Multi-relational networks and multidimensional time series clustering tools for biomedical applications

- Abstract -

BABEŞ-BOLYAI UNIVERSITY
Faculty of Mathematics and Computer Science

PhD Student: Ramona-Carmen Stoica
Scientific Supervisor: Bazil Pârv

Faculty of Mathematics and Computer Science

Babeş-Bolyai University

A thesis submitted for the degree of

PhD

2017

1

Thesis contents

Contents

2

Introduction

2.1 The motivation of the research

Nowadays, the experimental data available to researchers and scientists is growing exponentially. The decrease of cost for sequencing data from biology and medicine provides a huge opportunity for data analysis, modelling, and simulation. Bioinformatics, systems, biology and biomodel engineering are considered promising disciplines aimed to study and understand biological systems. Natural sciences, like physics or chemistry, as well as engineering, use mathematical and computational modeling to better understand and analyze existing real-world systems, by creating more and more complex models and by using simulations in order to reproduce their dynamical behavior. However, most systems are so complex that all existing hypotheses can only partially explain their behavior. Therefore, a natural question to ask is: how much information about a system is needed in order to formulate a mathematical model that fully explains it?

The motivation of this thesis has come from an increasing need of developing better models for complex, large-scale systems, with the goal of gaining new insights into the behavior of those systems. Developing new methods for studying the interactions that occur within complex networks has a high impact on many disciplines, from social sciences to biological sciences. This work, however, focuses mostly on bio-entity interactions in biomedical networks. The challenges include the development of general-purpose frameworks that can be applied to many different types of biological networks, and practical tools that can provide new insights into the behavior of these networks under various conditions (such as mutations, for example).

2.2 The objectives of the research

The main objectives of my thesis deal with two bioinformatics problems.

The first area the thesis refers to is multi-relational networks (most of the biological networks are multi-relational). The second aspect envisages multidimensional time series clustering (as many of the biological data sets and experiments collect information over time).

The scope of this work concentrates around two problems and the tasks envisaged are the development and analysis of multi-relational network models, multidimensional time series clustering and pointing out some general directions for extending the models developed in this thesis that may yield additional insight. This thesis has presented two main ideas which are used herein for addressing biological problems. However, they are not specifically constructed for biology or medicine and can be applied to other fields as well.

The first subject deals with multi-relational networks (networks with multiple links of different types between two nodes). It addresses three aspects of these networks:

- Comparison of multi-relational networks,
- Finding graphlets and motifs in multi-relational networks, and
- Detecting communities in multi-relational networks.

The challenges include dealing with NP-complete problems (such as network isomorphism), finding efficient algorithms for graphlets and motifs and finding the right way to represent networks and their interactions (we have multiple networks, some of the link types might be missing in one or more of the networks, etc).

The second objective is to approach multidimensional time series types of data and to perform clustering for these entities. Most of the biological data are organized as time series. Many times, biological experiments are performed in parallel conditions or situations, or in different media. However, the final biological goal is to identify entities (genes, proteins, enzymes, etc) which work in a similar way in the same organism (or related organisms or strains) under different conditions. This involves looking at multiple time series which characterize an entity and clustering these entities based on these time series. The problem is interesting and it is useful to approach it in a different way (rather than aggregating all the time series in a single one).

An additional objective was to provide tool support for all research problems addressed in this thesis.

2.3 The thesis contributions

The contributions of this thesis can be grouped into two categories: theoretical / conceptual, and applicative / practical. In the first category we mention the MultiNetCom algorithm (described in Chapter 7), designed for detecting communities in multi-relational networks.

Also, some improvements of multi-dimensional time series clustering algorithms were presented in Chapter 4.

The second category of contributions is composed of four software tools operation on multirelational networks: (a) a network comparison tool, NCTool, (b) MultiMot tool, for finding motifs, (c) MultiNetCom tool for detecting communities, and (d) MDTSC tool, for multi-dimensional time series clustering.

NCTool takes as input two or more networks and applies some basic operations such as union, intersection, common sub-graphs, degree distribution. Such tools are designed for gene-gene interaction networks and protein-protein interaction networks but are easily extensible and adaptable. The user has the option of selecting one, two, more or all types of existing links and performing operations for the selected types of links.

MultiMot tool has to deal with motifs of a given size (sizes of 3, 4 and 5 nodes are implemented), being able to find custom design motifs (a set of such motifs at a time) and also to operate on both directed and undirected graphs.

The MultiNetCom tool implements an extension of Fruchterman's algorithm, proposed by us (originally designed for single-relational networks) for dealing with community detection problem in multi-relational networks.

MDTSC tool was designed and implemented to be able to cluster entities consisting of multiple time series. The tool is designed to have an easy to use interface.

Taking into account application areas, the thesis tackles computational tools which are mostly used to solve problems in biology and medicine, but are general enough to be applied to other similar areas as well.

Having in mind the main areas employed, we can group the contributions in two domains: *networks* (or *graphs*) and *clustering*. We look at specific aspects of these topics as multi-relational networks, which allow multiple edges of different types between two nodes. We also tackle the concept of clustering multidimensional time series, which are entities composed of multiple time series.

The main contributions of my thesis can be outlined as follows:

- Definitions of the main concepts related to multi-relational networks and extension of single-relational network attributes to multi-relational ones.
- Development, implementation and testing of algorithms for comparing multi-relational networks from biology (56).
- Development, implementation and testing of algorithms for motifs and graphlets detection in multi-relational networks. This tool has two parts:
  - motifs of a certain size (given by the number of nodes) can be found using a sampling algorithm (sizes of 3, 4 and 5 are currently implemented, but the motifs are easily scalable to larger sizes), and
  - custom-designed motifs can be found using a backtracking algorithm. Multiple motifs can be searched at once.
- Development, implementation and testing of a new algorithm for detecting communities in multi-relational networks (55).

3

Complex networks

Chapter summary: Complex systems are often analyzed as networks because of their common topology. This chapter is meant to define the concept of network and its attached notations in a more scientific approach, also discussing their classification and representation. Basic notions about networks include: network definitions, classifications, structure and representation (as matrix, adjacency list and incidence list). Some of the measurements used in networks are described (degree centrality, degree distribution, clustering coefficient, centrality). Models of networks, such as Erdos-Renyi (2) model, the Watts-Strogatz (40) model and Barabasi-Albert (2) model, required in the following chapters are also mentioned and briefly described. The last section describes some properties of local networks: networks motifs and graphlets.

3.1 Network definitions

A wide variety of natural and social systems can be described by networks with complex topology. Usually represented as random graphs, it has been widely admitted that their evolution and their topology in the real world are driven by strong rules and principles (1). This subsection introduces basic concepts of the complex networks and provides an overview of the main models covering small-world, scale-free and random networks (40).

Because of their similarity with graphs, networks considered as being much larger graphs are defined by the same data structures and relations:

- Let $G = (V, E)$ be a representation of a graph, where V represent the set of nodes (or vertices) and E represents the set of edges (or connections, links between the nodes) (10).
- Let $|V|$ be the number of vertices (cardinal of V) in the graph. Let $|E|$ be the number of edges in the graph.
- Let $Deg(v), v \in V$, be the number of links (edges) connected to v.
- A path between the vertices $s, t \in V$ is defined by an alternating sequence of nodes and links, starting with s and ending with t. Each link connects its adjacent vertices (10). The path length is determined by the sum of the weights of its links according to the weight function (10).

Networks classifications are made by at least three criteria: application area, structure, and their model:

1. Their applicability in complex systems such as:
   - Computer networks
   - Transport networks
   - Neural synaptic connections networks
   - Brain functional networks
   - Disease (disease gene)
   - Ecological networks which synthesize ecosystems
   - Telecommunications networks
   - World Wide Web
   - Social Networks
   - Biological Networks, etc.
2. Their structure:
   - Let $\omega(e), e \in E$, be the weight function representing its links.
   - An un-weighted, simple network has $\omega(e) = 0$ for all links and only one link e $\in E$ between two vertices $u, v \in E$ (10).

  - A weighted network has the same structure for links as the un-weighted one but with $\omega(e) \in R$ (61).
  - A multiple network allows multiple links between the same pair of vertices and it can also be classified by the edge weights: it either can be weighted or un-weighted (61).
  - A directed network is described by the set of edges which have a direction associated with them; it can also be weighted or un-weighted (61).

3. Their nodes and degree distribution which represent models of networks:

  - A small-world network has a special vertices distribution: the distance D between two vertices grows proportionally to the logarithm of the number of vertices $|V|$ in the network (61).
  - A scale-free network (most biological networks are alike) respects a power-like distribution of edges in which each nodes degree respects a power formula (61).
  - A random network has no degree restriction: no law of the nodes degrees is respected (61).

There may be a more detailed classification possible with more information to offer, but these classifications are general and cover the most popular types of networks.

3.2 Network representations

As defined above, a network is defined by $G(V, E)$ having the set of vertices V and the set of edges E. Networks can be represented in multiple ways, depending on how the links between nodes are structured:

- a matrix representation,
- an adjacency list representation,
- an incidence list representation,
- an incidence matrix representation.

4

Biological networks

Chapter summary: This chapter describes different types of biological networks and presents some common aspects between biological and real-world networks. There are several types of biological networks and not all of them will be studied in this work. This chapter insists on some types of networks which will be used in experiments or networks similar to those, and could be approached with further extensions of the tools implemented in this thesis. Some of the networks of interest include the interactions between genes, interactions between proteins, interactions between biochemical elements (chemical reactions) and metabolic networks.

4.1 Metabolic networks

The activity of a cell is established on complex networks of interacting chemical reactions precisely organized in space and time which produce observable cellular functions.The process that helps us to identify the total reactions that compose a network is called network reconstruction. The sum of physical and metabolic processes that determine the characteristics of the cells, all biochemical and physiological, form a metabolic network.By its very nature, these networks compose the chemical reactions of metabolism, the metabolic pathways, and in the same time the regulatory interactions that guide these reactions (20, 24, 28, 48, 52). Intermediary metabolism can be considered as an chemical tool that converts the basic materials into energy as well as the building blocks needed to produce biological structures, support cells and maintain the different functions of the cells. This chemical tool is extremely changing, accept the laws of chemistry and physics, and for this reason it is limited by differently physicochemical constraints. In the same time, it has an complicated regulatory structure which allows it to answer to a mix of external perturbations. Metabolic imbalance is the root to all main human diseases as in heart diseases, cancer, diabetes and obesity. Metabolism is composed by two different types of chemical transformations such as: catabolic pathways which decomposed substrates into simple metabolites, and anabolic pathways which synthesize the aminoacids, nucleic acids, fatty acids and other necessary building blocks. Over these processes, there is a complicated exchange of different chemical groups and reductionoxidation (redox) potentials show up over a set of carrier molecules. These transporter molecules and the properties that they transfer thus link the metabolic network tight together.

