IMA Annual Program Year Workshop:
BIOLOGICAL SYSTEMS AND NETWORKS

Graph Theory for Systems Biology:
Interval Graphs, Motifs, and Pattern Recognition

John R. Jungck
Professor of Biological Sciences and Mathematical Sciences
Computational Biology and Bioinformatics
Interdisciplinary Science and Engineering Lab
University of Delaware
‘Two-node feedback’ is the most significant motif in all three species. By considering the sign of each two-node feedback interaction, we examined the enrichment of the three types of two-node feedbacks:

1. positive–positive (PP),
2. negative–negative (NN), and
3. positive–negative (PN).

We found that PN is enriched in the network of *A.thaliana*, NN in the network of *S.cerevisiae* and PP and NN in the network of *H.sapiens*. Each feedback type has characteristic features of robustness, multistability and homeostasis.
Modular organization of cellular networks

Alexander W. Rives and Timothy Galitski*
Institute for Systems Biology, 1411 North 34th Street, Seattle, WA 98103

Communicated by Larry Hood, Institute for Systems Biology, Seattle, WA, December 3, 2002 (received for review August 24, 2002)

We investigated the organization of interacting proteins and protein complexes into networks of modules. A network clustering method was developed to identify modules. This method of network structure determination was validated by clustering known signaling module proteins and by identifying modules in an unbiased high-throughput protein-protein interaction data set with high accuracy and low significance. The signaling network controlling the yeast developmental transition to a filamentous form was clustered. Abstraction of a modular network structure model identified transduction modules for several signaling proteins. The functions of these proteins suggest that they are important for modulating transcription and intracellular communication.

Protein molecules bind to each other to form stable complexes that often can be purified. At a higher level of structure, proteins and protein complexes interact with preferred partners reversibly, transiently, or conditionally to form a biological module or a specific collective function (1). For example, MAPKK (mitogen-activated protein kinase) cascade, together with its scaffold proteins and various regulators and effectors, forms a signal amplification module. As another example, the spindle-pole body is a complex of proteins that forms a hub for the attachment and organization of microtubules. Driven by the acquisition of whole-genome, single-cell data sets from complex biological systems, our conception of biomolecular organization is evolving from metabolic and signaling pathways to networks of evolutionarily conserved modules (2-4).

With the abundance of protein-protein interaction data produced by genomewide screening (5-9), it is possible to create a global representation of the protein-interaction network of the yeast cell (4-6). This network has been shown to have a nonrandom power-law distribution of node connectivity (number of interactions of each protein) and low frequency of direct connections between high-connectivity nodes (10). These observations suggest modular organization consistent with the insights of Kogut and (1). Various heuristics for network clustering have been developed and applied to identify modules in various biological systems, including genome coexpression network (11), a food-web (12), and the Escherichia coli metabolic network (13). We developed and evaluated a network-clustering method using the modular network of yeast signaling proteins. In addition, we identified modules in an interaction network consisting of electron transfer and data with very high accuracy and very high fidelity in functional identification. Moreover, we integrated functional gene-identification data with protein-protein interaction data to provide a modular abstraction of the organization of a complex network controlling a biological response.

Materials and Methods

Network Clustering. For each biological network investigated, vertices were assigned to clusters (each interaction among them) were assembled as described in the text. Each edge in the network was assigned a length of 1. An all-pairs shortest-path distance matrix was calculated using standard algorithms. The all-pairs shortest-path distances were defined as the length of the shortest path (distance) between every pair of vertices in the network. Each distance in the all-pairs shortest-path matrix was transformed into an "association," defined as 1/2, where d is the shortest-path distance. This transformation emphasizes local associations (short paths) in the subsequent clustering. The resulting associations range from 0 to 1. The association of a vertex with itself was defined as 1. The association of vertices that have no connecting path was defined as 0. Hierarchical agglomerative average-linkage clustering was employed using the uncentered correlation distance as the metric. The (14) was applied to the association matrix. The t test procedure (15) was used to view the results. For the clustering of the signaling-protein network, the Munich Information Center for Protein Sequences (MIPS) (15) database list of proteins of the signaling category and its pathway subcategories was obtained in August 2001.

Identification of Filamentation-Netwok Proteins. We searched the Yeast Protein Database (16) for proteins with annotations matching the search query "involved OR filamentous OR pseudohyphae." The list was screened manually for relevance to filamentation, invasive growth, or pseudohyphal development. The list was supplemented with proteins implicated in reviews (17, 18).

Results

Network Clustering. We sought to compute the modular organization of cellular networks controlling specific biological responses. We represented yeast protein-interaction networks as graphs of vertices and edges (nodes and links corresponding to proteins and interactions), and developed a network-clustering method based on the following ideas: (i) the shortest path between any two vertices is likely to be the most relevant one for functional associations and information transmission; (ii) each vertex in a network has a unique profile of shortest-path distances through the network to every other vertex; and (iii) module boundaries are likely to have similar (clustered) shortest-path-distance profiles. The method is described in Materials and Methods.

