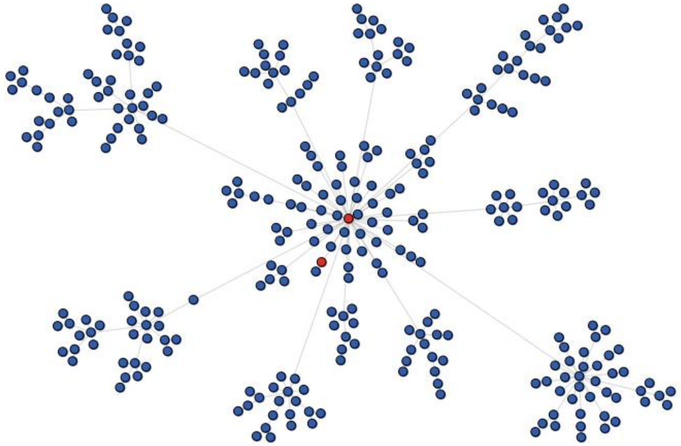
دانشگاه کردستان
University of Kurdistan
زانکۆی کوردستان

Popular Methods of biological networks analysis

Sadegh Sulaimany



Biological network analysis course

Agenda

- Neighborhood Statistics
- Paths and Traversals
- Connectivity
- Centrality and Ranking
- Subgraph Isomorphism and pattern
  - Motif finding
- Clustering

Introduction

- Graph analysis
  - to extract information from a graph that is not evident from individual entities or relationships
  - Example
    - Degree of a single node vs. degree of all nodes
      - We can determine how this node is popular and how important it is, what communities it apart, ...
  - We will enumerate the methods based on different applications

Neighborhood statistics

- These statistics are computed for
  - Each vertex in a graph
  - and depend only on the immediate neighborhood of a vertex

- Such as
  - *clustering coefficient, degree distribution, local efficiency, local subgraph,*

- To quantify a graph's structural properties

Neighborhood statistics

- Degree distribution & Clustering coefficient
  - to support the hypothesis that a network’s structure is different from a random graph
  - The performance of neighborhood analysis is often
    - sensitive to the structure of the input graph
  - If a vertex’s neighborhood is small
    - Analysis can benefit from exploiting data locality
  - the presence of several highly connected vertices
    - Such as power-law
    - Can lead to a significant imbalance in computation required per vertex

Paths and traversal

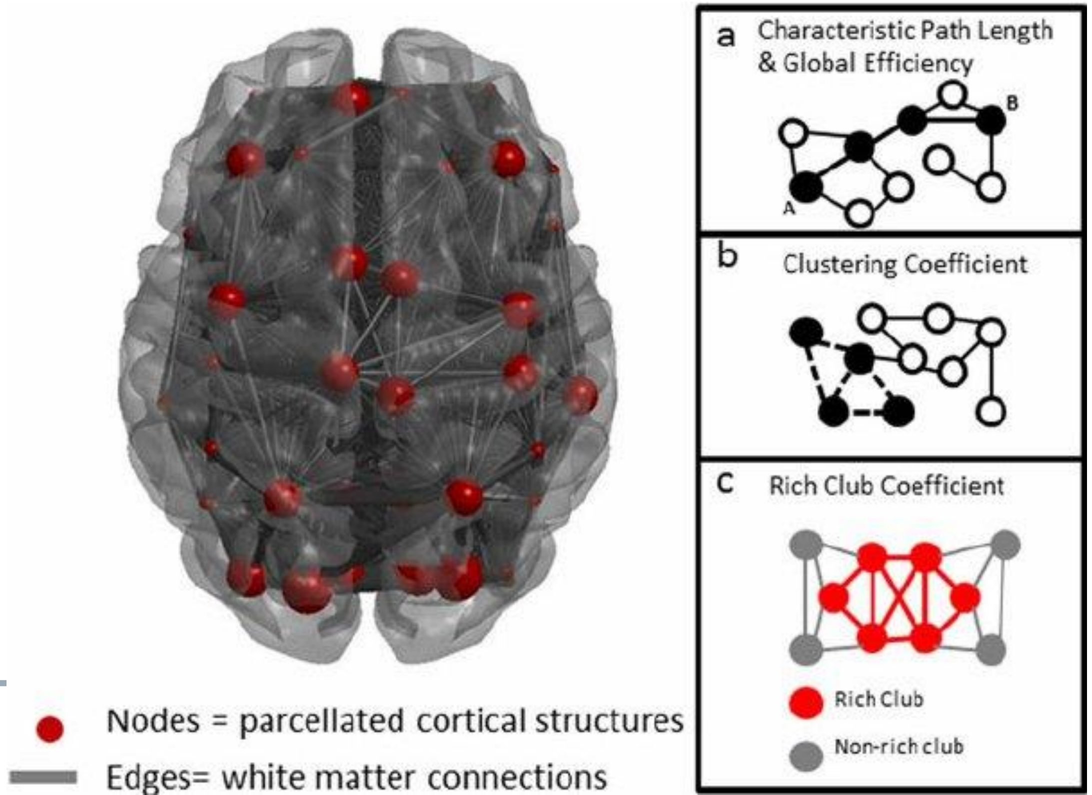
- Such as
  - *path length, shortest path, breadth-first, depth-first, traversal, global efficiency, random walk*

- Many applications are not restricted to the neighborhood of vertices
  - but instead involve the analysis of paths between vertices

- Path between two or more vertices
  - Dijkstra's algorithm vs. Floyd-Warshall algorithm

Paths and traversal

- Average path length
  - can be used to identify “small-world" properties
- Global efficiency
  - the inverse of the average path length
  - is used to determine the efficiency of communication between brain regions

