Analyzing the Social Networks using Blockmodeling Technique

Thesis submitted in partial fulfillment of the requirements for the award of degree of

Master of Engineering
in
Computer Science & Engineering

By:
Pawan Kumar
(800832009)

Under the supervision of:
Mrs. Shalini Batra
Assistant Professor, CSED

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR UNIVERSITY
PATIALA – 147004

JUNE - 2010

Certificate

I hereby certify that the work which is being presented in the thesis entitled, **"Analyzing the Social Networks using Blockmodeling Technique"**, in partial fulfillment of the requirements for the award of degree of Master of Engineering in Computer Science and Engineering submitted in Computer Science and Engineering Department of Thapar University, Patiala, is an authentic record of my own work carried out under the supervision of Shalini Batra and refers other researcher's works which are duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

Pawan K

(Pawan Kumar)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

(Mrs. SHALINI BATRA)
Assistant Professor,
Computer Science and Engineering Department,
Thapar University, Patiala.

Countersigned by

09/07/2010

(Dr. RAJESH BHATIA)
Professor & Head,
Computer Science & Engineering Department,
Thapar University,
Patiala.

13/7/10

(Dr. R.K.SHARMA)
Dean (Academic Affairs),
Thapar University,
Patiala.

i

Acknowledgement

No volume of words is enough to express my gratitude towards my guide **Mrs. Shalini Batra**, Department of Computer Science & Engineering, Thapar University, Patiala, who has been very concerned and has aided for all the materials essential for the preparation of this thesis report. He has helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research-oriented venture.

I am also thankful to **Dr. Rajesh Bhatia,** Head of Department, Computer Science & Engineering Department and **Dr. Inderveer Channa**, P.G. Coordinator, for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues who were always there at the need of the hour and provided with all the help and facilities, which I required, for the completion of my thesis work.

Most importantly, I would like to thank my parents and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope.

Pawan K

Pawan Kumar

(800832009)

Abstract

With the advent of the Internet, social networks have grown enormously and Social Network Analysis (SNA) has come up as an important field for research. Social networks are represented as graphs where each node called an actor or vertex in a graph represents an individual person or a group of persons and the fundamental component of SNA is the relationship defined by these linkages among units or nodes in the network. In SNA, statistical analysis of relational data is derived using various social network modeling techniques. One of the well known technique blockmodeling groups vertices into clusters and determine the relations between these clusters using matrices as computational tools. It is grounded on different structural concepts like equivalence and positions which are related to the theoretical concepts of social role and role sets.

In this thesis, the intent is to generate social networks of varying size using various network tools and analyze the relationship between participants using blockmodeling techniques. The data, generated in binary form, has been analyzed and visualized with varying cluster sizes.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

Understanding the nature of relationships and connections between entities is a key towards understanding a variety of phenomena throughout multiple disciplines. Network analysis is becoming increasingly popular as a general methodology for understanding complex patterns of interaction. Social Network Analysis (SNA) [1] is the study of relations between individuals including the analysis of social structures, social position, role analysis, etc. SNA is based on an assumption of the importance of relationships among interacting units or nodes. These relations defined by linkages among units or nodes are a fundamental component of SNA [2].

1.1 Introduction to Network

Networks found everywhere can range from small sociograms such as those introduced by Moreno [3] with only a few units to huge networks with practically billions of units for example a network of all computers connected to the Internet or a network of all people and their relationship or attributes. Units can be people, organizations, countries, words *etc.* and for each of these units there are a number of possible ties between pairs of these units. Commonly available networks are networks among people (friendship or communication), trade networks among organizations, countries, citation networks, genealogies, organic molecules in chemistry, ties among words in text, transportation networks *etc.*

1.2 Introduction to Social Network

Social network is a map of all of the relevant ties between all the nodes being studied. These concepts are often displayed in a social network diagram, where nodes are the points and ties are the lines.

Social networks have been defined by graphs representing social relationships between people or organizations. Each node also called an actor or vertex in a graph represents an individual person or a group of persons. An edge connecting two nodes called a tie represents relationship between the objects represented by these two nodes. Using graphs to represent social data enables social analysts to completely and rigorously describe and analyze the structural information embedded in social relationships. In a general social networks can be used to represent, identify, and measure any type of correlations between any kind of entities such as words, web pages, people, organizations, animals, cells, computers, attribute and other information or knowledge processing entities [4].

Social network analysis focuses on relationships between social entities. It offers the methodology to analyze social relations and it tells us how to conceptualize social networks. Social network analysts assume that interpersonal ties matter as do ties among organizations or countries, because they transmit behavior, attitudes, information, or goods. These networks can also be used to measure social capital the value that an individual gets from the social network. It is used widely in the social and behavioral sciences, as well as in political science, economics, organizational science, and industrial engineering.

Social network analysts argue that causation is not located in the individual but in the social structure. While people with similar attributes may behave similarly, explaining these similarities by pointing to common attributes misses the reality that individuals with common attributes often occupy similar positions in the social structure [5]. Their similar outcomes are caused by the constraints, opportunities, and perceptions created by these similar network positions.

1.2.1 Social Network Modeling

Social network modeling focuses on social behavior the model takes into account relationships derived from statistical analysis of relational data. This type of modeling facilitates an awareness and understanding of the connections among people, whether they are political leaders, specific groups, and/or cliques in organizations. It also allows for an explanation of a flow of information or trafficking of small arms, or the identification of outliers or individuals who are isolated from the group. Grouping patterns (algorithms) allow for the separating of large networks into smaller subsets. The members who share identifying marks or attributes are drawn closer. Perhaps they may be all from the same religious sect, are of the same age, or have the same political ideology. Integral to social network modeling is analysis or disciplined inquiry into the patterns of relationships that develop and exist among the members in the social system. The relationships among members at different levels of the analysis are also included. Because this type of modeling relies on actors who are concrete and observable, the relationships within the social network are usually social or cultural. These types of relationships bind the actors or entities together, making them interdependent entities.

1.2.2 Examples of Social Network

The example given below in Table 1.1 is a simple graph used to understand basic concepts in social networks. This graph is a network of friendships between students in a small class. Assume that data about friendship among students is available as in the Table 1.1. The diagonal elements are all blanks and other elements in the table are binary, considering the fact that in friendship analysis there is no need to consider if a person is a friend of himself/herself. Hence, data in the table represents whether two persons are friends or not.

The network obtained by the Table1.1 is shown in Figure 1.1. In this network each node represents a student in the class. Two nodes are connected with an edge if they are friends. In this example two students are friends if the value of the corresponding element in the Table 1.1 is 1 i.e. a 0 value indicates that two nodes are not friends and hence not directly connected.

	John	Susan	Tom	Jack	Alice	Jeff	Mike	Tiger	Jane
John	--	0	1	1	0	0	0	0	0
Susan	0	--	0	0	0	0	0	0	0
Tom	1	1	--	1	0	0	0	0	0
Jack	1	0	1	--	0	1	1	1	0
Alice	0	0	0	0	--	1	0	0	0
Jeff	0	0	0	1	1	--	1	1	0
Mike	0	0	0	1	0	1	--	1	0
Tiger	0	0	0	1	0	1	1	--	0
Jane	0	0	0	0	0	0	0	0	--

Table 1.1: The Social Data about Friend about Friendship between Students in Small Class [6]

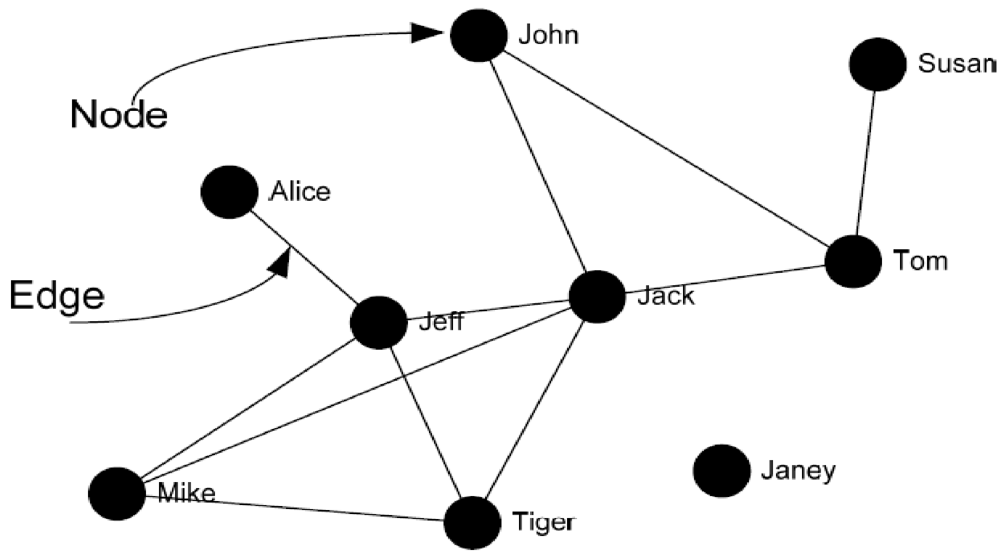


Figure 1.1: Friendship Network of a Small Class [6]

1.2.3 Social Network Data Collection

As discussed in previous section, the objective of the data collection is to provide a data set that could help analyze the effects of Virtual Social Networks in the different aspects of social activities such as like their generation, their spatial distribution, *etc.* [7]. However before building social networks the most important problem researchers have to face is how to acquire elementary data elements for building social networks. To get a complete and accurate description of interaction between individuals a lot of work has been done on social data gathering techniques focusing on how to identify the population, how to measure relationships *etc.* Currently there are mainly two kinds of approaches for social network data gathering: elicitation and registration [8]. Elicitation acquires interaction information via the questionnaire/survey. Registration acquires interactions through extracting from registered information such as membership lists email records, author records of scientific articles *etc.*

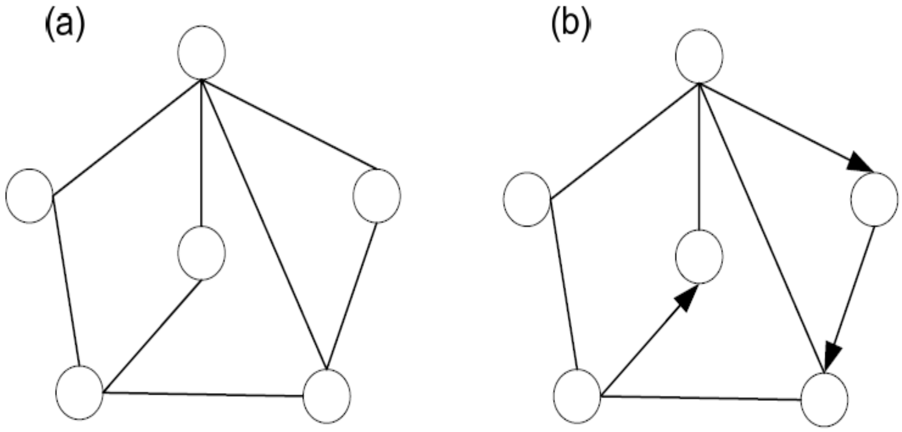
In the early SNA research, questionnaire/survey was the method primarily used for collecting data which had high-labor cost. It required social scientists and network researchers a myriad of efforts to gather data for a network of even middle size (several thousands of nodes). This intensive labor cost considerably limited the size of networks to be studied. Through fast developments of computer technologies and universal applications of computers, automated data acquisition are found in most, if not all, fields. Interactions between objects can be stored as or implicated by electronic data *for e.g.* co-authorship of research articles can represent the collaboration between research scientists. If two authors appear on the same paper, there will be a collaboration connection between them. Through rapid growth of network size and data-sharing techniques, huge databases of social interactions have emerged in various fields. For instance, there are many large databases that maintain records of article authors in publications of miscellaneous research fields [9]. For example, MEDLINE the database which covers published papers on biomedical research has about 2 million records from 1995 to 1999 [10].