Hierarchy in metabolic networks There are four levels of functional decomposition of metabolism, as follows(43):

The metabolic networks are difficult to understanding for human mind because genome-scale reconstructions of metabolic networks is composed of large processes which consists of multitude of metabolites and once in a while over a thousand reactions. Hence, we need mathematical models to study their properties and simulate their functions. Nonetheless, we can see the properties of a network in a hierarchical mode to make easy (streamline) the approach (theory) of network functions. There are four levels of functional decomposition of metabolism, as follows (43):

- Level 1: the entire cell
- Level 2: the metabolic sectors
- Level 3: pathways
- Level 4: the reactions which are individual

There are two reconstruction methods, as follows:

- Genome-scale metabolic reconstruction
- Multiple Genome-scale networks

In metabolic networks we have metabolites and metabolic pathways. Metabolites are microscopic particles like glucose as well as aminoacids, or macromolecules like polysaccharides as well as glycan (carbohydrates). Metabolic pathways are sequences of consecutive biochemical reactions for a particular metabolic function, e.g. glycolysis or penicillin synthesis, that convert one metabolite into another. Enzymes are proteins that catalyze (accelerate) chemical reactions. In this way, in a metabolic pathway the metabolites and the enzymes are the nodes correspondents and the metabolic reactions correspond to directed links. The simpler approaches state that nodes represent metabolites and directed edges are reactions that convert one metabolite into another. Examples of a part of a glycolysis pathway, a metabolite-centric representation, and reactions and metabolites are given in figures 4.1, 4.2, and 4.3:

In this way, in a metabolic pathway there are nodes that correspond to metabolites and enzymes. Directed edges correspond to metabolic reactions. The simpler approaches state that nodes represent metabolites and directed edges are reactions that convert one metabolite into another. Examples of a part of a glycolysis pathway, a metabolite-centric representation, and reactions and metabolites are given in figures 4.1, 4.2, and 4.3:

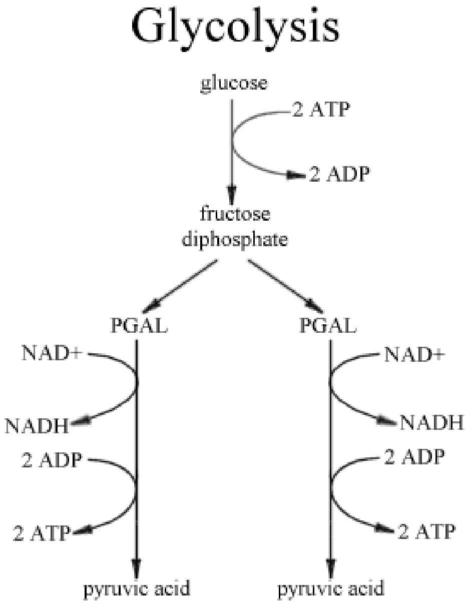


Figure 4.1: Part of the glycolysis pathway.
source: `http://library.thinkquest.org/27819/ch4_4.shtml`

All metabolic pathways of a cell form a metabolic network. Such a network is com-

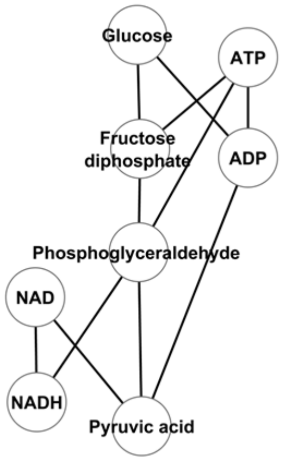


Figure 4.2: Metabolite-centric representation.

posed of complete view of cellular metabolism and material or mass flow. Cells rely on this network for digesting substrates from the environment, generate energy and synthesize needed components from the environment for their growth and survival. These networks are also useful for curing human metabolic disease through better understanding of metabolic mechanisms, or for controlling infections of pathogens by understanding the metabolic differences between the human and pathogens (59).

Metabolic pathways are constructed partially experimentally and partially from the genome sequence (homology). These networks are general, and they are used for many organisms, from bacteria to human(29). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a large collection of online databases have to do with genomes, enzymatic pathways and biological chemicals.

KEGG is a database created with the goal of understanding the functions and the utilities of the biological systems. It helps to learn biological systems like the organism, the cell and the ecosystems from molecular and genomic point of view. It consists of the following databases: chemical, network information and genomic (30).

The biological system is depicted in the Figure 4.4. The representation include molecular building blocks of the genes and proteins (i.e. genomic information) and chemical substances (chemical information) that are combined with the knowledge of

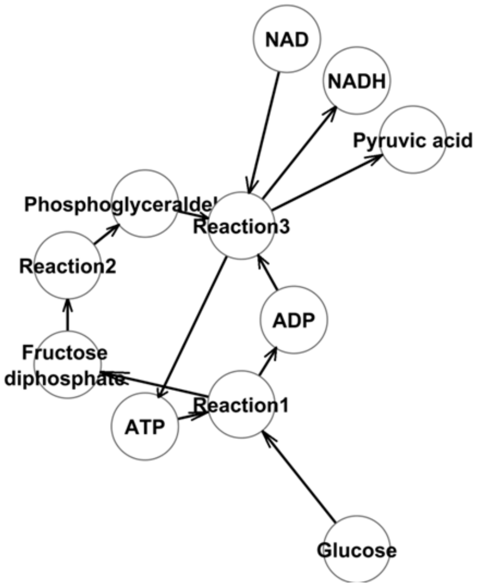


Figure 4.3: Reactions and metabolites.

molecular wiring diagrams of interaction, reactions and the relation between networks (systems information) (29).

4.2 Gene-Gene interaction networks

Gene-Gene interaction (GGI) networks are the networks in which the nodes are genes (from a certain organism) and the links are connections between them. These connections are obtained from various studies, from experiments and from the numerous existing databases. There are a couple of GGI tools but one of the most common and flexible among them is Genemania (39, 45). The interactions between genes included in Genemania and in most of the other GGI networks are the following (60):

- **Co-expression.** Two genes are co-expressed if their expression levels are similar according to certain conditions in a gene expression study.
- **Physical Interaction.** Two genes physically interact if they were found to interact in a protein-protein interaction study or database.
- **Genetic interaction.** Two genes have a genetic interaction between them if changes to one gene have an effect on changes to the other gene.

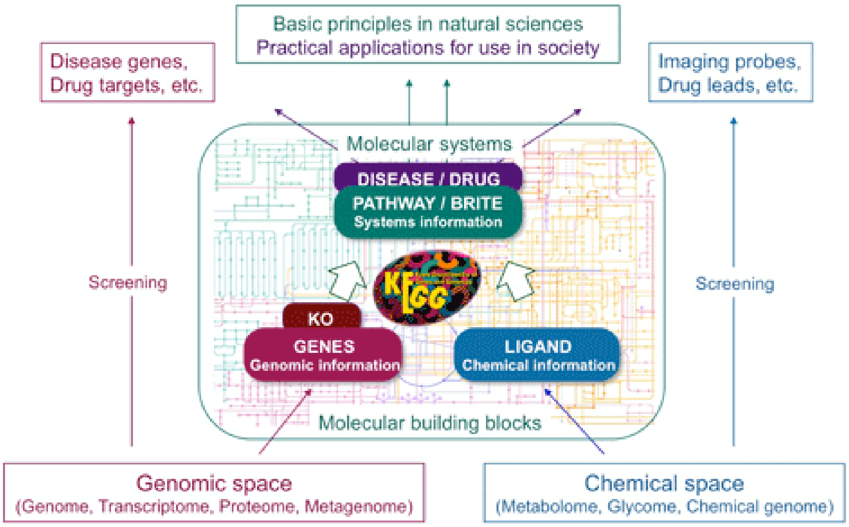


Figure 4.4: Digital representation of the biological system source: `http://www.genome.jp/kegg/kegg1a.html` (29)

- **Shared protein domains.** Two gene products (i.e. proteins) have this interaction type if they have the same protein domain.

- **Co-localization.** Two genes are co-localized if they are both expressed in the same tissue or if their gene products are both identified in the same cellular location.

- **Pathway.** Two gene products (i.e. proteins) are linked if they are part of the same reaction within a pathway.

- **Predicted.** Two genes have a predicted functional relationship if there are known functional relationships from another organism via orthology.

- **Other.** All other relationships: chemical genomics data, phenotype correlations from Ensembl, disease information from OMIM, etc.

5

Clustering in complex netwoks

This chapter presents the basic clustering notions, the measures for similarity and dissimilarity and several clustering algorithms. It then describes time series clustering and some existing methods for this. The multidimensional time series clustering is further presented and the extension from single time series clustering to multidimensional time series clustering is presented, together with the practical importance and need for this. An algorithm for dealing with multidimensional time series data is designed and implemented. The tool tested on some real benchmarks.

5.1 Basic notions

Clustering is the process of splitting objects in a set into distinct subsets, in a way that makes similar objects end up in the same subset. It belongs to the class of unsupervised learning methods as there is no label or class assigned to the objects and no indication on how the objects should be grouped.

A *cluster* contains objects that satisfy the following properties:

- the objects are very similar among them in the same cluster
- the objects are very different (dissimilar) from the objects in the other clusters

The literature contains many definitions of similarity and dissimilarity measures (15). These are building upon on:

- the type of the data dealing with

- the desired similarity type

Usually, the similarity and dissimilarity measures are formulated in terms of a distance function $d(x, y)$. Ideally, the chosen function must consider the next restrictions (19, 64):

- $d(x, y) \geqslant 0$
- $d(x, y) = 0 \iff x = y$
- $d(x, y) = d(y, x)$
- $d(x, z) \leqslant d(x, y) + d(y, z)$

5.2 Time series clustering

Time series are sequences of real numbers that correspond to the values observed for some parameters at equal time intervals.

Time series can be continuous (the variable is defined at all points in time), or discrete. The time series involved in cluster analysis are usually discrete and are a mixture of the next components (7):

1. a trend (the long-term movement) (7),
2. trend fluctuations of greater or lesser regularity (7),
3. a seasonal component (7),
4. a residual or random effect (7).

Clustering time series has many real-world application in a variety of fields. Time series data are prone to outliers, especially when the data sets are very large. Their elements have a temporal ordering and operations with such data types are common in data mining. A lot of research has gone into developing algorithms for clustering time series, and the effectiveness of each approach is tested on various real life applications, in order to encourage the development of even better algorithms (3, 6, 13, 21, 23, 34, 50, 58).

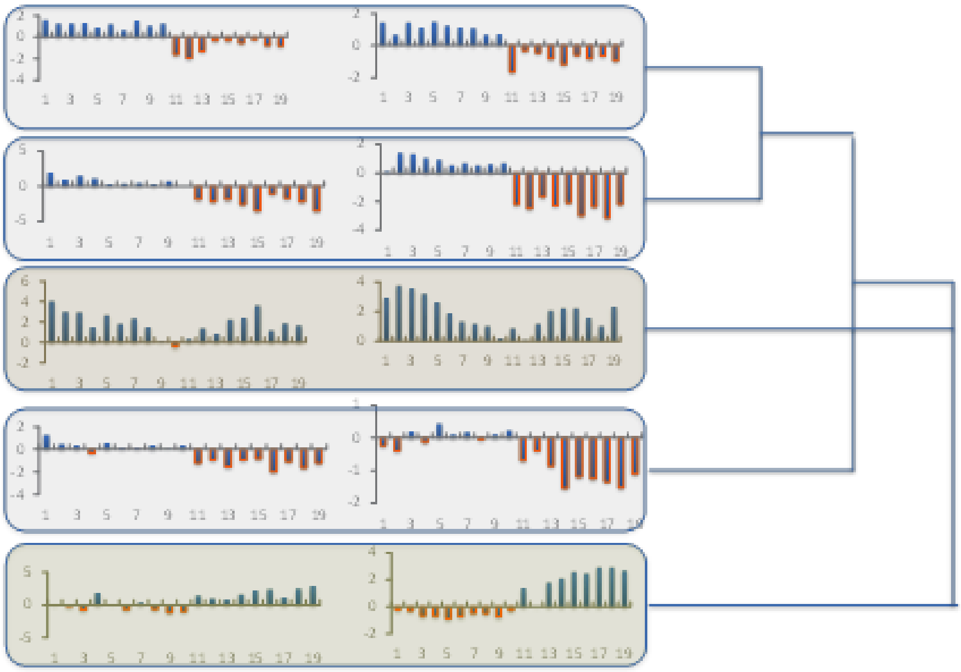


Figure 5.1: Two dimensional time series: example of hierarchical clustering

5.2.1 The multidimensional time series clustering

A time series is defined as an array $X = (x_1, x_2, \ldots, x_n)$ of measurements in time for a given parameter (or variable).

A *multidimensional time series* is represented as:

$$\mathrm{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}$$

where each $X_i, 1 \leq i \leq N$, is a time series on its own. Each individual time series can have a different size.

Clustering multidimensional time series involves grouping entities of form X. Figure 5.1 presents 5 entities that are hierarchically clustered as 2-dimensional time series. Multidimensional time series are particular in situations when the entities present multiple simultaneous measurements which are taken into account at the same time.

When clustering time series data, one has to group together series of time points: for instance, clustering a number of cities based on the temperature measurements for each city during a period of two years. Each data is formed of 730 two-dimensional time points (two years of 365 days). While transposed in the multidimensional case, each data is composed of multiple time series. For example, we also want to cluster cities based on daily measurements for wind speed, pressure, precipitation volume, etc, and not on temperature measurements alone. In the figures below we present examples of entities which have two time series each. Some of the instances are more similar with consideration to the measurements in the first time series while the others are more similar with consideration to the second time series. In the multidimensional case, we want to cluster entities which are in the same time similar with respect to both time series, overall. This example is illustrated in Figure 5.2.

The common approach in the multidimensional case is to concatenate all the time series into a single one and to transform the problem in a single dimensional time series problem. But this can lead to loss of general aspects of the problem. Therefore, it's advantageous to deal with multidimensional time series without transforming them: on one hand, they offer a global point of view and show some critical pathologies arising from evident discrepancies, and, on the other hand, they permit the integration of the information contained in each one-dimensional time series of X and therefore it is useful when each array is sparse and short (17).

5.2.2 A variant of multidimensional time series data clustering

A similarity measure is usually employed to calculate the similarity between two time series. In this chapter, we treat the difference between each time series of a multidimensional time series instance as an objective function that has to be minimized.