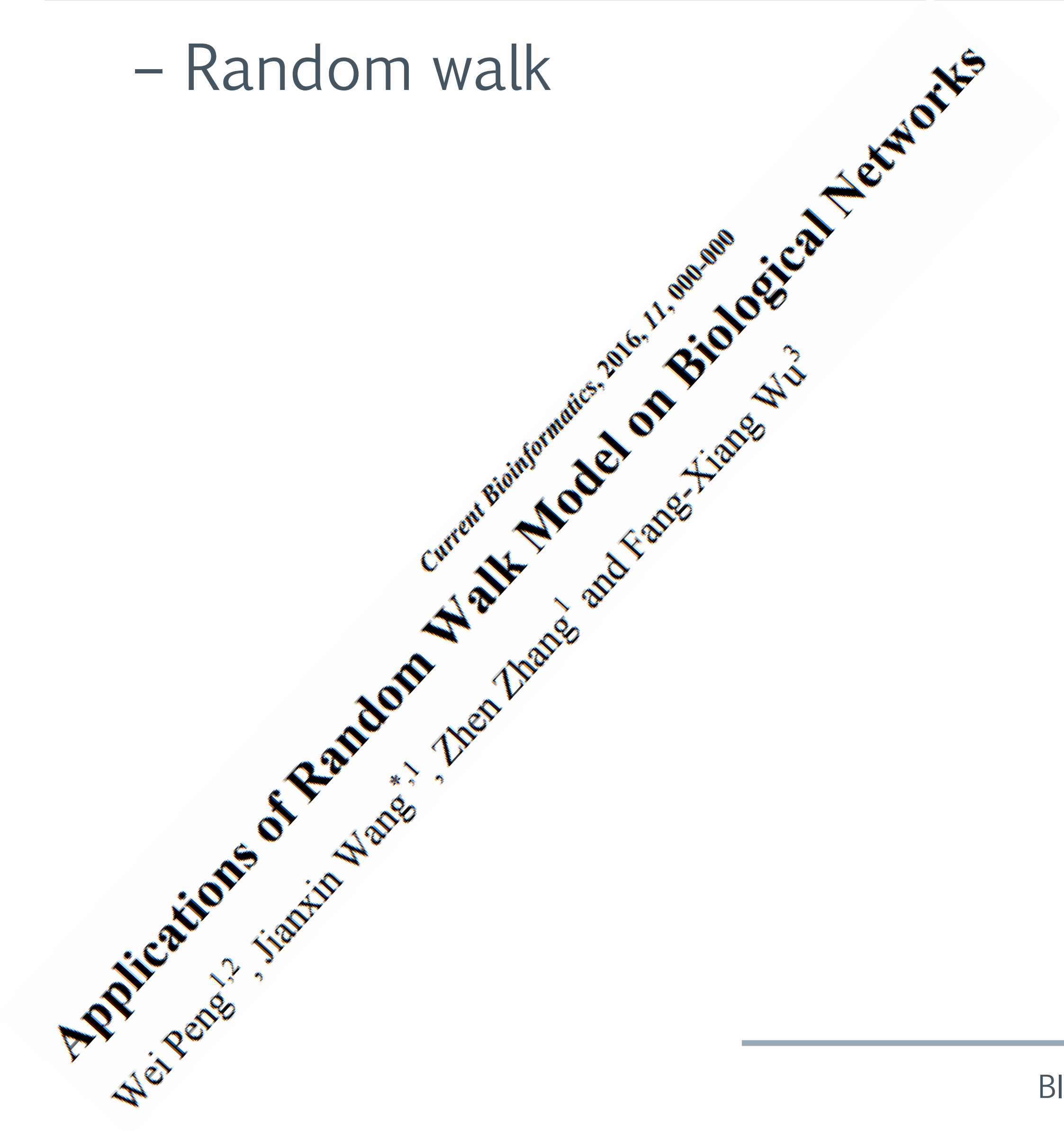


Paths and traversal

- computation of number of shortest paths between two vertices (i.e., the path multiplicity of a vertex pair)
  - is computed as a measure of flexibility in communication between brain regions
- Random walk
  - Various applications in biological networks