1.2.4 Type of Social Network

As types of social networks are growing, the applications of SNA are also growing. Social networks can be classified based on the combination of attributes and measurements of nodes and ties. In social networks there could be different kinds of nodes or the same kind of nodes with various weights. For example affiliation networks [1] contain two kinds of nodes: events (such as corporations/organizations) and actors. Ties between events and actors usually represent relationships/participation [11].

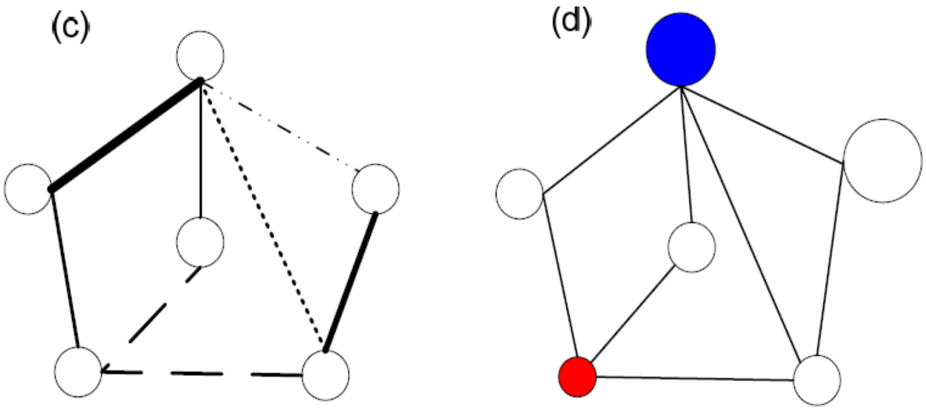
Similarly a social network could contain multiple types of ties or the same type of ties with different weights. A network with multiple relations are called multi-relational network. A multi-relational network may contain relationships as friendship collaboration and comember-ship. These relationships could have different importance or strength which is represented as the edge weight in a graph. Taking the example of friendship network people may use a number from 0 to 5 to indicate the strength of friendships between them and person ‘A’ who considers person ‘B’ as a friend does not necessary requires that ‘B’ consider ‘A’ as a friend too. Thus ties may have many directions. Symmetric ties are directed ties on both directions.

Another important classification is based on the number of distinct sets of units in the networks. This defines the mode of the network. If all units are from the same set the network is referred to as being one-mode, if there are two sets then it is two-mode and so on. In two-mode networks, all the units can not initiate ties *i.e.* all arcs usually lead from one set to another. More than two-mode networks are quite rare. In this thesis main focus is on the two mode data social networks. A real social network may be a combination of different networks, as shown in figure 1.2. The examples provided in this figure are simple and fundamental social networks.



(a) Social Network with Single Type of Nodes and Undirected Ties

(b) Social Network with Directed and Undirected Ties



(c) Social Network with Various Types and Weights of Ties

(d) Social Network whose Nodes have Different Types and Weights

Figure 1.2: Examples of Different Types of Social Networks [6]

1.3 Social Network Analysis

According to Wasserman & Faust [1] social network data can be viewed as a social relational system characterized by a set of actors and their social ties. Additional information in the form of actor attribute variables or multiple relations can be part of the social relational system.

Social network analysis aims at understanding the network structure by description visualization and statistical modeling. Social network data consist of various elements. In the analysis of complete networks a distinction can be made between:

- Descriptive methods, through graphical representations
- Analysis procedures often based on a of the adjacency matrix; and
- Statistical models based on probability distributions.

Visualization by displaying a sociogram as well as a summary of graph theoretical concepts provides a first description of social network data. Wasserman and Faust [1] stated that network analysis is based on the assumption of the importance of relationships (relations) among the interacting units. They also provided four principles that distinguish network analysis from similar approaches:

- Actors (units) and their actions (relations) are viewed as interdependent rather than independent and autonomous units.
- Relational ties (linkages) between actors (units) are channels for transfer or flow of resources (either material or nonmaterial).
- Network models focusing on individuals (individual units) view the network structural environment as providing opportunities for or constraints on individual action.
- Network models conceptualize structure (social, economic, political, and so forth) as lasting patterns of relations among the actors (units).

As discussed in Section 1.2.1 various techniques are available for modeling of social networks. One such model called Block Model is used in the analysis of social networks when the individual actors are combined into discrete subgroups. The subgroups are then linked with one or more link types, each expressing a different relationship. Blockmodeling is a flexible method for analyzing social networks. Blockmodeling technique is used to reduce a large potentially incoherent network to a smaller comprehensible structure that can be interpreted more readily. Blockmodeling aims to provide a slightly more generalized form of social network analysis based on groups rather than individual actors.

1.3.1 Social Network Analysis Applications

Social network analysis techniques can be applied to study structures of any types of interactions/relationships between any kinds of entities. From late 1970s, SNA techniques have gained massive attentions considerable developments, and successful applications in broad fields [1]. In companies and government agencies, there is lot of information sharing between workers. Using SNA tools on collaboration and/or information-sharing networks, managers can easily find the “important person” and build appropriate management strategies to improve efficiency.

Combating terrorism is another field where SNA techniques have important and successful applications. Terrorist organizations have special structures on recruitment, evolution and radical ideas diffusion [12]. SNA tools can be used to identify these unique organization structures and provide critical information for terrorist detection and terrorism prediction.

Social Network Analysis techniques also have been successfully applied in epidemiology. A lot of researchers try to analyze the spread of diseases based on the interactions between people.

A SNA researcher, Valdis Krebs, listed a number of recent successful applications of SNA in [4]. A selected set of applications include:

- Examining a network of farm animals to analyze how disease spreads from one cow to another
- Discovering emergent communities of interest amongst faculty at various universities
- Revealing cross-border knowledge flows based on research publications
- Determining influential journalists and analysts in the IT industry
- Unmasking the spread of HIV in a prison system
- Map executive's personal network based on email flows
- Discovering the network of Innovators in a regional economy

1.4 Structure of the Thesis

The rest of the thesis is organized in the following order:

Chapter 2 - Provides literature review of social network and techniques use for analyzing social networks.

Chapter 3 - Gives the problem statement and methodology used to solve it.

Chapter 4 - Gives a detailed introduction about blockmodeling technique used in social network analysis and tools used for applying blockmodeling.

Chapter5 – Provides the experiment performed using KEDIT Editor, UCINET, PAJEK, GSview and the results evaluated.

Chapter 6 - Gives the conclusion of thesis and suggestions for future work.

Thesis concludes with references.

Chapter 2
Literature Review

The scientific study of social networks has been ongoing for decades but in the past few years it has seen tremendous growth in its application and publicity. Social network analysis is an emerging area of research for computer scientists as it employs different concepts from graph theory, probability, and statistics to solve problems in a wide range of disciplines. Since social network analysis can be performed on many real world networks from different domains. A Number of social network analysis methods have been designed for various tasks.

Social network analysis goes back to Jacob L. Moreno and his psychodrama studies in the 1930’s [13]. He was among the first that operationalized the concept of social network and represented interpersonal relations in small groups using graphical methods in order to visualize channels of information. This visual device was called sociograph and the branch that describes and analyzes this kind of network configurations was denominated sociometry. By the end of the 1940's important advances were made in the research of the structural properties of networks.

As research in this area grew, network analysis was distinguished from traditional social science by the dyadic nature of the standard data set. These dyadic attributes (social relation) may be represented in matrix from by a square 1-mode matrix. But the data in traditional social science are represented as 2-mode matrices. However network analysis is not completely divorced from traditional social science and often has occasion to collect and analyze 2—mode matrices. Some of the methods developed in network analysis are used in analyzing non-network data. In [14] authors discussed ways of applying and interpreting traditional network analytic techniques to 2-mode data and further presented an idea on developing new techniques. In their paper three areas were covered in detail: displaying 2-mode data detecting network clusters and measuring centrality.

A structural technique called blockmodeling is applied to networks using matrices to represent the network data and permitting analyses of significantly more number of actors and various types of relations than the simple graph diagrams. All of this progress was possible due to advances in computer power that permitted to carry out complex analyses of relational data sets.

The two most widely used equivalences structural and regular equivalence are given special attention. Doreian *et al*. [22] introduced the concept of generalized equivalence. Generalized equivalence is defined by a set of allowed block types.

Patrick Doreian et al. in [15] examine blockmodeling procedures reviewing both structural equivalence and regular equivalence approaches. Realizing that few empirical examples of exact partitioning exist, it was argued that the lack of fit between model and reality can be measured and used as a way of comparing the adequacy of different models. This idea was combined this with a generalization of the blockmodeling method that permits many types of models to be constructed and compared. Sets of "permitted" ideal blocks are constructed and the model that shows minimum inconsistency is sought.

Batagelj Doreian and Ferligoj [16] developed a generalized approach to blockmodeling and methods where a set of observed relations are applied to a pre-specified blockmodel. This generalized blockmodeling approach was implemented in program Pajek.

Patrick Doreian & Vladimir Batagelj [17] extend the direct approach for block modeling one-mode data to two-mode data. The idea is that the rows and columns are partitioned simultaneously but in different ways. Many but not all of the generalized block types can be mobilized in block modeling two-mode network data. These methods were applied to some 'voting' data from the 2000–2001 terms of the Supreme Court and to the classic Deep South data on women attending events. The insight that rows and columns can be partitioned in different ways can also be applied to one-mode data.

There are many circumstances in which binary relations are defined between pairs of objects in sociology. For example, there are social relations between people in business, there are trading relations between firms, and in design there are functional dependencies between components. In all of these situations, the clustering of objects into densely interconnected blocks discovers the actual structure of the system. Alan Jessop [18] presented a method which permits the construction of blocks to be formulated as a quadratic programme. The method was applied to two illustrative cases: the pattern of elective choices by MBA students and the performance assessment of British universities.