5.2.2.1 Similarity measure

To compare the similarity of two objects X and Y, where X and Y are given by:

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}, Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix}$$

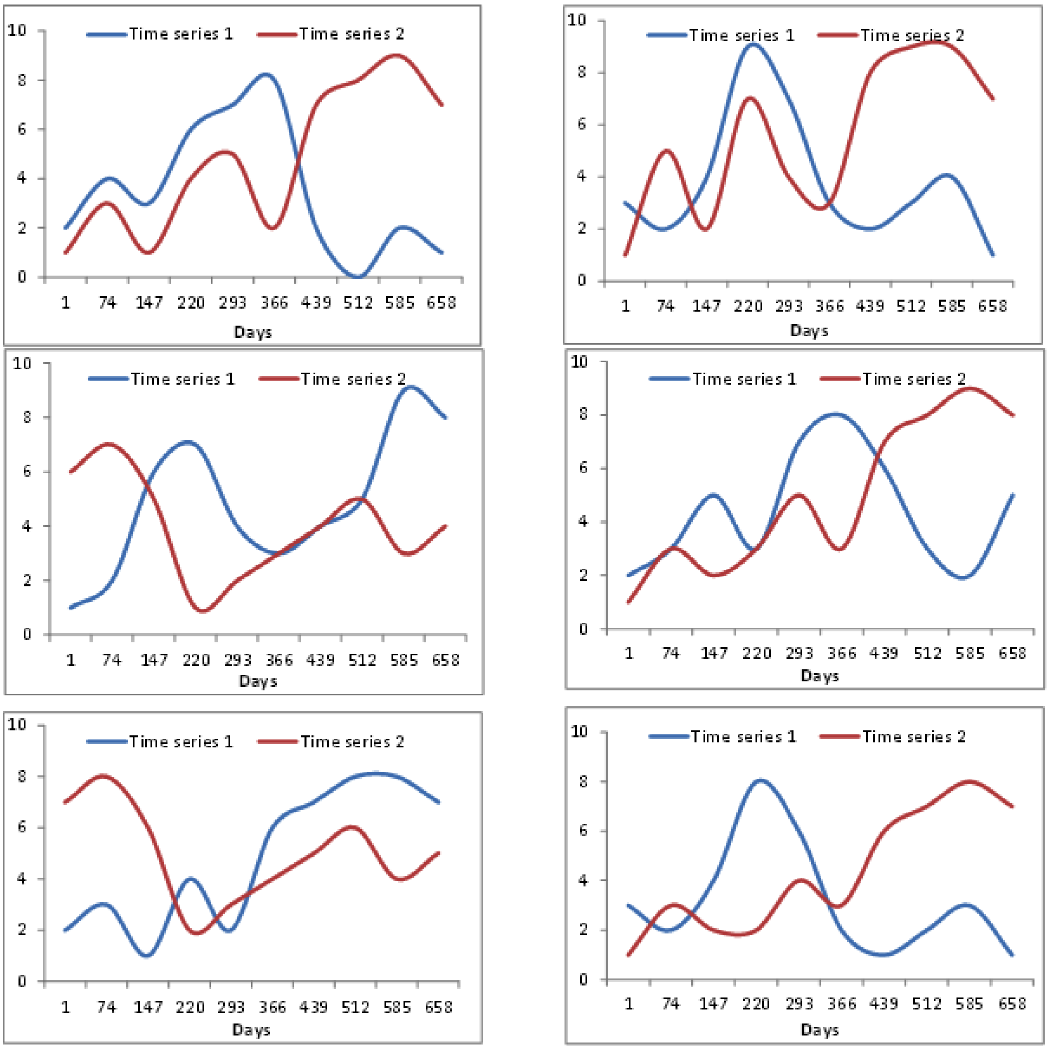


Figure 5.2: The two sets of measurements (there are two time series) over the course of two years (730 days)

we define an N dimensional objective function $F = (f_1, f_2, \ldots, f_N)$ as:

$$F = \begin{pmatrix} f_1 = d\,(X_1, Y_1) \\ f_2 = d\,(X_2, Y_2) \\ \vdots \\ f_N = d\,(X_N, Y_N) \end{pmatrix}$$

where $d(\cdot)$ defines a similarity measure.

We cluster multidimensional time series using the k-means clustering algorithm. We use function F to determine the cluster to which each item should be assigned. We compute the similarity value as a linear combination of $f_i, 1 \leq i \leq N$, and we denote the result by d_{sim}:

$$d_{sim} = \sum_{i=1}^{N} w_i f_i$$

where w_i are the weights that determine the importance of each time series in the clustering. In the experiments performed, the time series have equal importance ($w_i = 1, 1 \leq i \leq N$), and the next mentioned distances were implemented $d(\cdot)$:

- Euclidean distance
- Manhattan distance
- Maximum distance
- Average distance

5.2.2.2 Parameter setting: the value of k

The user can select one of these four measures from the main menu, as well as the Maximum Distance Percent (set to 0.6 by default in our experiments). The initial value of k is set to the total number of items in the data set. After every iteration, seeds that do not have any data point assigned to them are removed. The algorithm goes through as many iterations as needed for the clusters to stabilize (low distances between points in the same cluster, and large distances between any two points that belong to different clusters).

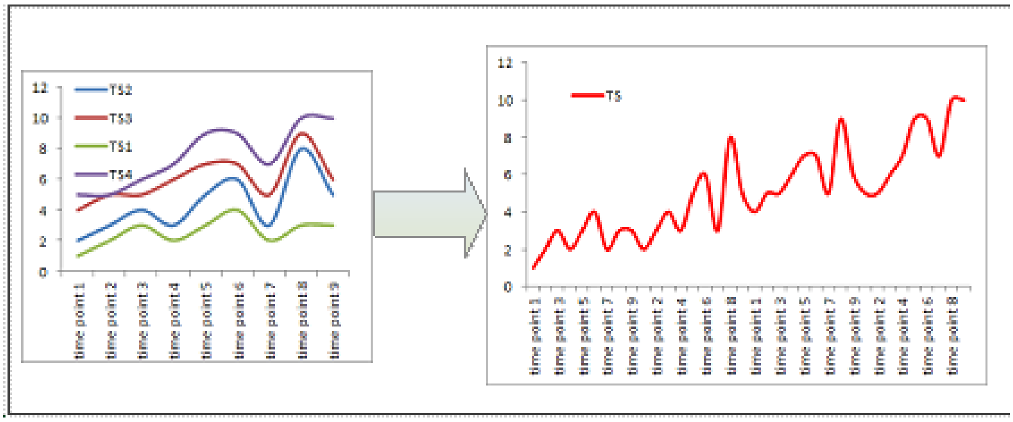


Figure 5.3: Comparison with traditional methods

5.3 The comparison with the traditional methods

The classical methods usually pool the data: they infer a single parameter that characterizes all time series (18). So essentially a multidimensional time series is converted to a single dimensional one, as show in Figure 5.3. We implemented this technique too, compared the results with those yielded by our approach, and noticed that the traditional method usually tends to split the data into more clusters than expected. This was especially obvious when using the Average and Manhattan similarity metrics on the first data set, or when using the Average measure on the second data set.

5.4 Summary

Multidimensional time series are a generalization of the single-dimensional time series. They have more parameters that are used to describe each data instance, and therefore, clustering them is harder than clustering single-dimensional time series. In this chapter, we approached the multi-dimensional time series clustering problem as a multiobjective problem and implement several geometrical distance measures in the k-means clustering algorithm in order to test the outcome of the similarity measure along with process of clustering. We validated the results on three data sets, and concluded that the most commonly used distance for clustering single-dimensional time series (the

Euclidean distance) might not be the most appropriate measure for clustering multidimensional time series. These results were published at the 2013 Intelligent System Design and Applications (ISDA 2013) conference (54) (poster publication) and in the Studia Universitatis Babeş Bolyai journal in 2013 (53).

6

Comparison of multi-relational networks

6.1 Introduction

In this chapter we introduce the definitions and concepts related to multi-relational networks as compared to single-relational ones. Some useful properties of these networks are presented and a tool developed for network comparison is described, together with its main operations. The last section makes an analysis of experiments on comparing gene-gene interaction networks for various types of cancers performed in order to validate the ideas and concepts included in this chapter.

6.2 Definitions

The term of network is interpreted as a group of people, organizations, places, etc. that are connected or work together (8). In computer science, a network is described as a large graph that has a set of vertices connected by edges. The first distinction between simple graphs and multigraphs is the number of interactions between the vertices: simple graphs have at most one interaction (edge), while in multigraphs there can be multiple interactions between any two nodes. A multigraph has a set of vertices and a multiset of unordered pairs of vertices defining the edges. A multigraph has all edges of the same type. In the case of (biological) networks, there can be many types of interactions between two vertices, so a biological network is a generalization of multigraphs.

Combining these definitions in bioinformatics, we can say that a biological network is made of a set of vertices (genes, proteins, metabolites, etc.), and a set of edges representing the interactions (of different types) between vertices.

6.3 Multi-relational networks

Analyzing and studying networks has become increasingly important in a variety of fields such as biology, computer science and sociology. In this chapter, we are focusing on biological networks. We the recent advancements in biology, a large amount of data has become available. However, due to the size of this data, mining and studying it comes with its challenges. One has to come up with ways to abstract this data into models that are easier to understand. One such model that has proven to be very effective at describing complex relations and interactions in the biological systems is biological networks.

A network is described as a set of nodes (vertices) and a set of links (edges) having various types of relations between the nodes. In biology, networks are used to describe both the structure and the dynamics of a biological system, and therefore, identifying biological networks is essential in systems biology.

Biological networks are composed of biological entities and the interactions between them (41), and can be used to describe many different types of relations. Some common examples of biological networks are: transcriptional regulatory networks (GRNs),metabolic and biochemical networks and the protein-protein interaction networks (PPI). These biological networks are usually multi-relational: two nodes can have multiple types of interactions between them, and those interactions are represented by links in the network. For example, the metabolic pathways can be described as multi-relational networks; the molecules are the nodes, and the enzyme activities, signal transduction or chemical reactions are the edges (41). In contrast to single-relational networks that allow a single type of edges between the vertices, multi-relational networks are much more suitable for representing real-world interactions, but are also much harder to analyze.

6.4 Experiments and results

6.4.1 Datasets for multi-relational networks comparison

Our tool was tested on three datasets of genes that are representative of endometrial, ovarian and breast cancers. We used the Genemania application to create the networks and save them to files, and we converted these files to the format accepted by our tool: each edge is represented in the form $(gene1, gene2, relation\ type)$; where *gene1* and *gene2* represent the nodes and *relation type* defines the type of link between the two nodes.

6.4.2 Results

We performed experiments on three types of cancer: breast, ovarian and endometrial. The input data was taken from the database made available by the Sanger Institute (5), the Human Gene Compendium at the Weizmann Institute of Science (47) and by a few other studies. The interactions for the most significant genes related to each type of cancer were extracted from Genemania (4, 41) and are shown in Figure 6.1 (the visualization uses a degree sorted circle layout and was drawn using the Cytoscape tool (11, 47, 51)).

The details for each of the three networks are shown in the Table 6.1.