Paths and traversal

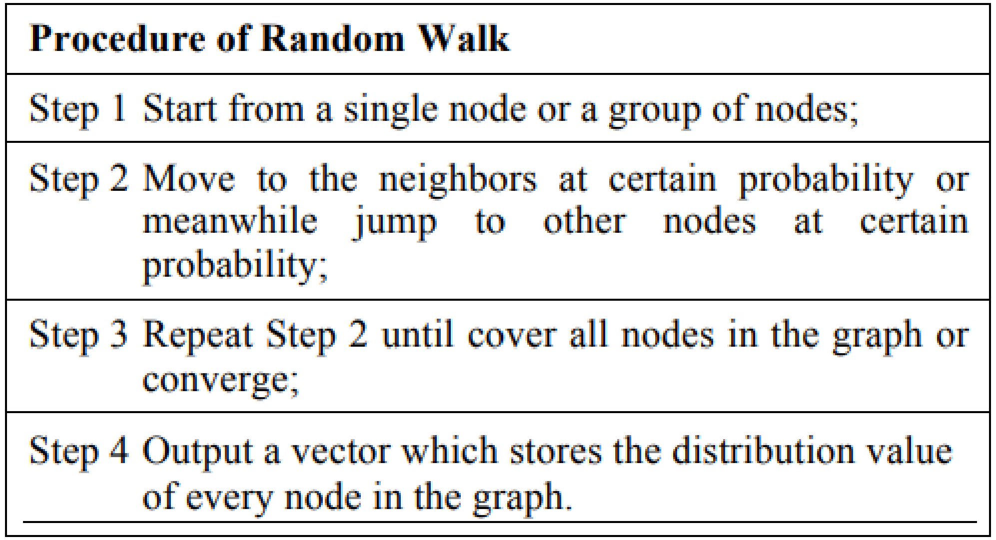
- Random walk



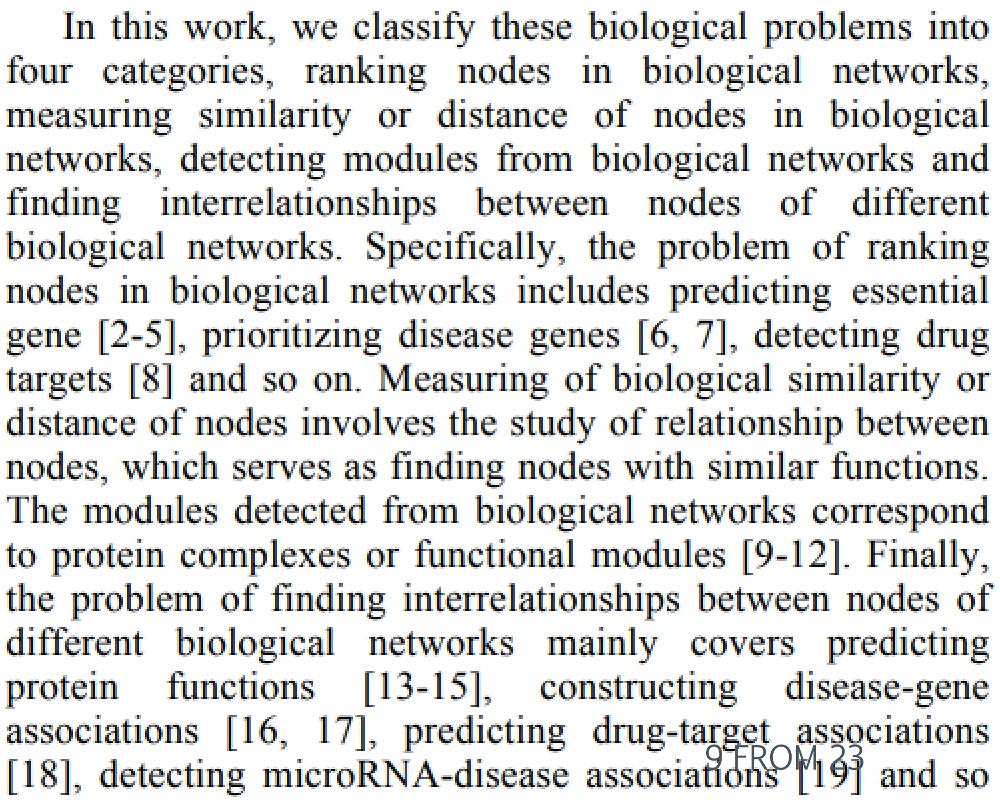
Current Bioinformatics, 2016, 11, 000-000

Applications of Random Walk Model on Biological Networks

Wei Peng[1,2], Jianxin Wang[*,1], Zhen Zhang[1] and Fang-Xiang Wu[3]



Procedure of Random Walk	
Step 1	Start from a single node or a group of nodes;
Step 2	Move to the neighbors at certain probability or meanwhile jump to other nodes at certain probability;
Step 3	Repeat Step 2 until cover all nodes in the graph or converge;
Step 4	Output a vector which stores the distribution value of every node in the graph.



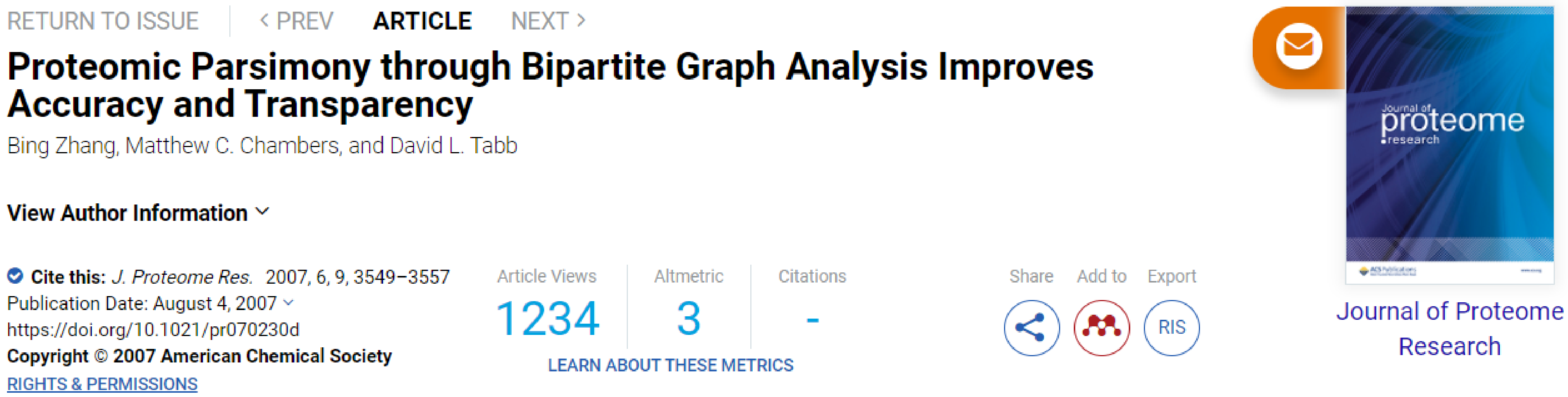
In this work, we classify these biological problems into four categories, ranking nodes in biological networks, measuring similarity or distance of nodes in biological networks, detecting modules from biological networks and finding interrelationships between nodes of different biological networks. Specifically, the problem of ranking nodes in biological networks includes predicting essential gene [2-5], prioritizing disease genes [6, 7], detecting drug targets [8] and so on. Measuring of biological similarity or distance of nodes involves the study of relationship between nodes, which serves as finding nodes with similar functions. The modules detected from biological networks correspond to protein complexes or functional modules [9-12]. Finally, the problem of finding interrelationships between nodes of different biological networks mainly covers predicting protein functions [13-15], constructing disease-gene associations [16, 17], predicting drug-target associations [18], detecting microRNA-disease associations [19] and so