Borgatti and Everett [19] stated that blockmodeling is a valuable technique in which redundant elements in an observed system are reduced to yield a simplified model of relationships among types of elements (units). The aim is to combine more or less identical units and thus reduce the complexity or size of the network. They understood positions as a set of actors that have the same ties with the same types of others and the relationships that these positions have as roles.

Everett and Borgatti [20] also presented a number of other equivalences and colorations and showed that most of them are special cases of regular equivalence or coloration. They further developed exact colorations especially in combination with perfect and ecological colorations in Everett and Borgatti. Several other equivalences have been proposed.

Winship and Mandel [21] show that blockmodeling is only capable of uncovering positions and not roles in a social network. In their paper they proposed an alternative (nonblockmodeling) approach to finding roles based on role sets. They actually defined position as a cluster of structurally equivalent units. On the other hand they associated roles with particular patterns of relations.

Edoardo Airoldi et al. [24] developed a model for examining data that consists of pair wise measurements for example presence or absence of links between pairs of objects. They introduced a class of latent variable models for pair wise measurements mixed membership stochastic blockmodels. Models in the class combined a global model of dense patches of connectivity (blockmodel) and a local model to instantiate node specific variability in the connections. They developed a general variational inference algorithm for fast approximate posterior inference.

Radoslaw Brendel & Henryk Krawczyk [25] considered social communities composed of several cohesive subgroups which they called compound communities. For such communities an extended generalized blockmodeling was proposed taking into account the structure of compound communities and relations with external actors. Using the extension, the community protection approach is proposed and used in detection of spam directed towards an e-mail local society.

After going through various research proposals in the area of analyzing social network, it was realized that blockmodeling can be used for analyzing complex data as it is based on the idea that units in a network can be grouped according to the extent to which they are equivalent under some meaningful definition of equivalence. In this thesis a comparative analysis of networks of varying sizes has been performed using direct approach of blockmodeling for various type two modes binary data to identify various clustering pattern discovered with different cluster sizes.

Chapter 3
Problem Statement

3.1 Problem Statement

With the growth of Internet, the size of social network is growing enormously. In social network all the relevant ties between all the nodes are often displayed in a social network diagram, where nodes are the points and ties are the lines and Social Network Analysis (SNA) focuses on relations between individuals and offers methodologies to analyze social relations existing between entities in a social network.

The grouping patterns (algorithms) available for analysis of such networks help in separation of large networks into smaller subsets, giving a clear picture of patterns of relationships that develop and exist among the members in the social system.

A technique called blockmodeling has been used in simplification of large and potentially incoherent networks in SNA. Blockmodeling is a framework for analyzing social structure and a set of procedures for the segmentation of a network. It is grounded on the different structural concepts like equivalence and positions which are related to the theoretical concepts of social role and role sets. It groups vertices into clusters and determine the relations between these clusters using matrices as computational tools. After going through the important applications of such networks in various fields including medical science, academia, research community, *etc*. we thought of analyzing and visualizing the relations between participants of social networks using blockmodeling on self designed data with varying cluster sizes and interpreting the output achieved.

3.2 Methodology

The step-by-step methodology to be followed in designing and analyzing of a social network is given below:

- Review of prevalent social networks and social network analysis techniques.
- Study of tools available for designing of Social network diagrams used for analyzing the network.

- Design of social networks of varying sizes.
- Analysis and interpretation of result achieved by using block modeling technique on the designed network.

Chapter 4

Blockmodeling Technique

4.1 Introduction

Blockmodeling is a framework for analyzing social structure and a set of procedures for the segmentation of a network. The main task of blockmodeling is to obtain a simplified version of the network in a model that exhibits its fundamental structure. With this technique, actors that share structural characteristics in terms of equivalent relations are grouped together into equivalent classes or clusters and their relations are partitioned into blocks hence the term blockmodeling. The clusters of equivalent actors define the positions of the network and the blocks indicate the positions, roles. The blockmodeling technique usually deals with large and potentially incoherent networks which need to be simplified. Blockmodeling is grounded on different structural concepts like equivalence and positions which are related to the theoretical concepts of social role and role sets. It group vertices into clusters and determine the relations between these clusters and it uses matrices as computational tools and for the visualization of results.

4.2 Blockmodel in a Network

Notations used for blockmodel are:

- The network N = (U, *R*), where U is a set of all units U = (u_1, u_2, …,u_n) and R is the relation between these units R $\subseteq$ U×U. (The network can have multiple relations N = (U, R_1, R_2, …, R_m), where m is the number of relations.
- Relation R can also be presented by the matrix R = $[r_{ij}]_{n\times n}$, where r_{ij} is the value (or weight) of an arc (or edge in undirected graphs) from unit i (or u_i) to unit j (or u_j).
- **C** = {C_1, C_2... C_k} is a partition (or clustering) of the set U in K clusters. Φ is a set of all feasible clusterings. A clustering C also partitions the relation *R* into blocks. R (C_i, C_j) = R ∩ C_i × C_j. Each such block consists of units belonging to clusters Ci and Cj and all arcs leading from cluster C_i to cluster C_j. If i = j, the block R (C_i, C_i) is called a diagonal block.
- n_i, – the number of units in cluster C_i

As a simple example [26], consider the sociogram (network) as shown in Figure 4.1 and Table 4.1 in which the vertices are in an arbitrary order and labeled. In general the essential structure of a network is not obvious when networks are large and/or complex. Blockmodeling attempts to discern and represent network structure and does so by applying to some form of equivalence. In this case, specifying regular equivalence leads to Table 4.2 where the vertices of Table 4.1 have been permuted into the order shown in Table 4.2 and a coherent partition imposed. There are 5 positions and 15 blocks. The blocks in this table are either all null (null blocks) or take the form where each row and column of the blocks contain a 1 (regular blocks).

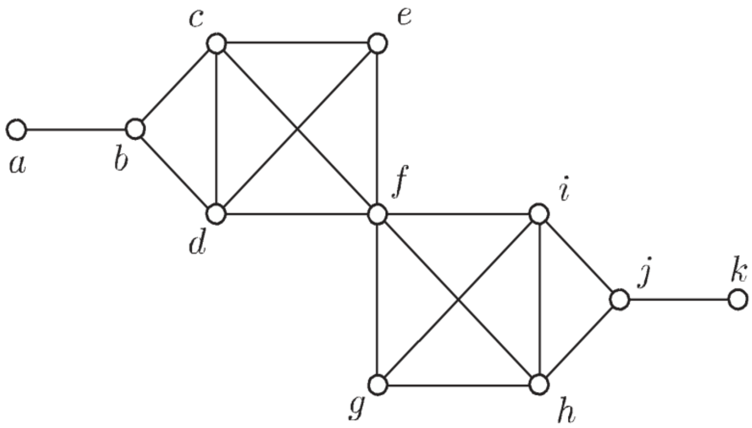


Figure 4.1: An Artificial Network [26]

	a	b	c	d	e	f	g	h	i	j	k
a	0	1	0	0	0	0	0	0	0	0	0
b	1	0	1	1	0	0	0	0	0	0	0
c	0	1	0	1	1	1	0	0	0	0	0
d	0	1	1	0	1	1	0	0	0	0	0
e	0	0	1	1	1	1	0	0	0	0	0
f	0	0	1	1	1	0	1	1	1	0	0
g	0	0	0	0	0	1	0	1	1	0	0
h	0	0	0	0	0	1	1	0	1	1	0
i	0	0	0	0	0	1	1	1	0	1	0
j	0	0	0	0	0	0	0	1	1	0	1
k	0	0	0	0	0	0	0	0	0	1	0

Table 4.1: Sociomatrix for the Artificial Network [26]

	a	k	b	j	f	c	d	h	i	e	g
a	0	0	1	0	0	0	0	0	0	0	0
k	0	0	0	1	0	0	0	0	0	0	0
b	1	0	0	0	0	1	1	0	0	0	0
j	0	1	0	0	0	0	0	1	1	0	0
f	0	0	0	0	0	1	1	1	1	1	1
c	0	0	1	0	1	0	1	0	0	1	0
d	0	0	1	0	1	1	0	0	0	1	0
h	0	0	0	1	1	0	0	0	1	0	1
i	0	0	0	1	1	0	0	1	0	0	1
e	0	0	0	0	1	1	1	0	0	0	0
g	0	0	0	0	1	0	0	1	1	0	0

Table 4.2: Blockmodel of the Artificial Network [26]

If regular blocks are labeled with 1 and the null blocks with 0 the image matrix achieved is shown in Table 4.3. This has a much simpler structure that captures the essence of Figure 4.1 where C_1 is {a, k}, C_2 is {b, j}, C_3 is {f}, C_4 is {c, d, h, i}, and C_5 is {e, g}. Authors [26] proved that every binary network has a class of regular equivalences that form a lattice. The network shown in Figure 4.1 has 21 regular equivalence partitions. Both result and the use of generalized blockmodeling make it clear that a network can have a (potentially large) number of optimal partitions rather than having a single blockmodel. This feature raises some interesting substantive and empirical tasks of assessing the set of such partitions in most empirical contexts when multiple optimal partitions are located.

	C_1	C_2	C_3	C_4	C_5
C_1	0	1	0	0	0
C_2	1	0	0	1	0
C_3	0	0	0	1	1
C_4	0	1	1	1	1
C_5	0	0	1	1	0

Table 4.3: Image of the Artificial Network

Blockmodeling tools were developed to partition network actors (units) into clusters, called positions, and to partition the set of ties into blocks that are defined by the position. Blockmodeling seeks clusters of equivalent units based on some notion of equivalence. Viewed as a method of data reduction, blockmodeling is a valuable technique in which redundant elements in an observed system are reduced to yield a simplified model of relationships among types of elements (units). The aim of blockmodeling is to combine more or less identical units and thus reduce the complexity or size of the network. Blockmodeling is usually seen 'as a tool for discovering roles and positions occupied by actors (units) in a social structure'.

4.3 Equivalence in Network

Equivalence is a cognitive operation that the sociologist perform to describe relations and social roles. In other words, it is a way of distinguishing subsets in a network to group individuals or social position according to the relations they keep with other.

Blockmodeling try to discover clusters of equivalent units based on some notion of equivalence. Equivalence has become a foundational concept in social network analysis. Each use of an equivalence concept has two components:

(i) The definition of equivalence
(ii) A computational algorithm for detecting equivalences

Several types of equivalences are used in social network analysis. The most widely used are structural and regular equivalence. Other equivalences include automorphic equivalence and prefect equivalence. The most commonly used equivalences are defined for binary networks.