Molecular Structure of the Yeast Signaling Network. The conception of the structure of cellular systems as a network of modules (1) is well supported by the results of the yeast signaling network. Accordingly, we used this system, for which there are many high-confidence individually validated interaction data to test our approach.

We asked a set of questions concerning 4,079 proteins and 6,764 protein-protein interactions from a global two-hybrid screen (7) and a composite data set (4) that includes global two-hybrid data and individually validated interaction data. The MIPS-database signal-protein category (15) includes 133 proteins. Of these, 64 had at least one interaction with another signaling protein. A network consisting of the 64 proteins and the interactions among them was constructed from the global set of interactions. Network clustering was applied to this signaling network.

The results are displayed as a grayscale representation of the values in the clustered protein-protein pairwise association.

PNAS (2003)
Could a little graph theory inform Rives & Galitski’s work?
Interval graphs:

An interval graph $G = (V,E)$ is an undirected graph obtained from a collection $C$ of intervals on the real line. To each interval in $C$ there corresponds a vertex in $G$. The edge $(u,v)$ is in $E$ if and only if their corresponding intervals intersect.
Benzer divided the region of the chromosome into subsections with Ordered overlapping deletions and then mapped point mutations in Standard three-point crosses to determine order within subsections.
Benzer’s map was entitled the first “Fine Structure of the Gene”

It clearly demonstrated that the gene was not the unit of function, recombination, and function.

Mutation was not uniform; there were hot spots.

Recombination could occur between adjacent nucleotides in DNA.
Graph Theory approach

Fulkerson & Gross 1964
Intersection Graph & Complementary Graph
No “$Z_4$" s
Transitively Oriented Complementary Graph
Maximal Cliques

A

B

C

D
Hamiltonian Path of Maximal Cliques

outDegree b(3) -> a(2) -> c(1) -> d(0)
Interval Graph Solution
Steps in the graph theory solution of "Benzer's problem"

1. Convert each deletion mutant (restriction fragment) into a vertex.
2. Construct the intersection graph by placing an edge between each pair of vertices which represent overlapping deletions.
3. Construct the complement of the intersection graph.
4. Check for absence of Z4's in the intersection graph.
5. Determine whether the complement of the intersection graph can be made transitive.
6. Find all the maximal cliques in the intersection graph.
7. Order these maximal cliques in the same way as in the transitive complementary graph.
8. Find the Hamiltonian path of all the ordered maximal cliques.
9. Construct the interval graph by assigning deletions to each interval of the line, which sequentially orders the maximal cliques, for all the cliques to which the deletion vertex belongs.

Thus, the algorithm is capable of processing the original recombination matrix data through each of these nine steps.
Set Theoretic approach

Shkurba 1965
original form
Move 5 one to the right
Move 1 three to the right
“Shkurba form”
<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

$2,3,4,7$  $1,3,4,7$  $1,4,6,7$  $5,6,7$
Two Alternative Approaches

BioGrapher

Visualization matters!

javaBenzer
The field of Network Design comprises a large class of problems:

- Low cost
- High Capacity
- Fault Tolerant
- Highly Connected (the remaining nodes will still be able to communicate)
- Low congestion
One-dimensional logic gate assignment and interval graphs


- This paper gives a graph-theoretic approach to the design of one-dimensional logic gate arrays using MOS or units. The incidence relation between gates and nets is represented by a graph, and a possible layout of gates and nets is characterized by an interval graph, where is called an augmentation. It is shown that the number of tracks required for between-gate wiring is equal to the clique number (chromatic number) of , and hence the optimum placement problem is converted to that of minimum clique number augmentation. This turns out to be an NP-complete problem. Instead a polynomial-time algorithm for finding a minimal augmentation is presented, where an augmentation is minimal if no proper subset of it is an augmentation. An algorithm for gate sequencing with respect to a given augmentation is also presented.
“This follows from a classic interval-graph coloring result:

if at most L subintervals of a line segment contain any point of the segment, then the subintervals can be colored with at most L colors so that overlapping subintervals have distinct colors. ...”
Decentralized Self-Adaptation Mechanism for Service Based Applications in Cloud using Spectral Clustering

E. NISHA

SCAD Engineering College, Cheranmahadevi, India.

(Received: March 20, 2014; Accepted: March 30, 2014)

ABSTRACT

Cloud computing, with its promise of (almost) unlimited computation, storage and bandwidth, is increasingly becoming the infrastructure of choice for many organizations. As cloud offerings mature, service-based applications need to dynamically recompose themselves, to self-adapt to changing QoS requirements. In this paper, we present a decentralized mechanism for such self-adaptation, using natural language processing. We use a continuous double-auction to allow applications to decide which services to choose, amongst the many on offer. The proposed scheme exploits concepts derived from graph partitioning, and groups together tasks so as to 1) minimize the time overlapping of the tasks assigned to a given resource and 2) maximize the time overlapping among tasks assigned to different resources. The partitioning is performed using a spectral clustering methodology through normalized cuts. Experimental results show that the proposed algorithm outperforms other scheduling algorithms for different values of the granularity and the load of the task requests.