Connectivity

- Instead of identifying *how* vertices in a graph are connected
  - i.e. discovering paths
  - some graph problems are concerned with determining
    - *If* vertices are connected
  - identifying connected components
  - For directed graphs
    - we distinguish two types of connected components:
      - weak and strong
  - weakly connected component
    - any pair of vertices is connected by an undirected path
  - strongly connected component
    - directed paths exist from each vertex to every other vertex

Connectivity

- Examples
  - Matching problem in a bipartite protein-peptide graph can be split into one smaller sub-problem per connected component



Proteomic Parsimony through Bipartite Graph Analysis Improves Accuracy and Transparency

Bing Zhang, Matthew C. Chambers, and David L. Tabb

Cite this: *J. Proteome Res.* 2007, 6, 9, 3549–3557
Publication Date: August 4, 2007
https://doi.org/10.1021/pr070230d
Copyright © 2007 American Chemical Society

Journal of Proteome Research

- Size of the largest component can be used as metrics to classify graphs
  - the fraction of vertices that belong to the largest connected component is a highly discriminative feature in classifying breast or bone tissue samples for cancer diagnosis

Centrality and Ranking

- Such as
  - Different Centrality measures
- Not all entities are of equal importance to the network
  - There are numerous applications for identifying the most important entities
    - Ranking entities
  - betweenness centrality, closeness centrality, and degree centrality

Centrality Measures in Biological Networks

Author(s): Mahdieh Ghasemi, Hossein Seidkhani, Faezeh Tamimi, Maseud Rahgozar, Ali Masoudi-Nejad

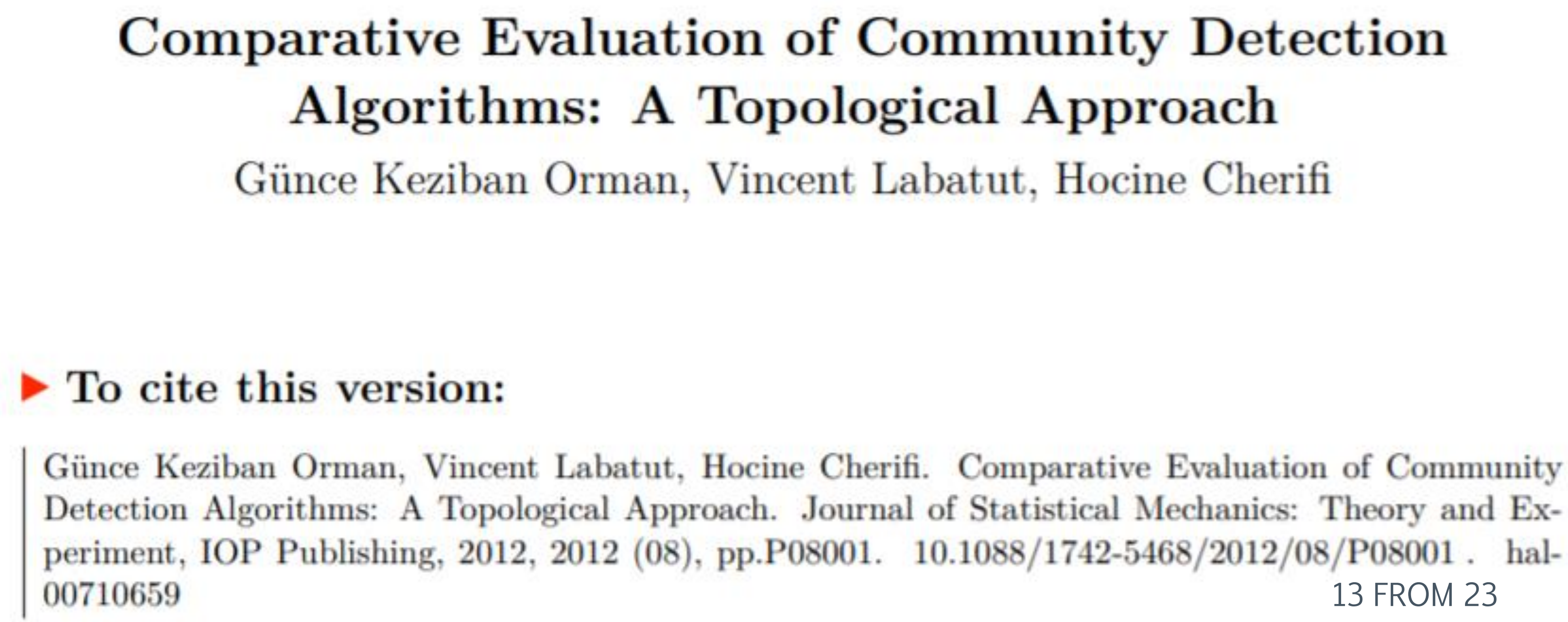
Journal Name: Current Bioinformatics

Volume 9 , Issue 4 , 2014

DOI : 10.2174/1574893611308660013

Clustering or community detection

- Used to identify groups of related entities from a graph based on the relationships
  - Relations is application-specific
    - So, there are many specialized clustering algorithms
  - Elbow method
    - Determining the number of clusters



Comparative Evaluation of Community Detection Algorithms: A Topological Approach

Günce Keziban Orman, Vincent Labatut, Hocine Cherifi

To cite this version:

Günce Keziban Orman, Vincent Labatut, Hocine Cherifi. Comparative Evaluation of Community Detection Algorithms: A Topological Approach. Journal of Statistical Mechanics: Theory and Experiment, IOP Publishing, 2012, 2012 (08), pp.P08001. 10.1088/1742-5468/2012/08/P08001 . hal-00710659

Subgraph Isomorphism and Patterns

- Task of identifying a subgraph of the input graph
  - that is isomorphic to a second graph (the *pattern*)
    - Examples
      - lookup chemical compounds by structure in a database

  - Popular terms
    - *subgraph, isomorphism, motif, pattern matching, cycle, clique, star, mesh*