4.3.1 Structural Equivalence

According to White and Reitz [28] structurally equivalent points (units) are related in the same way to each other and all other point (units). However this definition is not the only definition in existence. According to Lorain and White [23] 'Object 'a', 'b' of category C are structurally equivalent if for any morphism M and any object x of C, ${}_{A}M_{X}$ if and only if ${}_{B}M_{X}$, and ${}_{X}M_{A}$ if and only if ${}_{X}M_{B}$'. In other words 'a' is structurally equivalent to 'b' if 'a' relates to every 'x' object of C in exactly the same way as 'b' does. From logical view the structure 'a' and 'b' are absolutely equivalent since they are substitutable. Defined in other way, actors in a social network are structurally equivalent if they have identical ties to and from all the actors in a network.

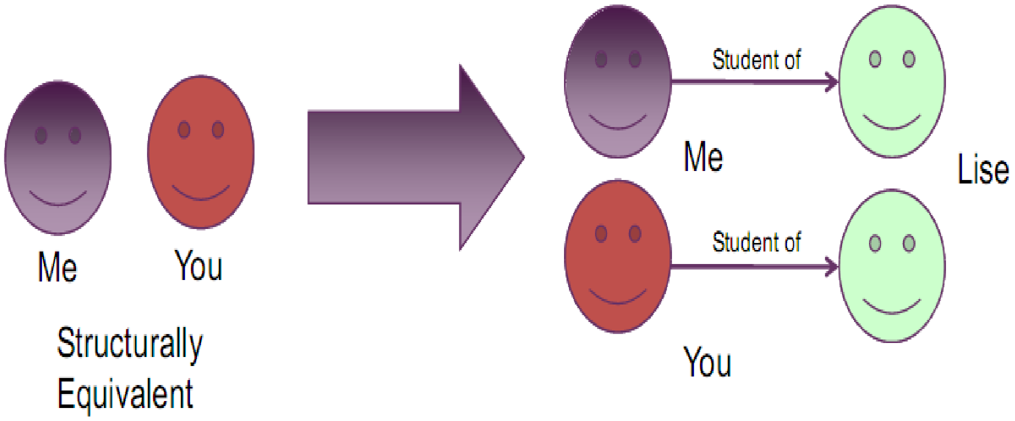


Figure 4.2: Structure Equivalence [29]

.

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

Figure 4.3: Ideal Blocks for Structural Equivalence [16]

Structural equivalence is well suited for combining identical units in a network and data reduction

4.3.2 Regular Equivalence

‘Two actor are regularly equivalent if they are related to equivalent other’. A more formal definition is given by [28]:

Definition: If G = (V, R) and ≡ is an equivalence relation on V then ≡ is a regular equivalence if and only if for all ‘a’, ‘b’ , c ϵ V, a ≡ b implies:

(i) aRb implies there exists d ϵ V such that bRd and d ≡ c

(ii) cRa implies there exists d ϵ V such that dRb and d ≡ c

According to Batagelj et al. [30] if **C** = {C_i} is a partition corresponding to a regular equivalence, then R(C_u, C_i) is either null (empty) or it has a property whereby there exists at least one 1 (tie) in each of its rows and in each of its columns.

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	1	0	0
0	0	1	0	1
0	1	0	0	0
1	0	1	1	0

Figure 4.4: Ideal Blocks for Regular Equivalence [16]

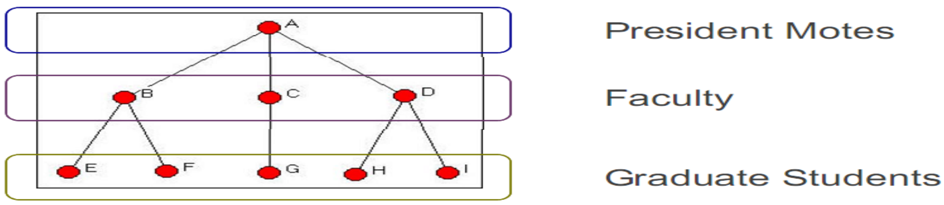


Figure 4.5: Regular Equivalence Levels [1]

4.3.3 Automorphic Equivalence

According to [20] automorphic equivalence is a natural generalization of (weak) structural equivalence. While two units that are structurally equivalent cannot be distinguished in a labeled network (if their labels are removed), two units that are automorphically equivalent cannot be distinguished in an unlabelled network.

4.4 Approaches to Blockmodeling

Two approaches used for blockmodeling are indirect methods and direct methods. Indirect methods are characterized by transforming one (or more) network(s) into some (dis) similarity matrix and then clustering this matrix by some clustering procedure. Direct methods skip the transformation(s) and work with the data directly.

4.4.1 Indirect Approaches

This approach involves two stages. First, a similarity (which might have to be converted to dissimilarity) or dissimilarity compatible with the selected equivalence has to be computed. This dissimilarity is then used as an input for standard data analysis techniques for finding clusters, usually hierarchical clustering. This approach can be used for any equivalence (not only structural) if a compatible similarity or dissimilarity measure exists.

4.4.2 Direct Approaches

Direct approaches directly search for a partition of the network that best fits the selected equivalence. The fit of the partition to the selected equivalence is measured by a selected criterion function. The selected criterion function should be computed using only network data (and the partition it evaluates) and should measure the extent to which a selected partition of units in a given network corresponds to the selected equivalence or other desired characteristics.

There are further two approaches in this class of blockmodeling, the first approach introduced by [31] says that structural equivalence in contingency tables is based on log-linear models. The second approach is the generalized blockmodeling approach, discussed in section 4.5.3.

4.4.3 Generalized Blockmodeling

Doreian *et al.* [26] state that there are three main characteristics of generalized blockmodeling.

- The direct approach is an optimizational one (the algorithm works directly with network data and does not transform them into some other form);
- A much broader set of block types is used to define equivalence instead of a few equivalence types.
- The model can be pre-specified (not only the allowed block types but also their positions).

There are several types of generalized blockmodeling techniques: binary blockmodeling valued blockmodeling, homogeneity blockmodeling and implicit blockmodeling. In the case of homogeneity blockmodeling two subtypes are defined: sum of squares blockmodeling and absolute deviations blockmodeling.

An appropriate generalization of the equivalence is one where each block, of a particular partition, is free to form to a different equivalence idea. The author [26] gives definition of several types of connection inside and between the clusters, or in other words, different types of ideal blocks as seen in Table 4.3.

The problem of establishing a partition of a network in terms of a considered equivalence is a special case of clustering problem that can be formulated as an optimization problem: determine the clustering C^* for which

$$P(C^*) = min_{c \in \emptyset} P(C)$$

$$\begin{array}{|ccccc|} 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0\\ 0&0&0&0&0 \end{array}$$

Null

$$\begin{array}{|ccccc|} 0&1&0&0&0\\ 1&1&1&1&1\\ 0&0&0&0&0\\ 0&0&0&1&0 \end{array}$$

Row-Dominant

$$\begin{array}{|ccccc|} 0&0&1&0&0\\ 0&0&1&1&0\\ 1&1&1&0&0\\ 0&0&1&0&1 \end{array}$$

Col-Dominant

$$\begin{array}{|ccccc|} 1&1&1&1&1\\ 1&1&1&1&1\\ 1&1&1&1&1\\ 1&1&1&1&1 \end{array}$$

Complete

$$\begin{array}{|ccccc|} 0&0&0&1&0\\ 0&0&1&0&0\\ 1&0&0&0&0\\ 0&0&0&1&0 \end{array}$$

Row-Functional

$$\begin{array}{|cccc|} 1&0&0&0\\ 0&1&0&0\\ 0&0&1&0\\ 0&0&0&0\\ 0&0&0&1 \end{array}$$

Col-Functional

$$\begin{array}{|ccccc|} 0&1&0&0&0\\ 1&0&1&1&0\\ 0&0&1&0&1\\ 1&1&0&0&0 \end{array}$$

Regular

$$\begin{array}{|ccccc|} 0&1&0&0&0\\ 0&1&1&0&0\\ 1&0&1&0&0\\ 0&1&0&0&1 \end{array}$$

Row-Regular

$$\begin{array}{|ccccc|} 0&1&0&1&0\\ 1&0&1&0&0\\ 1&1&0&1&1\\ 0&0&0&0&0 \end{array}$$

Col-Regular

Table 4.4 Generalized Ideal Blocks [26]

Where C is a clustering of a given set of units or units E, ∅ is the set of all possible clusterings and P: ∅→IR the criterion function. The criterion function must reflect the considered equivalence.

Criterion functions can be constructed:

- Indirectly as a function of a compatible (dis)similarity measure between pairs of units, or
- Directly as a function measuring the fit of a clustering to an ideal one with perfect relations within each cluster and between clusters according to the considered types of connections (equivalence).

4.5 2-Mode Concepts in Social Network Analysis

A typical data matrix has two dimensions or ways, corresponding to the rows and columns of the matrix. The modes of a matrix correspond to the distinct sets of entities indexed by the ways:

4.5.1 2-Mode Data

In social network analysis 2-mode data refers to data recording ties between two sets of entities. In this context, the term "mode" refers to a class of entities which called actors, nodes or vertices – whose members have social ties with other members (in the 1- mode case) or with members of another class (in the 2-mode case). Most social network analysis is concerned with the 1-mode case as in the analysis of friendship ties among a set of school children or advice giving relations within an organization. The 2-mode case arises when researchers collect relations between classes of actors such as persons and organizations or persons and events. For example a researcher might collect data on which students in a university belong to which campus organizations or which employees in an organization participate in which electronic discussion forums. These kinds of data are often referred to as affiliations. Co-memberships in organizations or participation in events are typically thought of as providing opportunities for social relationships among individuals. At the same time ties between organizations through their members are thought to be conduits through which organizations influence each other.

4.5.2 2-Mode Matrix

A (2-dimensional) matrix is said to be 2-mode if the rows and columns index different sets of entities e.g., the rows might correspond to persons while the columns correspond to organizations. In contrast a matrix is 1-mode if the rows and columns refer to the same set of entities, such as a city-by-city matrix of distances.

4.5.3 Unimodal Approaches to 2-Mode Data

One approach to handling 2-mode data in social network analysis is to convert the data in to 1-mode data. This is especially appropriate when the analytical interest focuses primarily only on one of the modes.

4.5.4 Unimodal Visualization of 2-Mode Data

A 2-mode matrix can be transformed into a 1-mode matrix by taking similarities among the rows (or columns) one can visualize the network using all the usual techniques for visualization of valued networks.