6.5 Summary

In this chapter we presented all basic definitions and notions required for the extension of a single-relational to a multi-relational network. Multi-relational networks can be applied to problems in many distinct areas, and biological networks are one of their most common applicability. The chapter also presented a tool which is designed for comparing

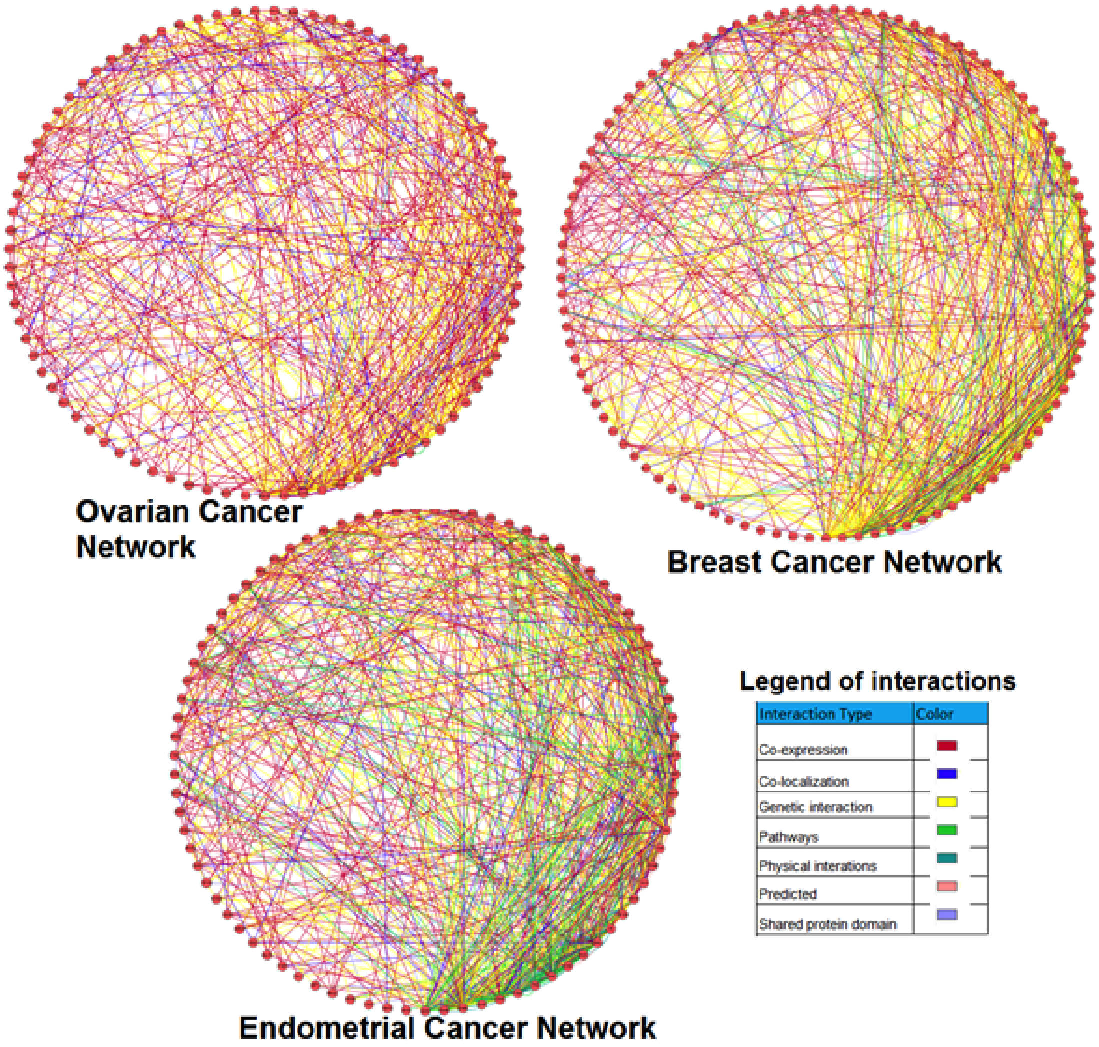


Figure 6.1: Representation of gene-gene interaction networks for the thee types of cancer: ovarian, breast and endometrial (Cytoscape tool (11, 47, 51) has been used to draw them).

	Ovarian cancer	Breast cancer	Endometrial cancer
Nodes	83	101	83
Edges	590	817	801
Multi-edge node pairs	47	74	108
Isolated nodes	0	0	0
The density of the network	0.157	0.14	0.19
The heterogeneity of the network	0.39	0.48	0.41

Table 6.1: Description of networks of interactions between the genes for the ovarian, breast and endometrial cancer.

multi-relational networks. Experiments on real data related to various types of female cancer were performed and described. The experiments included genes for three types of related cancers (ovarian, breast and endometrial) and were presented at International Conference on Intelligent Systems Design and Applications (ISDA) (56).

7

Network motifs and the MultiMot tool

7.1 Introduction

This chapter presents network motifs and graphlets and includes three sections. In the first section, the basic notions about motifs and graphlets are presented. We also describe the meaning of motifs and graphlets and their importance in biology. An extension of motifs and motif finding in the case of multi-relational networks is performed and the basic operator in this case is explained. A tool dealing with finding motifs in multi-relational networks, which is an extension of a famous tool developed in Uri Alons lab in Weizmann Institute is described in the second part of the chapter. The extension does not only refer to multi-relational networks (extended from single relational networks) but also to a user defined motif search tool. This user (or custom) defined motif finding part of the tool allows the usage of wildcards and also allows searching for more motifs or graphlets at the same time. Numerical experiments including real biological data are performed in the last part of this chapter in order to validate the efficiency of this tool.

7.2 Motifs in multi-relational networks

The amount of computational power that became available in the last few years has enabled the study of very large networks: the World Wide Web, social networks, and in particular, biological networks, among which networks for protein-protein interaction (PPI) (22, 26, 27), metabolic networks (28) and gene-gene interactions networks (GGI) (35, 38). These networks bring a wealth of data, but extracting meaningful signals from this data has its challenges. One has to develop algorithms which are efficient and at the same time resistant to errors found in the data being analyzed.

Unfortunately, due to their immense complexity, some networks cannot be fully analyzed even with all the computational power available today. The human genome is a classic example of such a network. These networks are usually extremely important too; therefore, gaining any new insight into their behavior is highly desirable. And one way to accomplish this is by drawing parallels with simpler networks that are easier to analyze. Once a subsystem in a simpler organism is well understood, that knowledge can be carried over to the more complex organism. This approach essentially makes it possible to isolate and analyze independently small subsystems in complex organisms.

One class of such algorithms is *finding network motifs* (38). These algorithms aim to detect patterns of connectivity that occur in a network significantly more often than expected, and also provide insights into the modularity and structure of a network (22, 35, 44, 63). It has been shown that the same motifs can be found in many different organisms (9, 25, 33, 46, 57), which makes them an essential tool for transferring knowledge from simpler organisms to subsystems of more complex ones.

In addition to their applications in biological networks, network motifs have been successfully applied to other areas too (36, 37, 42).

Biological networks present a large volume of data modeled as interaction networks, metabolic and signaling pathways, regulatory networks, etc. Multi-relational graphs are best used to show their complexity and large scale. The features of this kind of networks are of utmost importance in the analyzis of interactions between biological entities. They are called motifs in graphs.

In the general case, biological networks (and network motifs, too) are multi-relational. This complicates the process of motif finding.

There exist several tools (32, 62) which, in simple directed and nondirected graphs, are identifying motifs of different sizes and shapes within a network. It is not the case for multi-relational networks. We propose a web-based bioinformatics tool, which is straightforward and able to detect motifs in multi-relational networks. We provide two motif identification modes. First, in template matching approach, the end-user specifies one or multiple graph templates to search. As specified by the user, template matching works for both directed and nondirected graphs. Additionally, the user has the option to exclude link values. In this particular situation, the algorithm searches for sub-graphs that are matching the number of links specified by the user (i.e. two matching vertices are associated by at least the number of links chosen by the user), by ignoring the links labels. Second, the subgraph sampling approach is a variation of the algorithm for for multi-relational graphs (31). It estimates the frequency of size N subgraphs by using k subgraphs obtained randomly from the network.

The method is implemented for directed subgraphs. It considers the edge labels. MultiMot can run via a web interface and is also available as an executable jar package.

7.3 Experiments and results

7.3.1 Datasets for motifs finder in multi-relational biological networks - MultiMot tool

We consider the network defined by the following triples:

$node_a$ $node_b$ $edge_1$

$node_a$ $node_d$ $edge_3$

$node_e$ $node_a$ $edge_1$

$node_b$ $node_e$ $edge_2$

$node_d$ $node_a$ $edge_1$

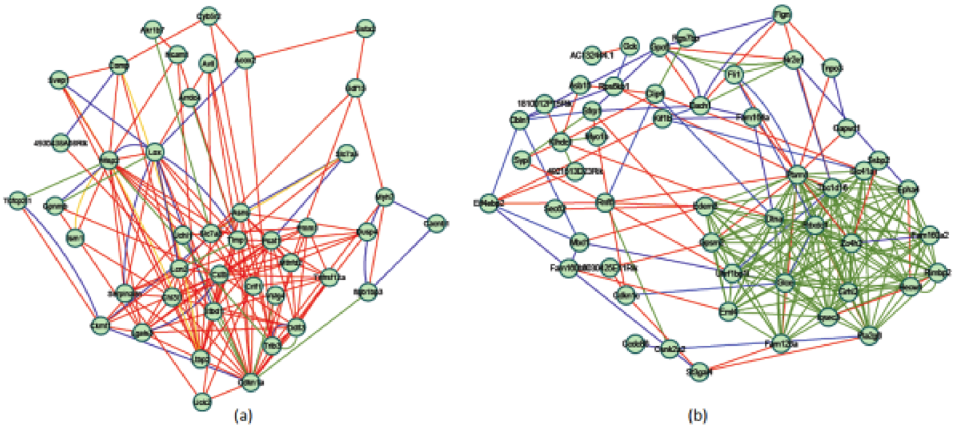


Figure 7.1: Representation gene-gene interaction networks of a heart tissue differentially expressed.

$node_e$ $node_f$ $edge_1$

where $node_x$ denotes a node (x being its label) and $edge_i$ denotes a relation between two nodes (i is the type of relation).

7.3.2 Results

To illustrate how MultiMot works, we selected a multi-relational gene-gene interaction network example. We considered top 50 unregulated genes of a cardiac tissue of two mouse models. The data is taken from the Gene Expression Omnibus database (14). The gene interaction networks (depicted in Figure 7.1) have been generated using Genemania.

Gene-gene interaction networks of the differently expressed heart tissue gene-gene interaction networks are represented in Figure 7.1. The link types represented by their color are given in Figure 7.2.

If the method is asked to find all motifs of size 3 in each of the networks, then the result is the one in the Figure 7.3 and Figure 7.4.

Links type legend are represented in Figure 7.2.

Motifs of size 3 found by MultiMot in the networks Figure 7.2 (a) and Figure7.2 (b)

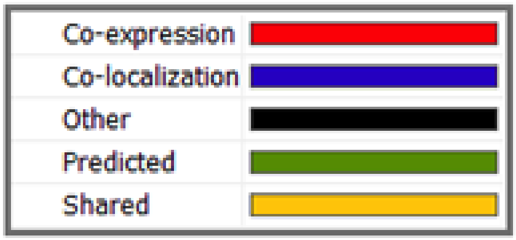


Figure 7.2: Representation of legend of gene-gene interaction networks of a heart tissue differentially expressed.

are represented in Figure 7.3 and Figure 7.4.

7.4 Summary

Detecting related and often repeated patterns (i.e. network motifs) over networks provides useful insights for a better understanding of the biology of the disease and is of immense impact in biological studies. This work introduces and presents a novel pattern finding algorithm which is an extension from single relational networks and deals with heterogeneous multi-relational graphs. It not only finds standard motifs of a certain size, but also looks for user (custom) defined motifs (which can have a higher degree of complexity). It considers both directed and undirected graphs, can search the motifs on a subset of edges and can employ wildcards on edge labels to increase generality, if required.

This work introduces and presents a novel pattern finding algorithm which is an extension from single relational networks and deals with heterogeneous multi-relational graphs. It not only finds standard motifs of a certain size, but also looks for user (custom) defined motifs (which can have a higher degree of complexity). It considers both directed and undirected graphs, can search the motifs on a subset of edges and can employ wildcards on edge labels to increase generality, if required.

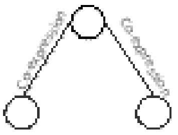
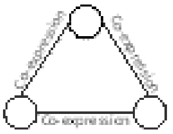
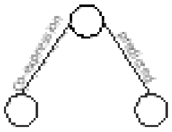
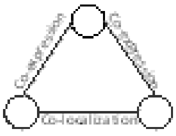
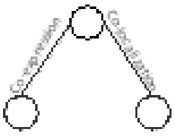
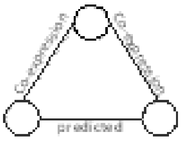
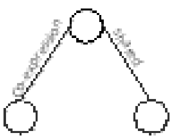
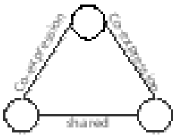
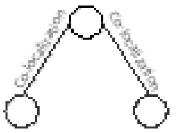
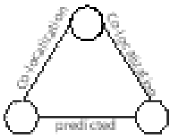
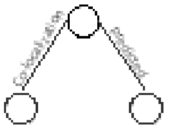
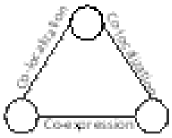
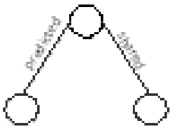
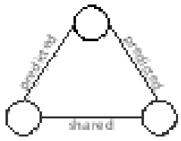
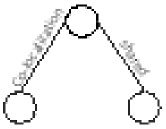
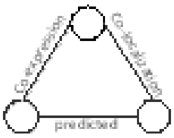
Motif	Network (a)	Network (b)	Motif	Network (a)	Network (b)
	223	22		63	5
	48	69		6	1
	60	19		2	4
	17			3	
	1	7		1	1
	4	19		4	4
	1			1	
	1			1	1

Figure 7.3: Frequency of motifs of size 3 found by MultiMot in the networks from Figure 7.1 (a) and Figure 7.1 (b).