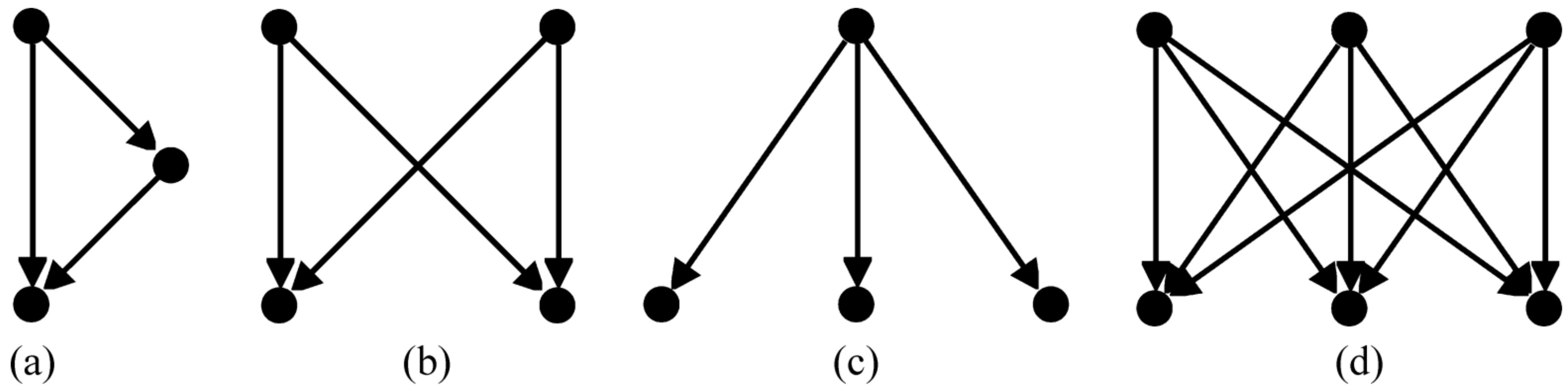
  - Pattern detection
    - subgraphs that match a more loosely defined pattern
      - e.g., a clique or a star
      - Example: identifying protein complexes as clique patterns

Motif finding

- *Network motif*
    - statistically significant overrepresented patterns of local interconnections in complex networks

– basic building blocks & design patterns of some bio. networks

- In biological networks: the structure has been shaped during evolution

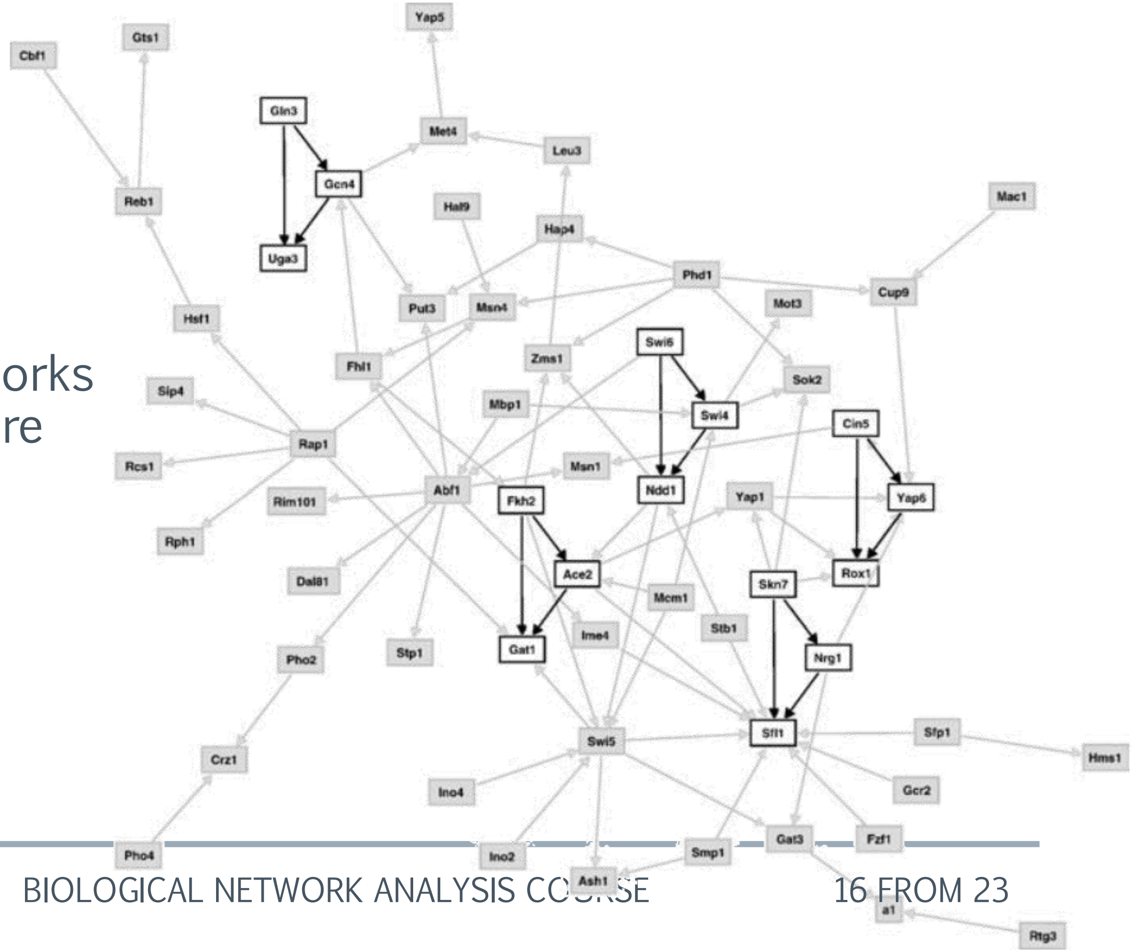


Network motifs that have been shown to be functionally relevant in biological networks: (*a*) feed-forward loop motif (*b*) bifan motif (*c*) single-input motif (*d*) multi-input motif