4.5.5 Bimodal Approaches to 2-Mode Data

Another approach to working with 2-mode data seeks to analyze both modes simultaneously. The data are seen to represent relations between two sets of nodes forming a bipartite graph GB(V1+V2,E) in which, or all u and v, (u , v) ∈ E if and only if *u* and *v* belong to different vertex sets. In other words, all ties are between vertex sets and none are within-group. The matrix representation of such a graph can be a rectangular incidence matrix X or a square bipartite adjacency matrix *B* with *n=n1+n2* rows representing both modes and an equal number of columns also representing both modes. In the latter case, the original matrix *X* forms a submatrix of the larger adjacency matrix *B* in which both rows and columns index the *V1+V2* entities.

4.5.6 Bimodal Analysis of 2-Mode Data

One approach to analyzing 2-mode data is to develop entirely new metrics and algorithms designed specifically for 2-mode data. Such techniques take awareness of the fact that the observed network is not just bipartite by coincidence but could not have been any other way. Few techniques of this kind have been developed, the exception being the area of 2-mode centrality measures, which has received significant attention [32].

4.5.7 Blocked Matrix

The blocked matrix rearranges the rows and columns (and inserted dividing lines or" partitions") to put actors with similar rows and columns together. The blocked matrix allows visualization of similarities that are defined for a group or groups.

4.6 Tools Used

During design and implementation phase following four tools were used. The description of each tool is given below:

KEDIT

KEDIT [33] for windows is a text editor for Microsoft windows. It provides many powerful and useful facilities for working with text files. KEDIT is typically used to computer programs, notes and memos, e-mail, lists of information and other textual data files.

With the help of KEDIT editor, a UCINET data file in DL format can be made. DL file consists of a set and phrases which describe the data. A DL file can consist only of the meta-data, along with a pointer to another file that contains the actual data.

UCINET

UCINET [34] is a comprehensive package for the analysis of social network data for both 1-mode and 2- mode data. It can read and write a multitude of differently formatted text files as well as Excel files. UCENET 6 for window runs on the platform with a minimum of 8 MB of RAM. It is available as a shareware from the web site of Analytic Technologies [34]. It can export data to MAGE and PAJEK.
UCINET' strengths include the ability to:

- Perform factor analysis of 2-mode data
- Construct bipartite graph from 2-mode data
- Plot graphical scatter plots and dendrograms
- Import and export Microsoft Excel files
- Accept additional input data format multiple output windows and more power powerful spreadsheet
- Handle large datasets limited only by the amount of memory available
- Construct bipartite graph from 2-mode data

PAJEK

PAJEK is a drawing package used by social network and expert for analyzing large network. It can handle very large network having some tens or hundreds of thousands of vertices and links. It was jointly developed in 1996 by Vladimir Batagelj and Andrej Mrvar from university of Ljubljana, Slovenia. PAJEK is implemented using Delphi program and runs on Windows operating system. It is freely available for non-commercial use and can be downloaded free of charge from its home pages at [35]. In PAJEK visualization and analysis of large networks are performed using six data type:

- **Network (graphs)**-the main object (vertices and lines). Default extension .net.

- **Partition**-(nominal or ordinal properties of vertices and ordinal properties). Default extension .per.
- P**ermutation** (re-ordering of vertices and ordinal properties). Default extension .per.
- **Cluster**-(subset of vertices e.g. a class from partition). Default extension .cls.
- **Hierarchies** (general tree structure on vertices)-hierarchically ordered clusters and vertices. Default extension .hie.
- **Vector** (numerical properties)- values of vertices. Default extension .vec.

PAJEK tool is used to analysis .net data files for given data set by using the blockmodeling technique. The steps used in this process are:

- Load the .net file in Pajek
- For high-quality printing matrix (without blockmodeling), save with command file->Network->Export matrix to EPS->Original
- Used the option menu command operation-> blockmodeling->random start
- Define the number of cluster (class) and equivalence
- To see the high-quality printing partition matrix by blockmodeling used two step
- Partition-> make permutation
- Save file as file->network->Export Matrix to EPS->Using Permutation

These files can be opened in GSview to see the partition in given network in the form of blockmodeling.

PAJEK strengths include the ability to:

- Contain several transformations that support different transitions among these data structures. The menu structure of the main PAJEK's window is based on them.
- Use its main window as a calculator paradigm with list-accumulator for each data type. The operations are performed on the currently active (selected) data and are also returned with the results through accumulators.
- Create special emphasis to automate generation of network layouts. Many standard algorithms for automatic graph drawing are implemented such as

spring embedders, layouts by eigenvalues/eigenvectors drawing in layers (genealogies and other acyclic structures), fish-eye view and block (matrix) representation. Properties of vertices and edges (given as data or computed) can be represented using colors, sizes and shapes of vertices and edges.

- Support drawing sequence of networks in its Draw window and exports sequences of network in suitable out graphic formats (Bitmap, Encapsulated PostScrpt (EPS), Scalable output vector Graphics (SVG), Virtual Reality (VRML), MDL MOL/Chime and Kinemages) that can be examined with special 2D or 3D viewers.

GSVIEW

GSview is a graphical interface for Ghostscript. Ghostscript is an interpreter for the PostScript page description language used by laser printers. For documents following the Adobe PostScript Document Structuring Conventions, GSview allows selected pages to be viewed or printed. GSview requires Ghostscript. GSview is available for Windows, OS/2 and Linux. In this thesis GSview is used to open the ESP PAJEK matrix files. This is downloaded from [36]. With GSview, we can:

- print selected pages using Ghostscript.
- convert pages to bitmap, PDF or PostScript.
- copy display bitmap to clipboard.
- save clipboard bitmap as BMP file.
- add bitmap or user preview to EPS file (Interchange, TIFF or Windows Metafile).
- graphically select and show bounding box for EPS file.
- extract bitmap preview or PostScript from DOS EPS file.
- read gzip and bzip2 compressed PostScript and PDF files.

Implementations and Result

Four tables have been considered for analysis:

1. CSED Data: Where 23 faculty members and their designations are considered.
2. CSED & Biotech dept combined data: Where 40 faculty members and their designations considered.
3. Game1: Where 10 games and 15 students are considered.
4. Game2: Where 15 games and 30 students are considered.

A comparative study of four set of 2-mode binary data has been used to analyze the social network by the blockmodeling technique. Tool KEDIT is used to make the data sets in dl format. These data sets are given in table 5.1, 5.2, 5.3 and 5.5. UCINET tool which discussed in section 4.6 is used to convert these data sets (network) from .dl format to .net format. This .net format data file is use as input file to analysis network (data set) in PAJEK.

Figure 5.1 show the binary network data file format which is save as .dl and used in UCINET tool to analysis the network data.

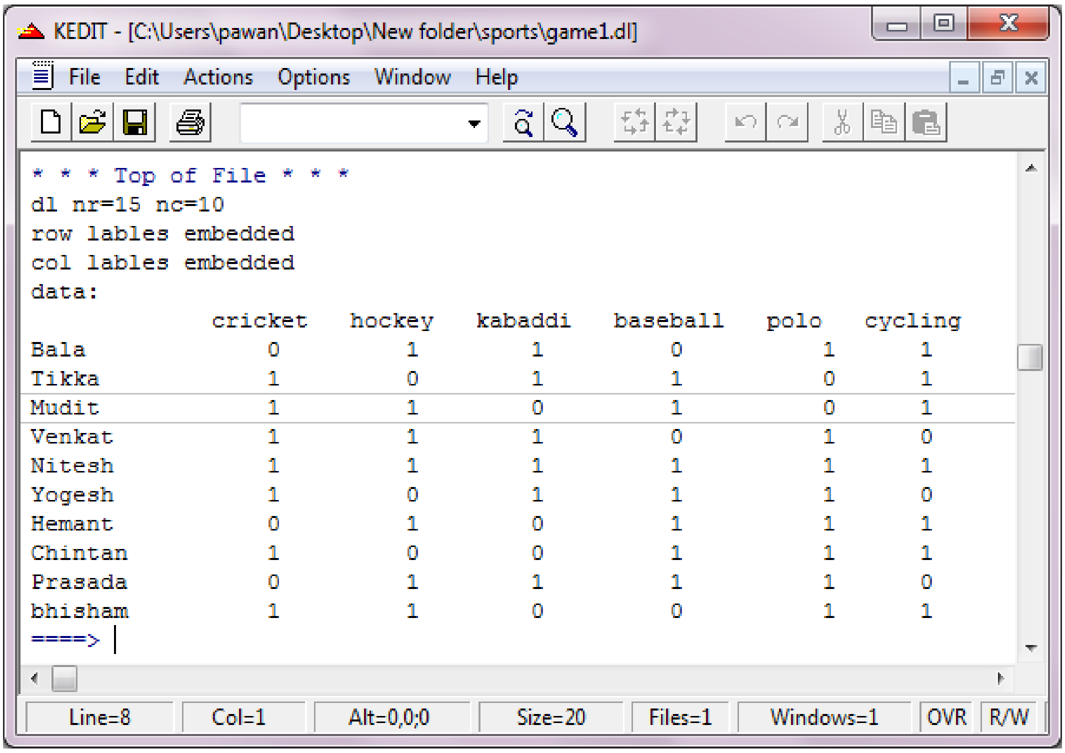


Figure 5.1: The .dl Format File for Matrix Data.

	HOD	PROFSSOR	ASSOCIATE PROFESSOR	ASSISTANT PROFESSOR	ADOH LECTURER	MALE	FEMALE	DOCTOR
Dr. Seema Bawa	0	1	0	0	0	0	1	1
Dr. Maninder Singh	0	0	1	0	0	1	0	1
Dr. Rajesh Kumar Bhatia	1	0	0	1	0	1	0	1
Dr. Deepak Garg	0	0	0	1	0	1	0	1
Dr. Varinder Pal Singh	0	0	0	1	0	1	0	1
Dr. Inderveer Chana	0	0	0	1	0	0	1	1
Mr. Anil Vashisht	0	0	0	1	0	1	0	0
Mrs. Rinkle Rani	0	0	0	1	0	0	1	0
Mr. Vinod Kumar Bhalla	0	0	0	1	0	1	0	0
Mrs. Shalini Batra	0	0	0	1	0	0	1	0
Mrs. Shivani Goel	0	0	0	1	0	0	1	0
Mr. Parteek Bhatia	0	0	0	1	0	1	0	0
Ms. Sushma Jain	0	0	0	1	0	0	1	0
Mr. Ravinder Kumar	0	0	0	1	0	1	0	0
Mr. Ajay Kumar	0	0	0	1	0	1	0	0
Ms. Ashima	0	0	0	1	0	0	1	0
Mr. Karun Verma	0	0	0	1	0	1	0	0
Mr. Sumit Miglani	0	0	0	1	0	1	0	0
Ms. Anju Bala	0	0	0	1	0	0	1	0
Mr. Ashish Aggarwal	0	0	0	1	0	1	0	0
Mr.Rajkumar Tekchandani	0	0	0	1	0	1	0	0
Mr. Vinay Arora	0	0	0	1	0	1	0	0
Ms. Navjot Kaur	0	0	0	0	1	0	1	0

Table 5.1: CSED Data in Binary Form

Table 5.1 represents the CSED data in 2- mode matrix in which row represent the faculty of department and columns represent the designation of faculty. This data is binary, small and sparse.