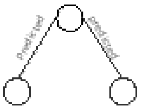
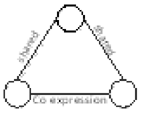
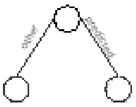
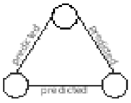
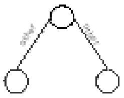
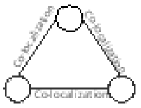
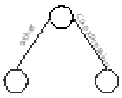
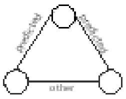
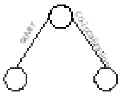
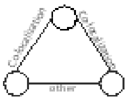
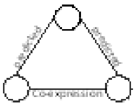
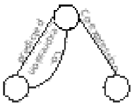
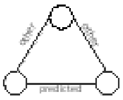
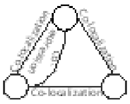
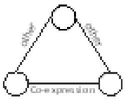
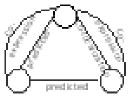
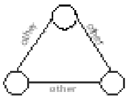
	2	219		1	
	31			0	20
	4			0	6
	9			0	1
	6			0	8
				0	10
	0	1		0	1
	3	0		0	1
 Co-expression predicted predicted Co-expression predicted	0	1		0	1

Figure 7.4: Frequency of motifs of size 3 found by MultiMot in the networks from Figure 7.1 (a) and Figure 7.1 (b) (continued).

8

Communities in multi-relational networks

8.1 Introduction

This chapter gives an overview of communities in single relational and multi-relational networks, and several algorithms for communities identifications in single relational networks. It also proposes a new algorithm for communities detection in multi-relational networks. Further, in this chapter we present a new tool implemented by us, based on the new algorithm proposed, and the results gathered by running the algorithm on several data sets.

8.2 Communities in multi-relational networks

Complex networks are mathematical models which are used to represent in a machine readable form many interaction phenomena that take place in the real world. The following networks are examples of large networks:

- The World Wide Web: vertices are depicted by web pages and links are represented by hyperlinks.

- Metabolic networks: vertices are depicted by enzymes and links are described by reactions between them.
- Facebook: vertices are represented by user profiles and links are represented by friendships.

Communities can be defined as disjoint groups of entities in a graph, such that each entity is "closer" to all other entities in the same group than to the entities outside it. Detecting communities in a network is an important step in identifying patterns in that network (12).

Detecting communities is very important in computer science and biology where information are represented as networks. Most networks in nature are multi-relational. However, detecting communities in multi-relational networks is hard, because the complexity increases with the size of the network and the number of connection types. Therefore, when looking for communities, it is often easier to reduce the multi-relational network to a single-relational one. This inevitably leads to loss of information, and the results are not always accurate.

8.3 Experiments and results

8.3.1 Datasets for detecting communities in multi-relational biological networks - MuliNetCom tool

We consider 11 datasets taken from (16) and from (49). They contain genes involved in 11 types of cancer such as: acute lymphoblastic leukemia (ALL), acute myeloid leukemia (AML), breast carcinoma (BC), colorectal adenocarcioma (CA), non hodgkin's lymphoma (NHL), non-small cell lung carcinoma (NSCLC), spitzoid tumour (ST), gastric cancer (GC), liver cancer (LIC), lung cancer (LUC), nervous system cancer (NSC). We then use Genemania (60), a software platform that generates gene-gene interaction networks from millions of publications available in various medical and biological databases. We obtain one multi-relational network for each type of cancer. Genes represent the nodes of the network and the interactions between genes represent the links or the edges. In the networks we obtained, the following types of interactions have been generated: predicted interactions, co-localization, co-expression, shared protein domains, physical interactions, genetic interactions, pathway interactions.

Table 8.1 contains the details of each of these networks: number of nodes, number of links, and the number of different types of links.

Cancer data			
Cancer type	No of nodes	No of edges	Types of edges
ALL	47	260	6
AML	92	822	7
BC	26	163	7
CA	39	411	7
NHL	19	49	6
NSCLC	35	166	6
ST	21	52	5
GC	511	28148	7
LIC	322	9694	7
LUC	74	687	7
NSC	47	20276	7

Table 8.1: Description of data sets used in experiments.

8.3.2 Results

We perform two tests for each network, separately: for the first one, we reduce the multi-relational network to a single relational one and then apply the original Fruchterman's algorithm to get the communities. For the second test, we consider the network as it is, multi-relational, and then apply the modified version of Fruchterman's algorithm proposed in this work to get the communities. Figure 8.1 shows the number of communities and the number of genes in each community obtained by each algorithm for all 11 datasets.

From the results presented here, it is obvious that the approaches obtain different communities. The number of communities obtained for the single-relational network is almost always different from the number of communities obtained by the multi-relational

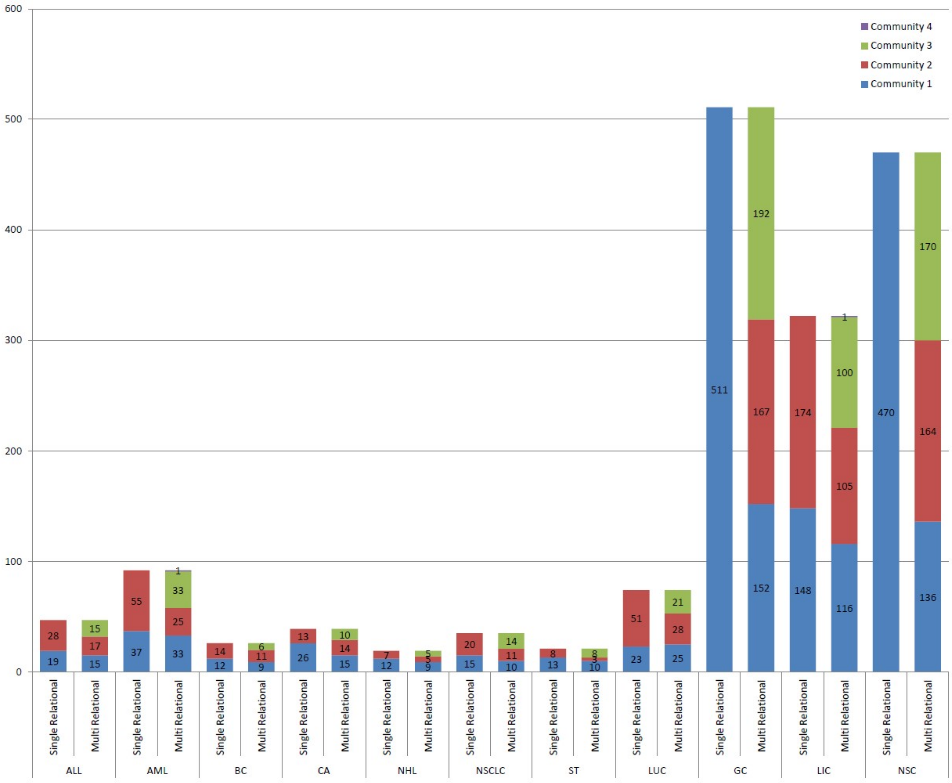


Figure 8.1: Number of communities and of genes in each community obtained by the two different approaches: single-relational transformation of the network and the original multi-relational network.

network as can be seen in Figure 8.1. Moreover, the overlap ratio between any two communities is, very low for some networks, with no genes in common (as in the case of ALL, AML, NHL, ST and LIC networks) or just one gene in common (as in the case of AML, BC, CA and LIC networks). In some cases, some of the communities obtained in the multi-relational setting are included in the communities obtained in the single-relational setting. It is worth mentioning that the multi-relational tests always generate more communities than the single-relational ones.

8.4 Summary

The work presented here has the purpose of demonstrating that multi-relational networks and the computational operations applied to them have to preserve the multi-relational aspect and cannot be reduced to single-relational ones. The transformation of a multi-relational network into a single-relational one can significantly reduce the computational cost, but usually leads to loss of information. In this work we propose an algorithm for finding communities in multi-relational networks, originally developed for single-relational networks. The algorithm was extended to multi-relational networks and comparisons between the two algorithms were performed on 11 datasets taken from cancer data which represent gene-gene interaction graphs. The results show significant differences between the two approaches which would be worth considered by researchers working in biomedical field. The framework we propose is a general one, so any algorithm (not just Fruchtermans algorithm presented here) can be extended from single-relational graphs to multi-relational ones.

In this chapter propose a new community detection algorithm for multirelational networks (55). We apply it to multirelational biological networks and compare the results to those produced by a similar community detection algorithm for single relational networks. We show that our algorithm is able to detect more fine-grained communities compared to a similar single-relational algorithm, because information is lost when treating the initial multi-relational graph as a single-relational one. Results presented in this chapter are included in a paper submitted to PlosOne journal.

9

Conclusions

This last chapter briefly review the efforts of developing and analyzing multi-relational network models and multidimensional time series clustering, and identify some general directions for extending the models developed in this thesis that may yield additional insights. Several open problems in multi-relational networks analysis and multidimensional time series clustering are also described below.

This thesis has presented two main ideas which are used herein for addressing biological problems. However, they are not specifically constructed for biology or medicine and can be applied to other fields as well.

The first subject deals with multi-relational networks (networks where there exist multiple links and more than one type of edges between two nodes). It addresses two aspects of these networks:

- Multi-relational network comparison and
- Graphlets and motifs finding in multi-relational networks.

The challenges encountered in this part of the thesis were:

- Dealing with some NP-complete problems (such as network isomorphism)
- Finding efficient algorithms for graphlets and motifs

- Finding the right way to represent networks and their interactions (we have multiple networks, some of the link types might be missing in one or more of the networks, etc.)
- Implementing the right measures for comparing networks (different users have different preferences).

Still, much work remains to be done for a comprehensive comparison of networks and for implementing even more efficient algorithms for motifs finding. The achievements so far are:

- Design and implementation of a network comparison tool:
  - It takes as input two or more networks
  - It is designed for gene-gene interaction networks and protein-protein interaction networks, but it is easily extensible and adaptable
  - It implements some basic operations such as union, intersection, common sub-graphs, degree distribution
  - It allows selection of one, two, more or all types of existing links and performs operations for the selected types of links
- Design and implementation of a graphlets and motifs finding tool:
  - It is designed for multi-relational networks
  - It finds motifs of a certain given size (sizes of 3, 4 and 5 nodes are implemented)
  - It finds custom design motifs (a set of such motifs at a time)
  - Works for both directed and undirected graphs (networks)
- Experiments and applications to biological problems:
  - Cancer networks for female related cancer types (ovarian, endometrial and breast) have been performed

  - Comparison of gene-gene interaction networks based on the contained motifs (i.e. basic building blocks)

- Design and implementation of a multidimensional clustering tool:

  - Entities formed of multiple time series are clustered

  - A comparison of some geometrical distance based similarity is performed.

10

Keywords

- bioinformatics
- complex networks
- single-relational networks
- multi-relational networks
- multi-relational biological networks
- clusters
- multidimensional time series clustering
- graph motif
- motifs in multi-relational networks
- communities detection
- communities in multi-relational networks
- gene-gene interaction networks
- networks analysis
- software tools

11

Thesis references

[1] AARTS/KORST. Simulated annealing and boltzmann machines. A stochastic approach to combinatorial opti- mization and neural computing. John Wiley., 1990.

[2] Gaurav Agarwal and David Kempe. Modularitymaximizing graph communities via mathematical programming. The European Physical Journal B, 66(3):409-418, 2008.

[3] Edoardo M Airoldi, David M Blei, Stephen E Fienberg, and Eric P Xing. Mixed membership stochastic blockmodels. Journal of Machine Learning Research, 9(Sep):1981-2014, 2008.

[4] Reka Albert and Albert-Laszlo Barabasi. Statistical mechanics of complex networks. Reviews of modern physics, 74(1):47, 2002.

[5] Reka Albert, Hawoong Jeong, and Albert-Laszlo Barabasi. Error and attack tolerance of complex networks. arXiv preprint cond-mat/0008064, 2000.

[6] Charles J Alpert and So-Zen Yao. Spectral partitioning: the more eigenvectors, the better. In Proceedings of the 32nd annual ACM/IEEE Design Automation Conference, pages 195-200. ACM, 1995.