Motif finding

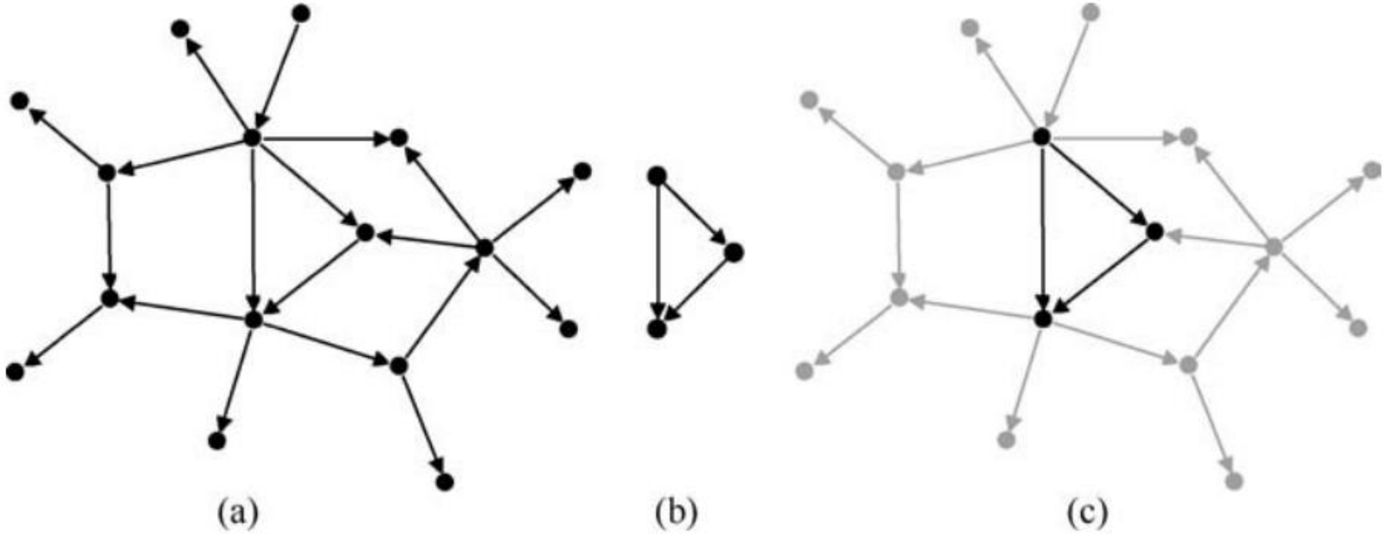
- Example
  - feed-forward loop motif
    - have particular information filtering capabilities within the process of gene regulation
  - *Saccharomyces cerevisiae*
    - *Gene-Regulatory network*

  - One non-bio example
    - example, the comparison of networks on the basis of their local structure



Motif finding

- Definition
  - *A motif is a small connected graph G ‘*
    - *the size of a motif is given in the number of vertices*
  - *A match of a motif within a target graph G is a graph G “*
    - *isomorphic to the motif G’ and a subgraph of G*
  - *Network motif analysis is generally for directed graphs but not limited to*
    - *GRN*



(a) a target graph G, (b) a motif G', and (c) a highlighted match G'' of the motif G' in the target graph G.

Motif finding

- Motif Frequency
    - number of different matches of this motif
  - Scenarios
    - *No restriction*
    - *Only vertices shared*
    - *Completely disjoint*

