sets such as Tong, DIP, Gavin 2002 and Gavin 2006 were used in biological experiments.

Graph theory in genomics

Graph theory is used analyze the genomics problem. It is used to analyze the pseudogenes alignments (Re et al., 2008). The first application is to characterize the "expressed" genes on chromosomes. The second application is to study the regulatory mechanism at the transcription level. The second application studies a potential new class of regulatory mechanisms at the level of the transcription process in the budding yeast *S. cerevisiae*. According to this hypothesis, pseudogenes would mainly act as regulators of their corresponding mRNAs coding genes (Corbin et al., 2002). Common fragile sites (CFS) are a fragment of DNA showing a high rate of recombination events. Such events imply both external DNA viral integration or intracellular DNA exchange. CFS are said to be "expressed" when they show one of the above mentioned events. These regions are mainly termed "common" since they exist in almost all living organisms, hence they do not denote by themselves a pathological status of the cell. They have been widely studied in mice and humans (Matsuyama et al., 2003) and all eukaryotes. These CFS regions are conserved by evolution however the functional role of these regions is not completely elucidated. CFS region has negative or positive role in tumour development. Functional analysis was carried out using the Gene Ontology database. Gene Ontology provides a controlled and dynamic annotation framework for describing gene products (Camon et al., 2004).

References

van Rooij ACM, Wilf HS. The interchange graph of a finite graph. Acta Math Acad Sc. Hungar 1965; 16: 263–269.

Corbin S, Neilly ME, Espinosa R, Davis EM, Mckeithan TW, Le Beau MM. Identification of unstable sequences within the common fragile site at 3p14.2: Implications for the mechanism of deletions within fragile histidine triad gene/common fragile site at 3p14.2 in tumors, Cancer Res 2002; 62(12): 3477–3484.

Harary F, Norman RZ. Some properties of line digraphs. Rend Circ Mat Palermo 1960; 2 (9): 161–168.

Whitney H. Congruent graphs and the connectivity of graphs. Amer J Math 1932; 54: 150–168.

Krausz J. Démonstration nouvelle d'une théorème de Whitney sur les réseaux. Mat Fiz Lapok 1943; 50: 75–85 (Hungarian).

Bagga KS, Beineke LW, Varma BN. *Super line graphs*, Graph Theory, Combinatorics, and Algorithms, Vol. 1, 2 (Kalamazoo, Mich, 1992), Wiley-Interscience, New York, 1995, pp. 35–46.

Vikis HG, Guan KL. Glutathione-S-transferase-fusion based assays for studying protein-protein interactions. Methods Mol Biol 2004; 261:175-186.

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Seraphin B: The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 2001; 24(3): 218-229.

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci USA 2001; 98(8): 4569-4574.

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, et al: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002, 415(6868):141-147.

Stoll D, Templin MF, Bachmann J, Joos TO: Protein microarrays: applications and future challenges. Curr Opin Drug Discov Devel 2005; 8(2):239-252.

Willats WG: Phage display: practicalities and prospects. Plant Mol Biol 2002; 50(6): 837-854.

Rea A, Molinerisa I, Casellea M. Graph theory analysis of genomics

problems: Community analysis of fragile sites correlations and of pseudogenes alignments. Comput Mathemat Appl 2008; 55: 1034–1043

Matsuyama A, Shiraishi T, Trapasso F, Kuroki T, Alder H, Mori M, Huebner K, Croce CM. Fragile site orthologs fhit/fra3b and fhit/fra14a2: Evolutionarily conserved but highly recombinogenic. Proc Natl Acad Sci USA 2003; 100(25): 14988–14993.

Camon E, Magrane M, Barrell D, Lee V, Dimmer E, Maslen J, Binns D, Harte N, Lopez R, Apweiler R. The gene ontology annotation (goa) database: Sharing knowledge in uniprot with gene ontology. Nucl Acids Res 2004; 32: D262–D266.