[7] Khaled Alsabti, Sanjay Ranka, and Vineet Singh. An eficient k-means clustering

algorithm. 1997.

[8] Michael R Anderberg. Cluster analysis for applications.monographs and textbooks on probability and mathematical statistics, 1973.

[9] E Michael Azoff. Neural network time series forecasting of financial markets. John Wiley and Sons, Inc., 1994.

[10] Gary D Bader, Doron Betel, and Christopher WV Hogue. Bind: the biomolecular interaction network database. Nucleic acids research, 31(1):248-250, 2003.

[11] Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5(2):101-113, 2004.

[12] Alain Barrat, Marc Barthelemy, Romualdo Pastor- Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. Proceedings of the Na- tional Academy of Sciences of the United States of America, 101(11):3747-3752, 2004.

[13] Federico Battiston, Jacopo Iacovacci, Vincenzo Nicosia, Ginestra Bianconi, and Vito Latora. Emergence of multiplex communities in collaboration networks. PloS one, 11(1):e0147451, 2016.

[14] Anais Baudot, Gonzalo Gomez-Lopez, and Alfonso Valencia. Translational disease interpretation with molecular networks. Genome biology, 10(6):221, 2009.

[15] Laura Bennett, Aristotelis Kittas, Gareth Muirhead, Lazaros G Papageorgiou, and Sophia Tsoka. Detection of composite communities in multiplex biological networks. Scientific reports, 5, 2015.

[16] Laura Bennett, Songsong Liu, Lazaros G Papageorgiou, and Sophia Tsoka. Detection of disjoint and overlapping modules in weighted complex networks. Advances in Complex Systems, 15(05):1150023, 2012.

[17] Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino

Pedreschi. Foundations of multidimensional network analysis. In Advances in So- cial Networks Analysis and Mining (ASONAM), 2011 In- ternational Conference on, pages 485-489. IEEE, 2011.

[18] Bela Bollobas, Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. Time-series similarity problems and well-separated geometric sets. In Proceedings of the thirteenth annual symposium on Computational geometry, pages 454-456. ACM, 1997.

[19] David E Booth. Time series, 1992.

[20] Stefan Bornholdt and Heinz Georg Schuster. Handbook of graphs and networks: from the genome to the internet. John Wiley and Sons, 2006.

[21] Laurie A Boyer, Tong Ihn Lee, Megan F Cole, Sarah E Johnstone, Stuart S Levine, Jacob P Zucker, Matthew G Guenther, Roshan M Kumar, Heather L Murray, Richard G Jenner, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. cell, 122(6):947-956, 2005.

[22] Ulrik Brandes. A faster algorithm for betweenness centrality*. Journal of Mathematical Sociology, 25(2):163- 177, 2001.

[23] Rainer Breitling, David Gilbert, Monika Heiner, and Richard Orton. A structured approach for the engineering of biochemical network models, illustrated for signalling pathways. Briefings in Bioinformatics, 9(5):404-421, 2008.

[24] Karin Breuer, Amir K Foroushani, Matthew R Laird, Carol Chen, Anastasia Sribnaia, Raymond Lo, Geoffrey L Winsor, Robert EW Hancock, Fiona SL Brinkman, and David J Lynn. Innatedb: systems biology of innate immunity and beyondrecent updates and continuing curation. Nucleic acids research, 41(D1):D1228-D1233, 2013.

[25] Piotr Brodka, Tomasz Filipowski, and Przemys law Kazienko. An introduction to community detection in multi-layered social network. In World Summit on Knowledge Society, pages 185-190. Springer, 2011.

[26] Deng Cai, Zheng Shao, Xiaofei He, Xifeng Yan, and Jiawei Han. Community mining from multi-relational networks. In European Conference on Principles of Data Mining and Knowledge Discovery, pages 445-452. Springer, 2005.

[27] Muffy Calder, Stephen Gilmore, and Jane Hillston. Modelling the influence of rkip on the erk signalling pathway using the stochastic process algebra pepa. In Transactions on computational systems biology VII, pages 445-452. Springer, 2006.

[28] Melissa S Cline, Michael Smoot, Ethan Cerami, Allan Kuchinsky, Nerius Landys, Chris Workman, Rowan Christmas, Iliana Avila-Campilo, Michael Creech, Benjamin Gross, et al. Integration of biological networks and gene expression data using cytoscape. Nature pro- tocols, 2(10):2366-2382, 2007.

[29] Michele Coscia, Fosca Giannotti, and Dino Pedreschi. A classification for community discovery methods in complex networks. Statistical Analysis and Data Mining, 4(5):512-546, 2011.

[30] Nello Cristianini, John Shawe-Taylor, and Jaz S Kandola. Spectral kernel methods for clustering. In Advances in neural information processing systems, pages 649-655, 2002.

[31] Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. Finding similar time series. In European Symposium on Principles of Data Mining and Knowledge Discovery, pages 88-100. Springer, 1997.

[32] Darcy A Davis and Nitesh V Chawla. Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. PloS one, 6(7):e22670, 2011.

[33] Kara Dolinski, Andrew Chatr-aryamontri, and Mike Tyers. Systematic curation of protein and genetic interaction data for computable biology. BMC biology, 11(1):43, 2013.

[34] Benjamin S Duran and Patrick L Odell. Cluster analysis: a survey. Springer-Verlag

New York, 1974.

[35] Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. Nucleic acids research, 30(1):207-210, 2002.

[36] William H Elliott, Daphne C Elliott, John R Jefferson, and John Wheldrake. Biochemistry and molecular biology. Oxford University Press Oxford, 1997.

[37] Graphical Enumeration. F. harary and em palmer, 1973.

[38] Paul Erdos and Alfred Renyi. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci, 5(1):17- 60, 1960.

[39] Brian S Everitt. Unresolved problems in cluster analysis. Biometrics, pages 169-181, 1979.

[40] http://cancer.sanger.ac.uk/census/.

[41] Santo Fortunato. Community detection in graphs. Physics reports, 486(3):75-174, 2010.

[42] Andrea Franceschini, Damian Szklarczyk, Sune Frankild, Michael Kuhn, Milan Simonovic, Alexander Roth, Jianyi Lin, Pablo Minguez, Peer Bork, Christian von Mering, et al. String v9. 1: protein-protein interaction networks, with increased coverage and integration. Nucleic acids research, 41(D1):D808-D815, 2013.

[43] Marco Franciosi and Giulia Menconi. Multi-dimensional sparse time series: feature extraction. arXiv preprint arXiv:0803.0405, 2008.

[44] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. Software: Practice and experience, 21(11):1129-1164, 1991.

[45] Sylvia Fruhwirth-Schnatter, Christoph Pamminger, Rudolf Winter-Ebmer, and Andrea Weber. Model-based clustering of categorical time series with multinomial logit

classification. In AIP Conference Proceedings, volume 1281, pages 1897-1900. AIP, 2010.

[46] Guojun Gan, Chaoqun Ma, and JianhongWu. Data clus- tering: theory, algorithms, and applications, volume 20. Siam, 2007.

[47] Michael R Garey and David S Johnson. Crossing number is np-complete. SIAM Journal on Algebraic Discrete Methods, 4(3):312-316, 1983.

[48] Claude Gerard, Didier Gonze, and Albert Goldbeter. Effect of positive feedback loops on the robustness of oscillations in the network of cyclin-dependent kinases driving the mammalian cell cycle. FEBS Journal, 279(18):3411-3431, 2012.

[49] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12):7821-7826, 2002.

[50] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. Nature, 433(7028):895-900, 2005.

[51] Roger Guimera, Marta Sales-Pardo, and Luis A Nunes Amaral. A network-based method for target selection in metabolic networks. Bioinformatics, 23(13):1616-1622, 2007.

[52] Dimitrios Gunopulos and Gautam Das. Time series similarity measures (tutorial pm-2). In Tutorial notes of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 243-307. ACM, 2000.

[53] Jing-Dong J Han, Nicolas Bertin, Hao Tong, Debra S Goldberg, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature, 430(6995):88, 2004.

[54] Frank Harary and Edgar M Palmer. Graphical enumeration. Technical report,

DTIC Document, 1973.

[55] John A Hartigan. Clustering algorithms. John Wiley and Sons, Inc., 1975.

[56] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1):100-108, 1979.

[57] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. Social networks, 5(2):109-137, 1983.

[58] William M Holmes. Time series: Sir maurice kendall and j. keith ord, (edward arnold, great britain, 1990) pp. 296, 1992.

[59] Thomas C Ings, Jose M Montoya, Jordi Bascompte, Nico Bluthgen, Lee Brown, Carsten F Dormann, Francois Edwards, David Figueroa, Ute Jacob, J Iwan Jones, et al. Review: Ecological networks-beyond food webs. Journal of Animal Ecology, 78(1):253-269, 2009.

[60] Kuhn Ip, Caroline Colijn, and Desmond S Lun. Analysis of complex metabolic behavior through pathway decomposition. BMC systems biology, 5(1):91, 2011.

[61] Negin Iranfar, Danny Fuller, and William F Loomis. Transcriptional regulation of post-aggregation genes in dictyostelium by a feed-forward loop involving gbf and lagc. Developmental biology, 290(2):460-469, 2006.

[62] Shalev Itzkovitz, Reuven Levitt, Nadav Kashtan, Ron Milo, Michael Itzkovitz, and Uri Alon. Coarse-graining and self-dissimilarity of complex networks. Physical Re- view E, 71(1):016127, 2005.

[63] Ariel Jaimovich, Gal Elidan, Hanah Margalit, and Nir Friedman. Towards an integrated protein-protein interaction network: A relational markov network approach. Journal of Computational Biology, 13(2):145-164, 2006.

[64] Anil K Jain and Richard C Dubes. Algorithms for clus- tering data. Prentice-Hall, Inc., 1988.

[65] E Ferrell James Jr. Feedback loops and reciprocal regulation: recurring motifs in the systems biology of the cell cycle. Current Opinion in Cell Biology, 25(6):676- 686, 2013.

[66] Hawoong Jeong, Sean P Mason, Albert-Laszlo Barabasi, and Zoltan N Oltvai. Lethality and centrality in protein networks. arXiv preprint cond-mat/0105306, 2001.

[67] Hawoong Jeong, Balint Tombor, Reka Albert, Zoltan N Oltvai, and A-L Barabasi. The large-scale organization of metabolic networks. Nature, 407(6804):651-654, 2000.

[68] Stephen C Johnson. Hierarchical clustering schemes. Psychometrika, 32(3):241-254, 1967.

[69] Bjorn H Junker and Falk Schreiber. Analysis of biological networks, volume 2. John Wiley and Sons, 2008.

[70] Minoru Kanehisa. A database for post-genome analysis. Trends in genetics: TIG, 13(9):375, 1997.

[71] Minoru Kanehisa. The kegg database. In Novartis Found Symp, volume 247, pages 91-101, 2002.

[72] Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Yasushi Okuno, and Masahiro Hattori. The kegg resource for deciphering the genome. Nucleic acids re- search, 32(suppl 1):D277-D280, 2004. 23

[73] Ravi Kannan, Santosh Vempala, and Adrian Vetta. On clusterings: Good, bad and spectral. Journal of the ACM (JACM), 51(3):497-515, 2004.

[74] Tapas Kanungo, David M Mount, Nathan S Netanyahu, Christine D Piatko, Ruth Silverman, and Angela Y Wu. An eficient k-means clustering algorithm: Analysis

and implementation. IEEE transactions on pattern analysis and machine intelligence, 24(7):881-892, 2002.

[75] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Mfinder tool guide. Department of Molecular Cell Biology and Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot Israel, Tech. Rep, 2002.

[76] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Eficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. Bioinformatics, 20(11):1746-1758, 2004.

[77] Leonard Kaufman and Peter J Rousseeuw. Finding groups in data: an introduction to cluster analysis, volume 344. John Wiley and Sons, 2009.

[78] Jinki Kim and Gwan-Su Yi. Rmod: A tool for regulatory motif detection in signaling network. PloS one, 8(7):e68407, 2013.

[79] Mikko Kivela, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. Journal of complex networks, 2(3):203-271, 2014.

[80] Mark Kozdoba and Shie Mannor. Community detection via measure space embedding. In Advances in Neural Information Processing Systems, pages 2890-2898, 2015.

[81] Cho Kwang-Hyun, Shin Sung-Young, Kim Hyun-Woo, Olaf Wolkenhauer, Brian McFerran, and Walter Kolch. Mathematical modeling of the influence of rkip on the erk signaling pathway. In Computational methods in sys- tems biology, pages 127-141. Springer, 2003.