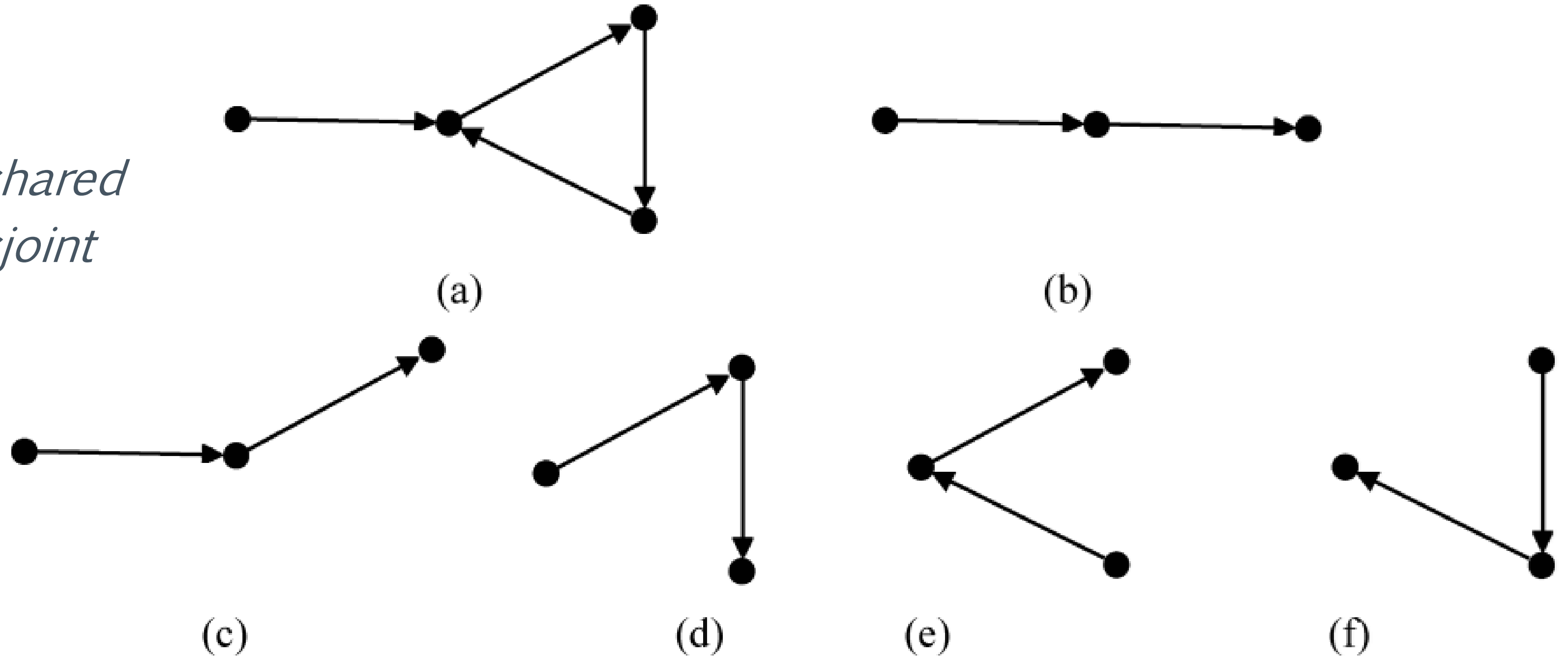
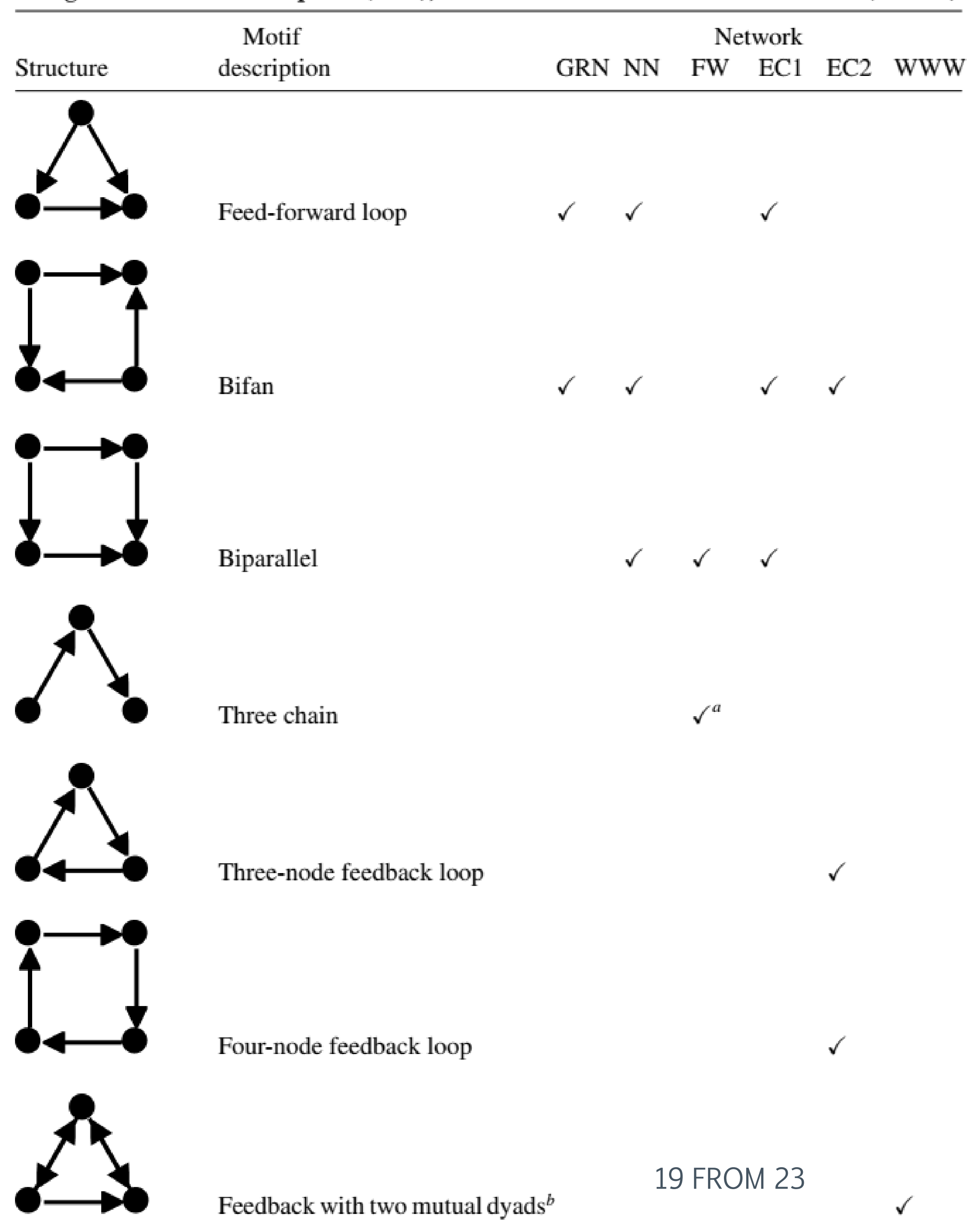


Illustration of the application of the different concepts of motif frequency on the basis of (*a*) a target graph G, (*b*) a motif G', and (*c*)–(*f*) all four matches G'' of the motif G' in the target graph G. The application of frequency concept $\mathcal{F}_1$ results in a frequency of four by counting all different matches (*c*)–(*f*), for $\mathcal{F}_2$ the frequency is two by counting the matches at (*c*) and (*f*), and for concept $\mathcal{F}_3$ the frequency is one as only one match out of the four matches (*c*)–(*f*) can be selected.