Table 5.2 represents the CSED & Biotech data in 2-mode matrix in which row represent the faculty of both department and column represent the designation.

	Hod	Professor	Associate professor	Assistant professor	Lecturer	Male	Female	Doctor	Csedept	BioTech &Extdept
Dr.N.Das	1	0	1	0	0	1	0	1	0	1
Dr.A.S.Reddy	0	0	0	1	0	1	0	1	0	1
Mr.Anoop Kumar	0	0	0	0	1	1	0	0	0	1
Dr.Abhijit Ganguli	0	0	0	1	0	1	0	1	0	1
Dr.Anita Rajor	0	0	0	1	0	0	1	1	0	1
Dr.Anil Kumar	0	0	0	1	0	0	1	1	0	1
Dr.Dinesh Goyal	0	1	0	0	0	1	0	1	0	1
Dr.M.S. Reddy	0	1	0	0	0	1	0	1	0	1
Dr.Moushumi Ghosh	0	0	0	1	0	0	1	1	0	1
Dr.Manju Anand	0	0	0	1	0	0	1	1	0	1
Dr.Sanjai Sexena	0	0	0	1	0	1	0	1	0	1
Mrs.M.Vasundhara	0	0	0	0	1	0	1	0	0	1
Dr.N.T. Prakash	0	0	0	1	0	1	0	1	0	1
Dr.BiswanathMohanty	0	0	0	0	1	1	0	1	0	1
Ms.Gurinderjit Kaur	0	0	0	0	1	0	1	0	0	1
Mr.Amit Dhir	0	0	0	0	1	0	1	0	0	1
Dr.Manoj Baranwal	0	0	0	0	1	1	0	0	0	1
Dr Seema Bawa	0	1	0	0	0	0	1	1	1	0
Dr Maninder Singh	0	0	1	0	0	1	0	1	1	0
Dr Rajesh Bhatia	1	0	0	1	0	1	0	1	1	0
Dr Deepak Garg	0	0	0	1	0	1	0	1	1	0
Dr Varinder Pal Singh	0	0	0	1	0	1	0	1	1	0
Dr Inderveer Chana	0	0	0	1	0	0	1	1	1	0
Mr Anil Vashisht	0	0	0	1	0	1	0	0	1	0
Mrs Rinkle Rani	0	0	0	1	0	0	1	0	1	0
Mr Vinod Bhalla	0	0	0	1	0	1	0	0	1	0
Mrs Shalini Batra	0	0	0	1	0	0	1	0	1	0
Mrs Shivani Goel	0	0	0	1	0	0	1	0	1	0
Mr Parteek Bhatia	0	0	0	1	0	1	0	0	1	0
Ms. Sushma Jain	0	0	0	1	0	0	1	0	1	0
Mr. Ravinder Kumar	0	0	0	1	0	1	0	0	1	0
Mr. Ajay Kumar	0	0	0	1	0	1	0	0	1	0
Ms. Ashima	0	0	0	1	0	0	1	0	1	0
Mr. Karun Verma	0	0	0	1	0	1	0	0	1	0
Mr. Sumit Miglani	0	0	0	1	0	1	0	0	1	0
Ms. Anju Bala	0	0	0	1	0	0	1	0	1	0
Mr. Ashish Aggarwal	0	0	0	1	0	1	0	0	1	0
Mr.RajkumarTekchandani	0	0	0	1	0	1	0	0	1	0
Mr. Vinay Arora	0	0	0	1	0	1	0	0	1	0
Ms. Navjot Kaur	0	0	0	0	1	0	1	0	1	0

Table 5.2: CSED & Biotech Data in Binary Form

	Cricket	Hockey	Kabaddi	Baseball	Polo	Cycling	Tennis	Basketball	Volleyball	Carrom	Snoker	Football	Swimming	Boxing	Shooting
Bala	0	1	1	0	1	1	1	0	1	1	1	1	0	0	1
Tikka	1	0	1	1	0	1	1	1	0	1	0	0	0	1	1
Mudit	1	1	0	1	0	1	1	0	1	0	1	1	1	0	1
Venkat	1	1	1	0	1	0	0	1	1	1	1	0	1	1	0
Nitesh	1	1	1	1	1	1	0	0	1	1	0	1	0	1	0
Yogesh	1	0	1	1	1	0	1	1	1	0	1	1	1	0	0
Hemant	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0
Chintan	1	0	0	1	1	1	1	0	1	0	0	1	1	1	1
Prasada	0	1	1	1	1	0	1	0	0	1	1	0	1	0	1
Bhisham	1	1	0	0	1	1	1	0	1	0	0	1	1	0	1
Satyavir	1	0	1	0	0	1	1	1	0	0	1	1	0	0	1
Ravinder	1	1	1	1	1	0	0	1	0	1	1	0	1	1	0
Virendra	1	0	1	0	0	1	1	0	1	0	0	1	0	1	1
Ravikiran	0	0	1	1	1	0	0	1	0	1	0	0	0	1	1
Gnendra	0	1	0	1	0	0	1	0	1	1	1	0	1	0	1
Jitesh	1	1	0	0	1	1	1	0	0	1	1	0	1	1	1
Deepak	0	1	1	0	1	0	1	1	1	0	0	0	1	0	0
Nitin	1	0	1	0	0	1	1	1	0	1	1	1	0	1	1
Pankaj	0	1	1	1	1	1	0	1	0	1	0	1	1	0	1
Sachin	1	0	1	1	0	1	0	0	1	1	1	0	0	1	1
Pawan	0	1	1	0	1	1	1	0	1	0	0	1	1	1	1
Manoj	1	0	0	1	1	0	1	1	0	0	1	1	0	0	0
Harpreet	0	1	0	1	1	0	0	0	1	1	1	0	1	0	0
Kushal	1	1	1	0	0	1	1	0	1	0	1	0	1	1	1
Vijay	1	1	1	0	1	1	1	0	1	1	1	1	0	0	1
Arun	0	1	0	1	1	1	1	1	1	1	1	1	0	0	0
Anand	1	1	1	0	1	0	0	1	1	1	1	0	1	1	0
Mohit	1	1	1	0	1	0	0	1	1	1	1	0	1	1	0
Ajit	1	0	1	1	0	1	0	1	0	0	1	1	0	0	1
Ankit	0	1	0	1	0	1	0	0	1	1	0	0	1	1	1

Table 5.3 Game2 Data in Binary Form

Table 5.3 represents 15 popular games as columns and 30 students were asked to fill their choice of whether they like to play a game or not. Here zero (0) show that the student does not like to play that game and one (1) show that student is interested in that game .Entire data has been represented as a binary table.

Table 5.4 shows the 15 student's Game1 data set in 2-mode matrix. This is also binary type data.

	Cricket	Hockey	Kabaddi	Baseball	Polo	Cycling	Tennis	Basketball	Volleyball	Carrom
Bala	0	1	1	0	1	1	1	0	1	1
Tikka	1	0	1	1	0	1	1	1	0	1
Mudit	1	1	0	1	0	1	1	0	1	0
Venkat	1	1	1	0	1	0	0	1	1	1
Nitesh	1	1	1	1	1	1	0	0	1	1
Yogesh	1	0	1	1	1	0	1	1	1	0
Hemant	0	1	0	1	1	1	1	1	1	1
Chintan	1	0	0	1	1	1	1	0	1	0
Prasada	0	1	1	1	1	0	1	0	0	1
Bhisham	1	1	0	0	1	1	1	0	1	0
Satyavir	1	0	1	0	0	1	1	1	0	0
Ravinder	1	1	1	1	1	0	0	1	0	1
Virendra	1	0	1	0	0	1	1	0	1	0
Ravikiran	0	0	1	1	1	0	0	1	0	1
Gnendra	0	1	0	1	0	0	1	0	1	1

Table 5.4: Game1 Data in Binary Form

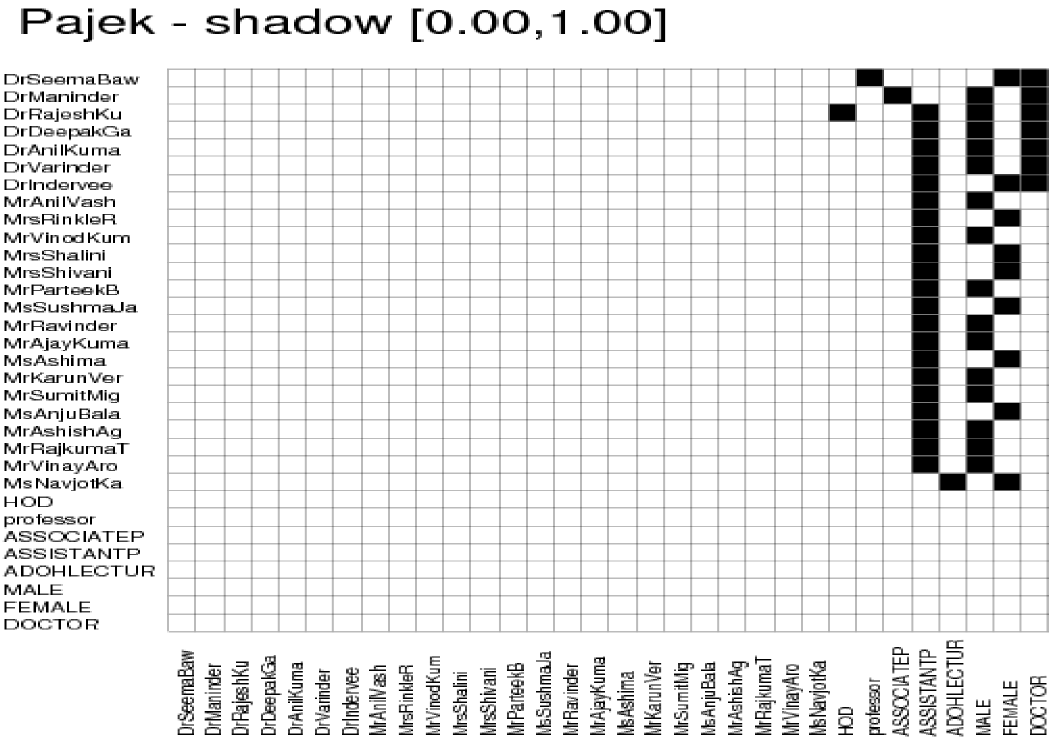


Figure 5.2 Original Matrix for CSED

Figure 5.2 displays the CSED data network of faculty and attributes in matrix form after analysis by the PAJEK without using the blockmodeling.