[82] Godfrey N Lance and William Thomas Williams. A general theory of classi catory sorting strategies 1. hierarchical systems. The computer journal, 9(4):373-380, 1967.

[83] Tong Ihn Lee, Nicola J Rinaldi, Francois Robert, Duncan T Odom, Ziv Bar-Joseph,

Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, et al. Transcriptional regulatory networks in saccharomyces cerevisiae. science, 298(5594):799-804, 2002.

[84] Xutao Li, Michael K Ng, and Yunming Ye. Multicomm: Finding community structure in multi-dimensional networks. IEEE Transactions on Knowledge and Data Engineering, 26(4):929-941, 2014.

[85] T Warren Liao. Clustering of time series dataa survey. Pattern recognition, 38(11):1857-1874, 2005.

[86] Luana Licata, Leonardo Briganti, Daniele Peluso, Livia Perfetto, Marta Iannuccelli, Eugenia Galeota, Francesca Sacco, Anita Palma, Aurelio Pio Nardozza, Elena Santonico, et al. Mint, the molecular interaction database: 2012 update. Nucleic acids research, 40(D1):D857-D861, 2012.

[87] M. Lichman. UCI machine learning repository, 2013.

[88] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. Pattern recognition, 36(2):451-461, 2003.

[89] Azi Lipshtat, Sudarshan P Purushothaman, Ravi Iyengar, and Avi Maayan. Functions of bifans in context of multiple regulatory motifs in signaling networks. Biophysical journal, 94(7):2566-2579, 2008.

[90] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, pages 281-297. Oakland, CA, USA., 1967.

[91] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. Proceedings of the National Academy of Sciences, 100(21):11980-11985, 2003.

[92] Sergei Maslov, Kim Sneppen, and Uri Alon. Correlation profiles and motifs in complex networks. Handbook of Graphs and Networks: From the Genome to the Internet, pages 168-198, 2003.

[93] Louis L McQuitty. Elementary linkage analysis for isolating orthogonal and oblique types and typal relevancies. Educational and Psychological Measurement, 1957.

[94] Marina Meila and Jianbo Shi. Learning segmentation by random walks. In NIPS, volume 14, 2000.

[95] Manuel Middendorf, Etay Ziv, and Chris H Wiggins. Inferring network mechanisms: the drosophila melanogaster protein interaction network. Proceedings of the National Academy of Sciences of the United States of America, 102(9):3192-3197, 2005.

[96] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. Science, 303(5663):1538-1542, 2004.

[97] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. Science, 298(5594):824-827, 2002.

[98] Alexander Y Mitrophanov and Eduardo A Groisman. Positive feedback in cellular control systems. Bioessays, 30(6):542-555, 2008.

[99] Sara Mostafavi, Debajyoti Ray, David Warde-Farley, Chris Grouios, Quaid Morris, et al. Genemania: a realtime multiple association network integration algorithm for predicting gene function. Genome Biol, 9(Suppl 1):S4, 2008.

[100] Panagiotis Moulos, Julie Klein, Simon Jupp, Robert Stevens, Jean-Loup Bascands, and Joost P Schanstra. The kupnetviz: a biological network viewer for multipleomics datasets in kidney diseases. BMC bioinformatics, 14(1):235, 2013.

[101] KA Abdul Nazeer and MP Sebastian. Improving the accuracy and efficiency of the

k-means clustering algorithm. In Proceedings of the World Congress on Engineering, volume 1, pages 1-3, 2009. 109 [102] Mark EJ Newman. Models of the small world. Journal of Statistical Physics, 101(3):819-841, 2000.

[103] Mark EJ Newman. Clustering and preferential attachment in growing networks. Physical review E, 64(2):025102, 2001.

[104] Mark EJ Newman. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2):404-409, 2001.

[105] Mark EJ Newman. The structure and function of complex networks. SIAM review, 45(2):167-256, 2003.

[106] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. Physical review E, 69(2):026113, 2004.

[107] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In Ad- vances in neural information processing systems, pages 849-856, 2002 [108] Michaek Kwok-Po Ng, Xutao Li, and Yunming Ye. Multirank: co-ranking for objects and relations in multirelational data. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1217-1225. ACM, 2011.

[109] BO Palsson. Properties of reconstructed networks. Cambridge: Systems Biology, 2006.

[110] Tony Pawson and Rune Linding. Network medicine. FEBS letters, 582(8):1266-1270, 2008.

[111] Matteo Pellegrini, Edward M Marcotte, Michael J Thompson, David Eisenberg, and Todd O Yeates. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proceedings of the Na- tional Academy of Sciences, 96(8):4285-4288, 1999.

[112] Steven J Phillips. Acceleration of k-means and related clustering algorithms. In Algorithm Engineering and Experiments, pages 166-177. Springer, 2002.

[113] V Pihur, Somnath Datta, and Susmita Datta. Finding common genes in multiple cancer types through meta- analysis of microarray experiments: A rank aggregation approach. Genomics, 92(6):400-403, 2008.

[114] Ivan Plavec, Oksana Sirenko, Sylvie Privat, Yuker Wang, Maya Dajee, Jennifer Melrose, Brian Nakao, Evangelos Hytopoulos, Ellen L Berg, and Eugene C Butcher. Method for analyzing signaling networks in complex cellular systems. Proceedings of the National Academy of Sciences of the United States of America, 101(5):1223-1228, 2004.

[115] Teresa Przytycka. An important connection between network motifs and parsimony models. In Research in Computational Molecular Biology, pages 321-335. Springer, 2006.

[116] Marcos G Quiles, Elbert EN Macau, and Nicolas Rubido.Dynamical detection of network communities. Scientific reports, 6, 2016.

[117] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America, 101(9):2658-2663, 2004.

[118] Silvia Rausanu, Crina Grosan, Zujian Wu, Ovidiu Parvu, Ramona Stoica, and David Gilbert. Computational models for inferring biochemical networks. Neural Computing and Applications, 26(2):299-311, 2015.

[119] Erzsebet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltan N Oltvai, and A-L Barabasi. Hierarchical organization of modularity in metabolic networks. science, 297(5586):1551-1555, 2002.

[120] F James Rohlf. 12 single-link clustering algorithms. Handbook of Statistics, 2:267-284, 1982.

[121] Michal Ronen, Revital Rosenberg, Boris I Shraiman, and Uri Alon. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proceedings of the national academy of sciences, 99(16):10555-10560, 2002.

[122] Jill A Rosenfeld, Dina Amrom, Eva Andermann, Frederick Andermann, Martin Veilleux, Cynthia Curry, Jamie Fisher, Stephen Deputy, Arthur S Aylsworth, Cynthia M Powell, et al. Genotype-phenotype correlation in interstitial 6q deletions: a report of 12 new cases. neurogenetics, 13(1):31-47, 2012.

[123] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences, 105(4):1118-1123, 2008.

[124] Peter J Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20:53-65, 1987.

[125] Louis A Saddic, Barbel Huvermann, Staver Bezhani, Yanhui Su, Cara M Winter, Chang Seob Kwon, Richard P Collum, and Doris Wagner. The leafy target lmi1 is a meristem identity regulator and acts together with leafy to regulate expression of cauliflower. Development, 133(9):1673-1682, 2006.

[126] Rintaro Saito, Michael E Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, Samad Lotia, Alexander R Pico, Gary D Bader, and Trey Ideker. A travel guide to cytoscape plugins. Nature methods, 9(11):1069-1076, 2012.

[127] Regina Samaga and Steffen Klamt. Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. Cell communication and signaling, 11(1):43, 2013.

[128] S Schuster, DA Fell, and T Dandekar. A general definition of metabolic pathways

useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology, 18(3):326-332, 2000.

[129] http://ccgd-starrlab.oit.umn.edu/search.php.

[130] Jimmy Shadbolt and John Gerald Taylor. Neural Net- works and the Financial Markets: Predicting, Combining, and Portfolio Optimisation. Springer, 2002.

[131] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research, 13(11):2498-2504, 2003.

[132] Junming Shao, Zhichao Han, and Qinli Yang. Community detection via local dynamic interaction. arXiv preprint arXiv:1409.7978, 2014.

[133] Peter HA Sneath. The application of computers to taxonomy. Journal of general microbiology, 17(1):201-226, 1957.

[134] Daniel A Spielmat and Shang-Hua Teng. Spectral partitioning works: Planar graphs and

nite element meshes. In Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on, pages 96-105. IEEE, 1996.

[135] Jorg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. Nature, 420(6912):190-193, 2002.

[136] RAMONA STOICA. Multiobjective approach of multidimensional time series clustering. Studia Universitatis Babes-Bolyai, Informatica, 59(1), 2014.

[137] Ramona Stoica, Mihaela Ola, and Mihai Paraschivescu. Multidimensional temporal clustering: geometrical similarity measures analysis in k-means. Intelligent Systems

Design and Applications, 2013.

[138] Ramona Stoica, Bazil Parv, and Crina Grosan. Communities detection in multi-relational networks. PlosOne, 26(2):299-311, 2017.

[139] Ramona Stoica and Liviu Stirb. A tool for comparing multirelational networks from biology. In Intelligent Sys- tems Design and Applications (ISDA), 2013 13th International Conference on, pages 242-246. IEEE, 2013.

[140] Gemma Swiers, Roger Patient, and Matthew Loose. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Developmental biology, 294(2):525-540, 2006.

[141] Damian Szklarczyk, Andrea Franceschini, Michael Kuhn, Milan Simonovic, Alexander Roth, Pablo Minguez, Tobias Doerks, Manuel Stark, Jean Muller, Peer Bork, et al. The string database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research, 39(suppl 1):D561-D568, 2011.

[142] Bosiljka Tadic, Miroslav Andjelkovic, Biljana Mileva Boshkoska, and Zoran Levnajic. Algebraic topology of multi-brain connectivity networks reveals dissimilarity in functional patterns during spoken communications. PloS one, 11(11):e0166787, 2016.

[143] Bing Tian Dai, Freddy Chong Tat Chua, and Ee-Peng Lim. Structural analysis in multi-relational social networks.In Proceedings of the 2012 SIAM International Conference on Data Mining, pages 451-462. SIAM, 2012.

[144] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. Random walk with restart: fast solutions and applications. Knowledge and Information Systems, 14(3):327-346, 2008.

[145] Ruey S Tsay. Analysis of financial time series. Financial econometrics, a wiley-interscience publication, 2002.

[146] Joshua R Tyler, Dennis M Wilkinson, and Bernardo A Huberman. E-mail as

spectroscopy: Automated discovery of community structure within organizations. The Information Society, 21(2):143-153, 2005.

[147] Ikuo Uchiyama. Mbgd: microbial genome database for comparative analysis. Nucleic acids research, 31(1):58- 62, 2003.

[148] Olga Vechtomova. Introduction to information retrieval christopher d. manning, prabhakar raghavan, and hinrich schutze (stanford university, yahoo! research, and university of stuttgart) cambridge: Cambridge university press, 2008, xxi+ 482 pp; hardbound, isbn 978-0- 521-86571-5, 2009.

[149] David Warde-Farley, Sylva L Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz, Chris Grouios, Farzana Kazi, Christian Tannus Lopes, et al. The genemania prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research, 38(suppl 2):W214-W220, 2010.

[150] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. nature, 393(6684):440- 442, 1998.

[151] Yen-Chuen Wei and Chung-Kuan Cheng. Towards efficient hierarchical designs by ratio cut partitioning. In Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on, pages 298-301. IEEE, 1989.

[152] Sebastian Wernicke. Efficient detection of network motifs. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 3(4), 2006.

[153] Dennis M Wilkinson and Bernardo A Huberman. A method for Finding communities of related genes. pro- ceedings of the national Academy of sciences, 101(suppl 1):5241-5248, 2004.

[154] WT Williams and JM t Lambert. Multivariate methods in plant ecology: V. similarity analyses and information-analysis. The Journal of Ecology, pages 427-445, 1966.

[155] Elisabeth Wong, Brittany Baur, Saad Quader, and Chun-Hsi Huang. Biological network motif detection: principles and practice. Briefings in bioinformatics, 13(2):202-215, 2011.

[156] Elisabeth A Wong and Brittany Baur. On network tools for network motif
nding: a survey study, 2010.

[157] Zhiang Wu, Zhan Bu, Jie Cao, and Yi Zhuang. Discovering communities in multi-relational networks. In User Community Discovery, pages 75-95. Springer, 2015.

[158] Stefan Wuchty. Scale-free behavior in protein domain networks. Molecular biology and evolution, 18(9):1694-1702, 2001.