Motif finding

- Examples

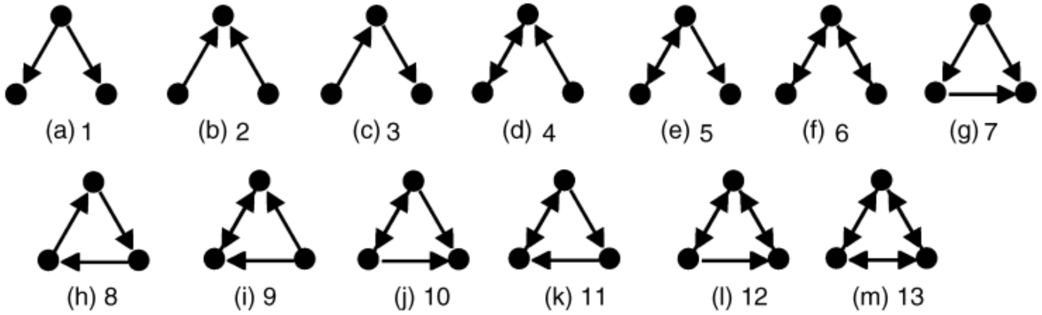
The Network Motifs Discovered in a Study on Biological and Technological Networks. The Analyzed Networks were Gene-Regulatory Networks (GRN) of *E. coli* and *S. cerevisiae*, a Neuronal Network (NN) of *Caenorhabditis Elegans*, Seven Food Webs (FW), Five Electronic Circuits of Forward Logic Chips (EC1), Three Electronic Circuits of Digital Fractional Multipliers (EC2), and a Network of the World Wide Web (WWW)



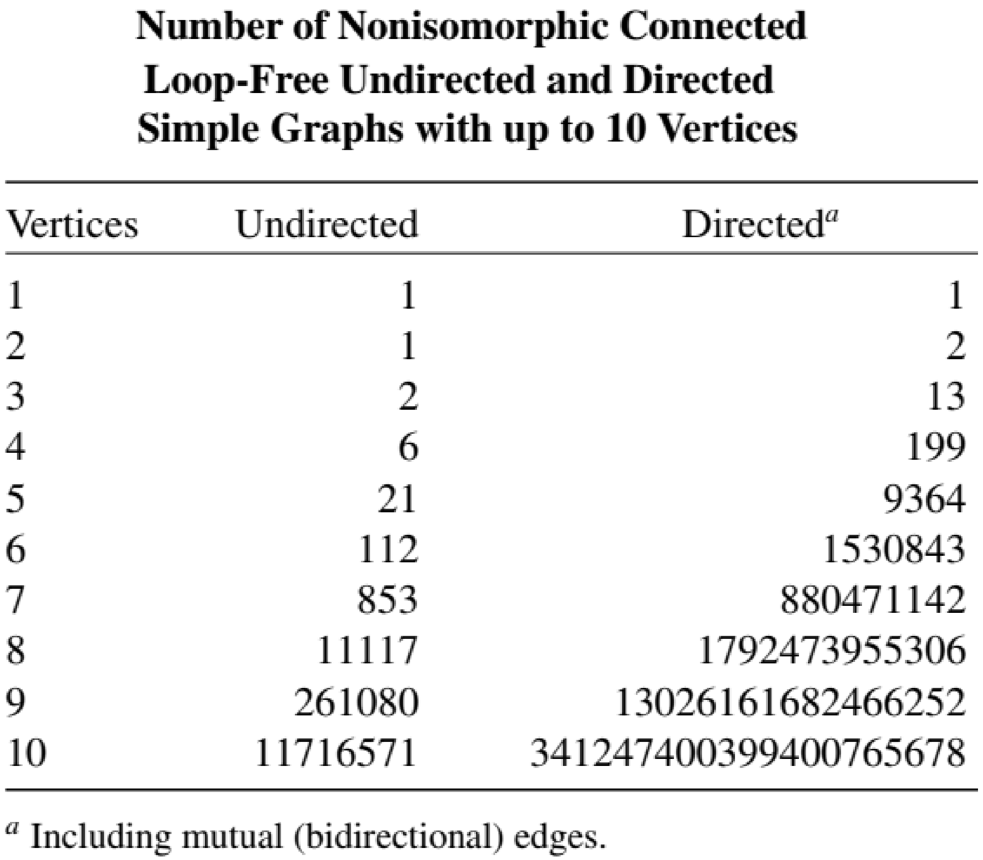
	Motif	Network					
Structure	description	GRN	NN	FW	EC1	EC2	WWW
	Feed-forward loop	✓	✓		✓		
	Bifan	✓	✓		✓	✓	
	Biparallel		✓	✓	✓		
	Three chain			✓[a]			
	Three-node feedback loop					✓	
	Four-node feedback loop					✓	
	Feedback with two mutual dyads[b]						✓

Motif finding

- NP hard problem



Structure of the 13 directed motifs with three vertices.



Number of Nonisomorphic Connected Loop-Free Undirected and Directed Simple Graphs with up to 10 Vertices

Vertices	Undirected	Directed[a]
1	1	1
2	1	2
3	2	13
4	6	199
5	21	9364
6	112	1530843
7	853	880471142
8	11117	1792473955306
9	261080	13026161682466252
10	11716571	341247400399400765678

[a] Including mutual (bidirectional) edges.

Motif finding

- Still there is a lot of room for research in this area

ELSEVIER

Computer Science Review

Volume 37, August 2020, 100267

Survey

Motif discovery in networks: A survey

Shuo Yu, Yufan Feng, Da Zhang, Hayat Dino Bedru, Bo Xu, Feng Xia

SBNA

Social & Biological Network Analysis Laboratory

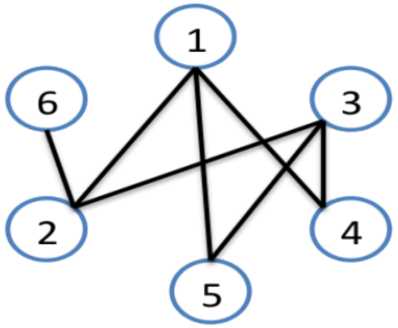
Question?

Bioinformation.ir

info@Bioinformation.ir

Ranking relations

- Link prediction



	$CN = \|\Gamma(x) \cap \Gamma(y)\|$	*Ranked List*
(1 , 3)	3	**(1 , 3)**
(1 , 6)	1	**(2 , 4)**
(2 , 4)	2	**(2 , 5)**
(2 , 5)	2	**(4 , 5)**
(3 , 6)	1	**(1 , 6)**
(4 , 5)	2	**(3 , 6)**
(4 , 6)	0	**(4 , 6)**
(5 , 6)	0	**(5 , 6)**

Predicted Edges