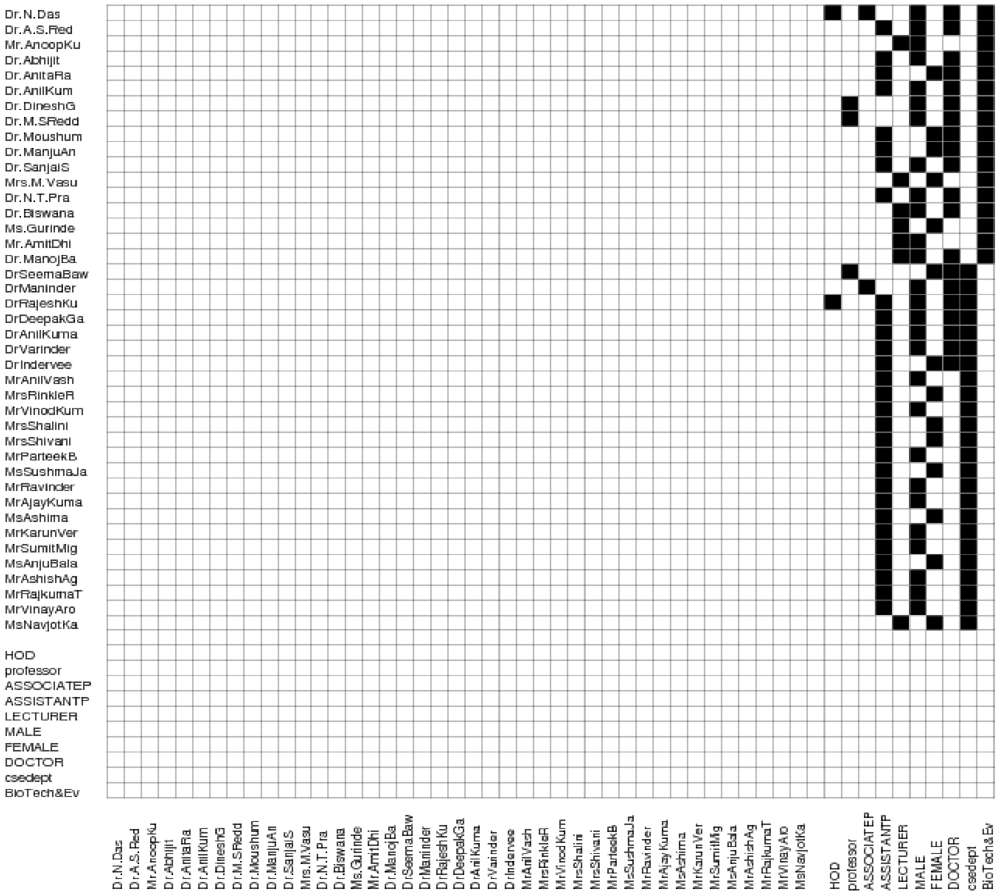


Figure 5.3: Original Matrix CSED & Biotech Data Dept.

Figure 5.3 displays the CSED & Biotech data network of faculty and designation in matrix form after analysis by the PAJEK without using the blockmodeling

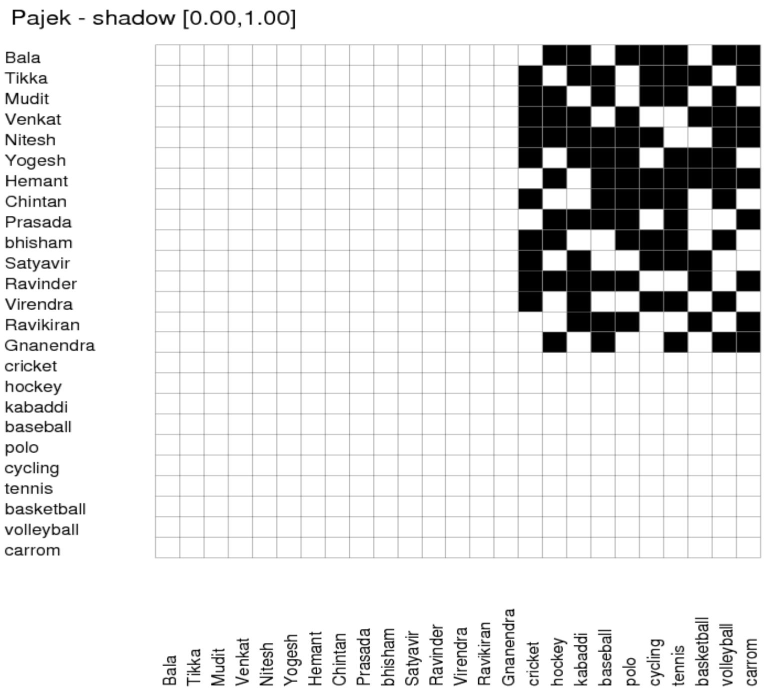


Figure 5.4 Original Matrix Game1 Data

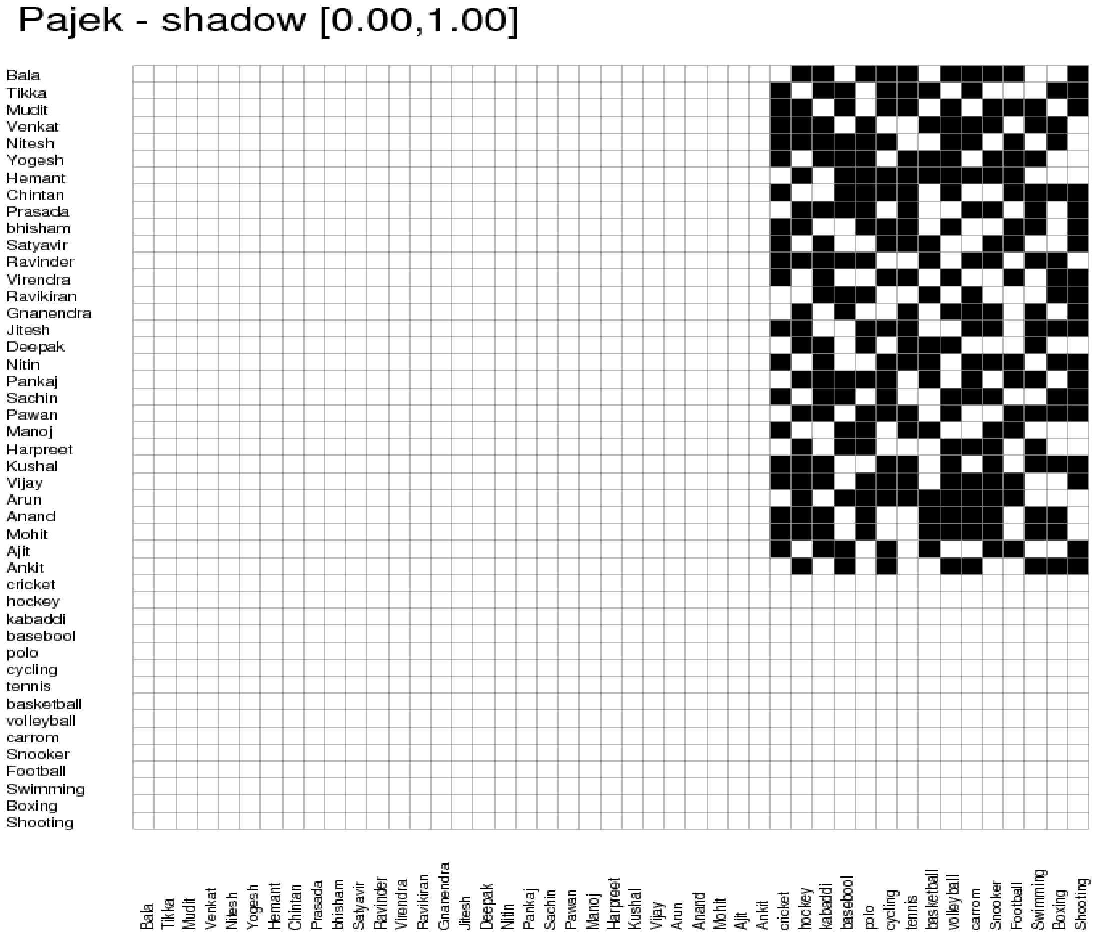


Figure 5.5 Original Matrix Game2 Data

Figure 5.4 displays the Game1 and Figure 5.5 display the Game 2 data network of student and games chosen by the student respectively, in matrix form, without blockmodeling.