[159] Ioannis Xenarios, Danny W Rice, Lukasz Salwinski, Marisa K Baron, Edward M Marcotte, and David Eisenberg. Dip: the database of interacting proteins. Nucleic acids research, 28(1):289-291, 2000.

[160] Gang Xu, Laura Bennett, Lazaros G Papageorgiou, and Sophia Tsoka. Module detection in complex networks using integer optimisation. Algorithms for Molecular Biology, 5(1):36, 2010.

[161] Yoshihiro Yamanishi, Michihiro Araki, Alex Gutteridge, Wataru Honda, and Minoru Kanehisa. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics, 24(13):i232-i240, 2008.

[162] Muhammed A Yildirim, Kwang-Il Goh, Michael E Cusick, Albert-Laszlo Barabasi, and Marc Vidal. Drug- target network. Nature biotechnology, 25(10):1119, 2007.

[163] Chang Hun You, Lawrence B Holder, and Diane J Cook. Application of graph-based data mining to metabolic pathways. In Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on, pages 169-173. IEEE, 2006.

[164] Alon Zaslaver, Avi E Mayo, Revital Rosenberg, Pnina Bashkin, Hila Sberro,

Miri Tsalyuk, Michael G Surette, and Uri Alon. Just-in-time transcription program in metabolic pathways. Nature genetics, 36(5):486, 2004.

[165] Bin Zhang and Sargur N Srihari. Properties of binary vector dissimilarity measures. In Proc. JCIS Intl Conf. Computer Vision, Pattern Recognition, and Image Processing, volume 1, 2003.

[166] Bin Zhang and Sargur N Srihari. Fast k-nearest neighbor classification using cluster-based trees. IEEE Transactions on Pattern analysis and machine intelligence, 26(4):525-528, 2004.

[167] Yan Zhang, Zhifeng Yang, and Xiangyi Yu. Ecological network and emergy analysis of urban metabolic systems: model development, and a case study of four chinese cities. Ecological Modelling, 220(11):1431-1442, 2009.

[168] Chunxiang Zheng, Chad R Weisbrod, Juan D Chavez, Jimmy K Eng, Vagisha Sharma, Xia Wu, and James E Bruce. Xlink-db: Database and software tools for storing and visualizing protein interaction topology data. Journal of proteome research, 12(4):1989-1995, 2013.

[169] Xiaowei Zhu, Mark Gerstein, and Michael Snyder. Getting connected: analysis and principles of biological networks. Genes and development, 21(9):1010-1024, 2007.

References

[1] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002. 11

[2] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *arXiv preprint cond-mat/0008064*, 2000. 11

[3] E Michael Azoff. *Neural network time series forecasting of financial markets*. John Wiley & Sons, Inc., 1994. 21

[4] Albert-László Barabási and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004. 30

[5] Anais Baudot, Gonzalo Gomez-Lopez, and Alfonso Valencia. Translational disease interpretation with molecular networks. *Genome biology*, 10(6):221, 2009. 30

[6] Béla Bollobás, Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. Time-series similarity problems and well-separated geometric sets. In *Proceedings of the thirteenth annual symposium on Computational geometry*, pages 454–456. ACM, 1997. 21

[7] David E Booth. Time series, 1992. 21

[8] Stefan Bornholdt and Heinz Georg Schuster. *Handbook of graphs and networks: from the genome to the internet*. John Wiley & Sons, 2006. 28

[9] Laurie A Boyer, Tong Ihn Lee, Megan F Cole, Sarah E Johnstone, Stuart S Levine, Jacob P Zucker, Matthew G Guenther, Roshan M Kumar, Heather L Murray, Richard G Jenner, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *cell*, 122(6):947–956, 2005. 34

[10] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, 25(2):163–177, 2001. 12

[11] Melissa S Cline, Michael Smoot, Ethan Cerami, Allan Kuchinsky, Nerius Landys, Chris Workman, Rowan Christmas, Iliana Avila-Campilo, Michael Creech, Benjamin Gross, et al. Integration of biological networks and gene expression data using cytoscape. *Nature protocols*, 2(10):2366–2382, 2007. 30, 31

[12] Michele Coscia, Fosca Giannotti, and Dino Pedreschi. A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining*, 4(5):512–546, 2011. 41

[13] Gautam Das, Dimitrios Gunopulos, and Heikki Mannila. Finding similar time series. In *European Symposium on Principles of Data Mining and Knowledge Discovery*, pages 88–100. Springer, 1997. 21

[14] Ron Edgar, Michael Domrachev, and Alex E Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic acids research*, 30(1):207–210, 2002. 36

[15] Brian S Everitt. Unresolved problems in cluster analysis. *Biometrics*, pages 169–181, 1979. 20

[16] `http://cancer.sanger.ac.uk/census/`. 41

[17] Marco Franciosi and Giulia Menconi. Multi-dimensional sparse time series: feature extraction. *arXiv preprint arXiv:0803.0405*, 2008. 23

[18] Sylvia Frühwirth-Schnatter, Christoph Pamminger, Rudolf Winter-Ebmer, and Andrea Weber. Model-based clustering of categorical time series with multinomial logit classification. In *AIP Conference Proceedings*, volume 1281, pages 1897–1900. AIP, 2010. 26

[19] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*, volume 20. Siam, 2007. 21

[20] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005. 14

[21] Dimitrios Gunopulos and Gautam Das. Time series similarity measures (tutorial pm-2). In *Tutorial notes of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 243–307. ACM, 2000. 21

[22] Jing-Dong J Han, Nicolas Bertin, Hao Tong, Debra S Goldberg, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430(6995):88, 2004. 34

[23] William M Holmes. Time series: Sir maurice kendall and j. keith ord, (edward arnold, great britain, 1990) pp. 296, 1992. 21

[24] Kuhn Ip, Caroline Colijn, and Desmond S Lun. Analysis of complex metabolic behavior through pathway decomposition. *BMC systems biology*, 5(1):91, 2011. 14

[25] Negin Iranfar, Danny Fuller, and William F Loomis. Transcriptional regulation of post-aggregation genes in dictyostelium by a feed-forward loop involving gbf and lagc. *Developmental biology*, 290(2):460–469, 2006. 34

[26] Ariel Jaimovich, Gal Elidan, Hanah Margalit, and Nir Friedman. Towards an integrated protein–protein interaction network: A relational markov network approach. *Journal of Computational Biology*, 13(2):145–164, 2006. 34

[27] Hawoong Jeong, Sean P Mason, Albert-Laszlo Barabasi, and Zoltan N Oltvai. Lethality and centrality in protein networks. *arXiv preprint cond-mat/0105306*, 2001. 34

[28] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000. 14, 34

[29] Minoru Kanehisa. The kegg database. In *Novartis Found Symp*, volume 247, pages 91–101, 2002. 17, 18, 19

[30] Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Yasushi Okuno, and Masahiro Hattori. The kegg resource for deciphering the genome. *Nucleic acids research*, 32(suppl 1):D277–D280, 2004. 17

[31] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Mfinder tool guide. *Department of Molecular Cell Biology and Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot Israel, Tech. Rep*, 2002. 35

[32] Jinki Kim and Gwan-Su Yi. Rmod: A tool for regulatory motif detection in signaling network. *PloS one*, 8(7):e68407, 2013. 35

[33] Tong Ihn Lee, Nicola J Rinaldi, François Robert, Duncan T Odom, Ziv Bar-Joseph, Georg K Gerber, Nancy M Hannett, Christopher T Harbison, Craig M Thompson, Itamar Simon, et al. Transcriptional regulatory networks in saccharomyces cerevisiae. *science*, 298(5594):799–804, 2002. 34

[34] T Warren Liao. Clustering of time series dataa survey. *Pattern recognition*, 38(11):1857–1874, 2005. 21

[35] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003. 34

[36] Manuel Middendorf, Etay Ziv, and Chris H Wiggins. Inferring network mechanisms: the drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3192–3197, 2005. 34

[37] Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, and Uri Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–1542, 2004. 34

[38] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002. 34

[39] Sara Mostafavi, Debajyoti Ray, David Warde-Farley, Chris Grouios, Quaid Morris, et al. Genemania: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*, 9(Suppl 1):S4, 2008. 18

[40] Mark EJ Newman. Models of the small world. *Journal of Statistical Physics*, 101(3):819–841, 2000. 11

[41] V Pihur, Somnath Datta, and Susmita Datta. Finding common genes in multiple cancer types through meta–analysis of microarray experiments: A rank aggregation approach. *Genomics*, 92(6):400–403, 2008. 29, 30

[42] Teresa Przytycka. An important connection between network motifs and parsimony models. In *Research in Computational Molecular Biology*, pages 321–335. Springer, 2006. 34

[43] Erzsébet Ravasz, Anna Lisa Somera, Dale A Mongru, Zoltán N Oltvai, and A-L Barabási. Hierarchical organization of modularity in metabolic networks. *science*, 297(5586):1551–1555, 2002. 15

[44] Michal Ronen, Revital Rosenberg, Boris I Shraiman, and Uri Alon. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceedings of the national academy of sciences*, 99(16):10555–10560, 2002. 34

[45] Jill A Rosenfeld, Dina Amrom, Eva Andermann, Frederick Andermann, Martin Veilleux, Cynthia Curry, Jamie Fisher, Stephen Deputy, Arthur S Aylsworth, Cynthia M Powell, et al. Genotype–phenotype correlation in interstitial 6q deletions: a report of 12 new cases. *neurogenetics*, 13(1):31–47, 2012. 18

[46] Louis A Saddic, Bärbel Huvermann, Staver Bezhani, Yanhui Su, Cara M Winter, Chang Seob Kwon, Richard P Collum, and Doris Wagner. The leafy target lmi1 is a meristem identity regulator and acts together with leafy to regulate expression of cauliflower. *Development*, 133(9):1673–1682, 2006. 34

[47] Rintaro Saito, Michael E Smoot, Keiichiro Ono, Johannes Ruscheinski, Peng-Liang Wang, Samad Lotia, Alexander R Pico, Gary D Bader, and Trey Ideker. A travel guide to cytoscape plugins. *Nature methods*, 9(11):1069–1076, 2012. 30, 31

[48] S Schuster, DA Fell, and T Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332, 2000. 14

[49] `http://ccgd-starrlab.oit.umn.edu/search.php`. 41

[50] Jimmy Shadbolt and John Gerald Taylor. *Neural Networks and the Financial Markets: Predicting, Combining, and Portfolio Optimisation*. Springer, 2002. 21

[51] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003. 30, 31

[52] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002. 14

[53] RAMONA STOICA. Multiobjective approach of multidimensional time series clustering. *Studia Universitatis Babes-Bolyai, Informatica*, 59(1), 2014. 27

[54] Ramona Stoica, Mihaela Ola, and Mihai Paraschivescu. Multidimensional temporal clustering: geometrical similarity measures analysis in k-means. *Intelligent Systems Design and Applications*, 2013. 27

[55] Ramona Stoica, Bazil Parv, and Crina Grosan. Communities detection in multi-relational networks. *PlosOne*, 26(2):299–311, 2017. 10, 44

[56] Ramona Stoica and Liviu Stirb. A tool for comparing multirelational networks from biology. In *Intelligent Systems Design and Applications (ISDA), 2013 13th International Conference on*, pages 242–246. IEEE, 2013. 10, 32

[57] Gemma Swiers, Roger Patient, and Matthew Loose. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Developmental biology*, 294(2):525–540, 2006. 34

[58] Ruey S Tsay. Analysis of financial time series. financial econometrics, a wiley-interscience publication, 2002. 21

[59] Ikuo Uchiyama. Mbgd: microbial genome database for comparative analysis. *Nucleic acids research*, 31(1):58–62, 2003. 17

[60] David Warde-Farley, Sylva L Donaldson, Ovi Comes, Khalid Zuberi, Rashad Badrawi, Pauline Chao, Max Franz, Chris Grouios, Farzana Kazi, Christian Tannus Lopes, et al. The genemania prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research*, 38(suppl_2):W214–W220, 2010. 18, 41

[61] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998. 13

[62] Elisabeth A Wong and Brittany Baur. On network tools for network motif finding: a survey study, 2010. 35

[63] Alon Zaslaver, Avi E Mayo, Revital Rosenberg, Pnina Bashkin, Hila Sberro, Miri Tsalyuk, Michael G Surette, and Uri Alon. Just-in-time transcription program in metabolic pathways. *Nature genetics*, 36(5):486, 2004. 34

[64] Bin Zhang and Sargur N Srihari. Properties of binary vector dissimilarity measures. In *Proc. JCIS Intl Conf. Computer Vision, Pattern Recognition, and Image Processing*, volume 1, 2003. 21