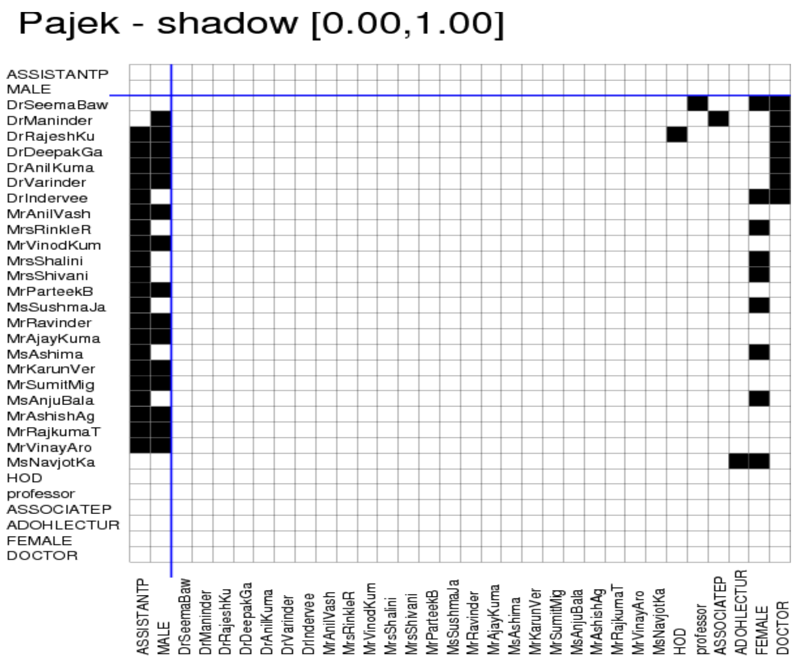


Figure 5.6 Blockmodel for CSED Data using 2 Clusters

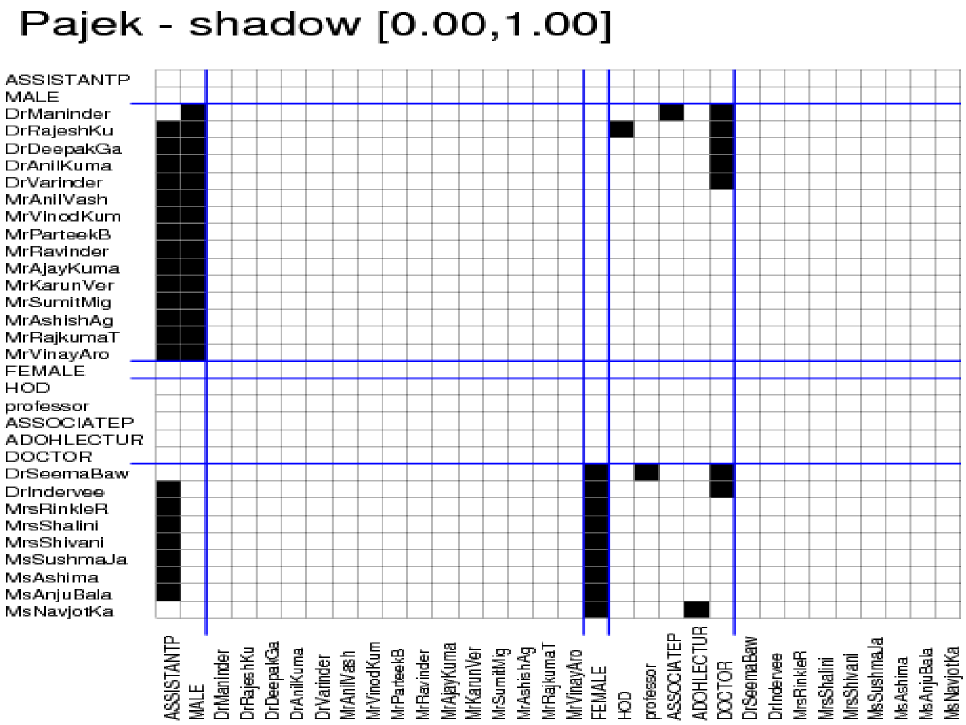


Figure 5.7 Blockmodel for CSED Data using 5 Clusters

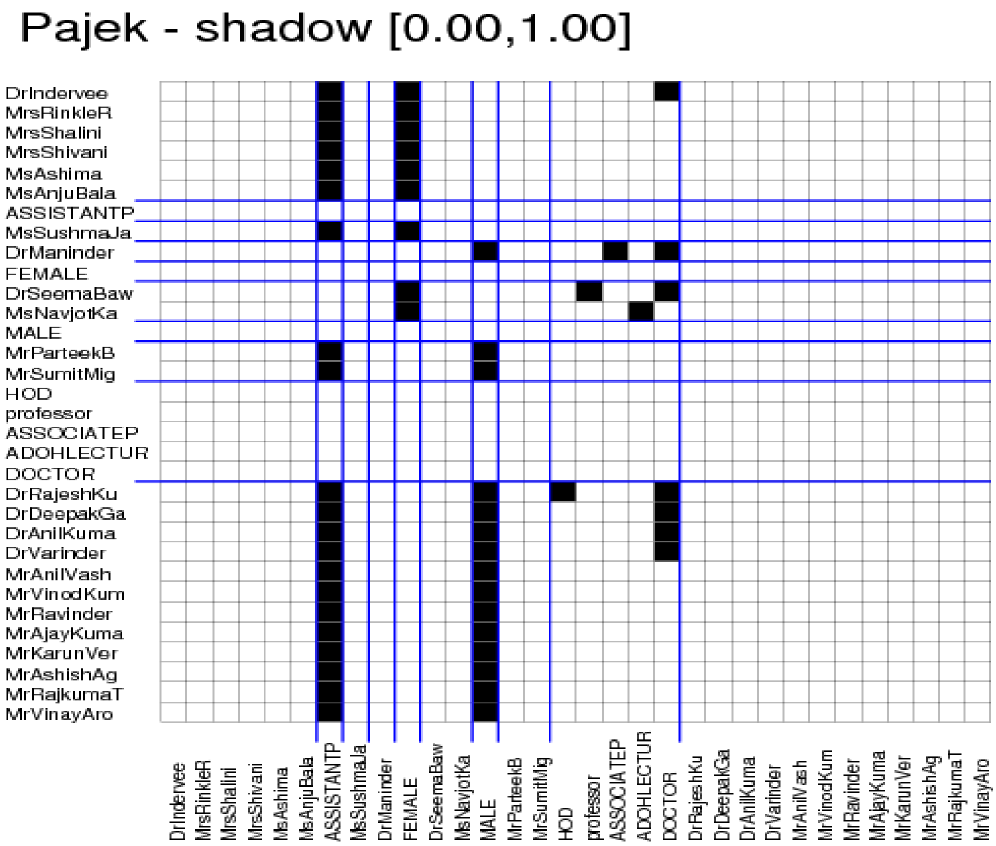


Figure 5.8 Blockmodel for CSED Data using 10 Clusters

Figure 5.6, 5.7 and 5.8 displays the CSED data network of faculty and attribute in matrix form using the blockmodeling.

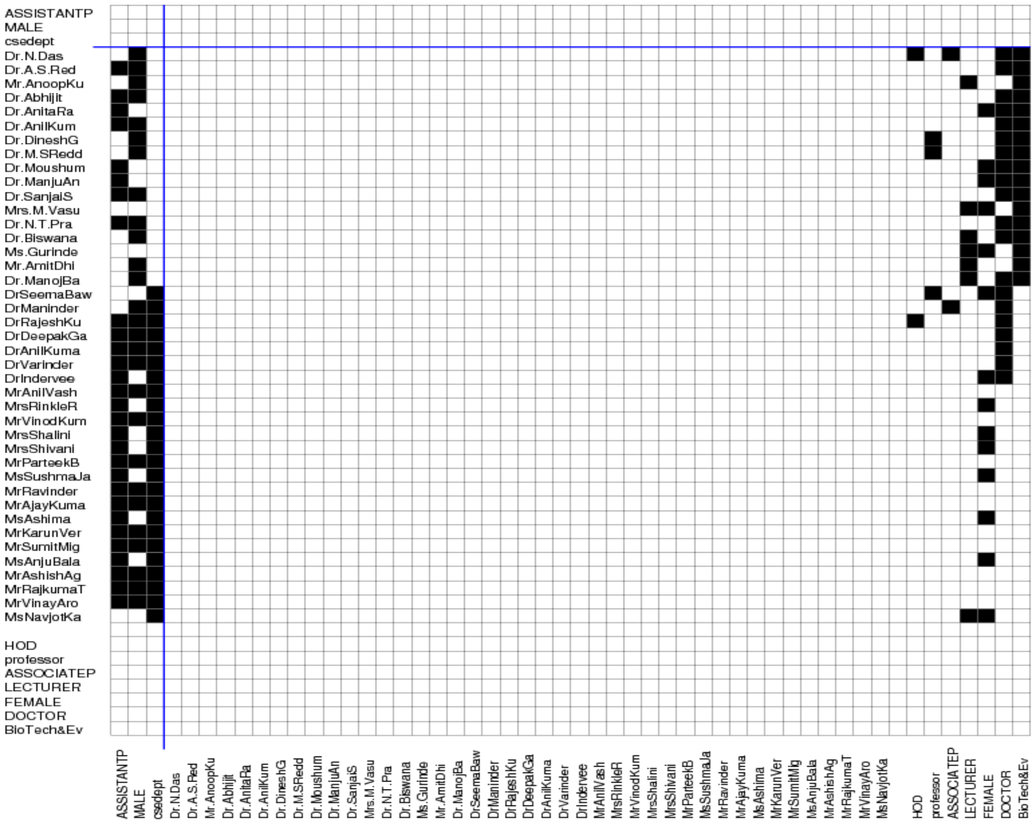


Figure 5.9 Blockmodel for CSED & Biotech Dept. using 2 Clusters

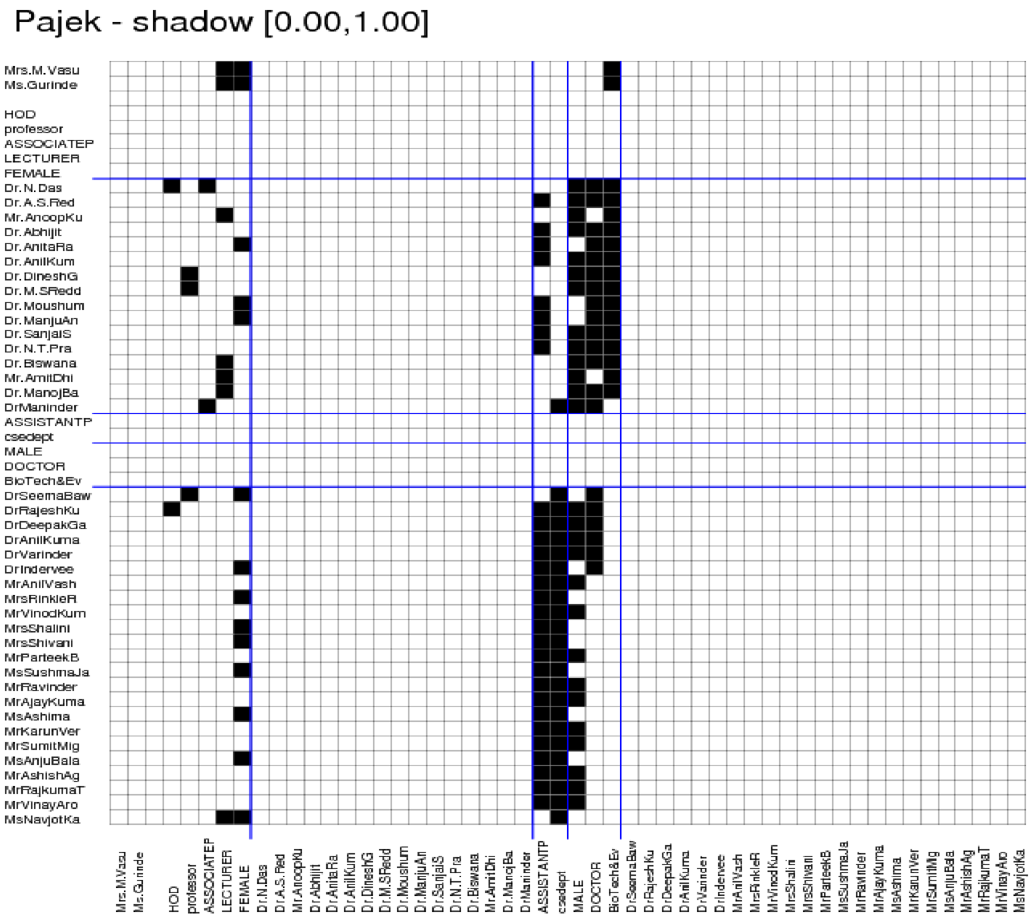


Figure 5.10 Blockmodel for CSED & Biotech Dept. using 5 Clusters

Figure 5.9, 5.10 and 5.11 display the CSED & Biotech data network of faculty and designation in matrix form after analysis by the PAJEK with using the blockmodeling.

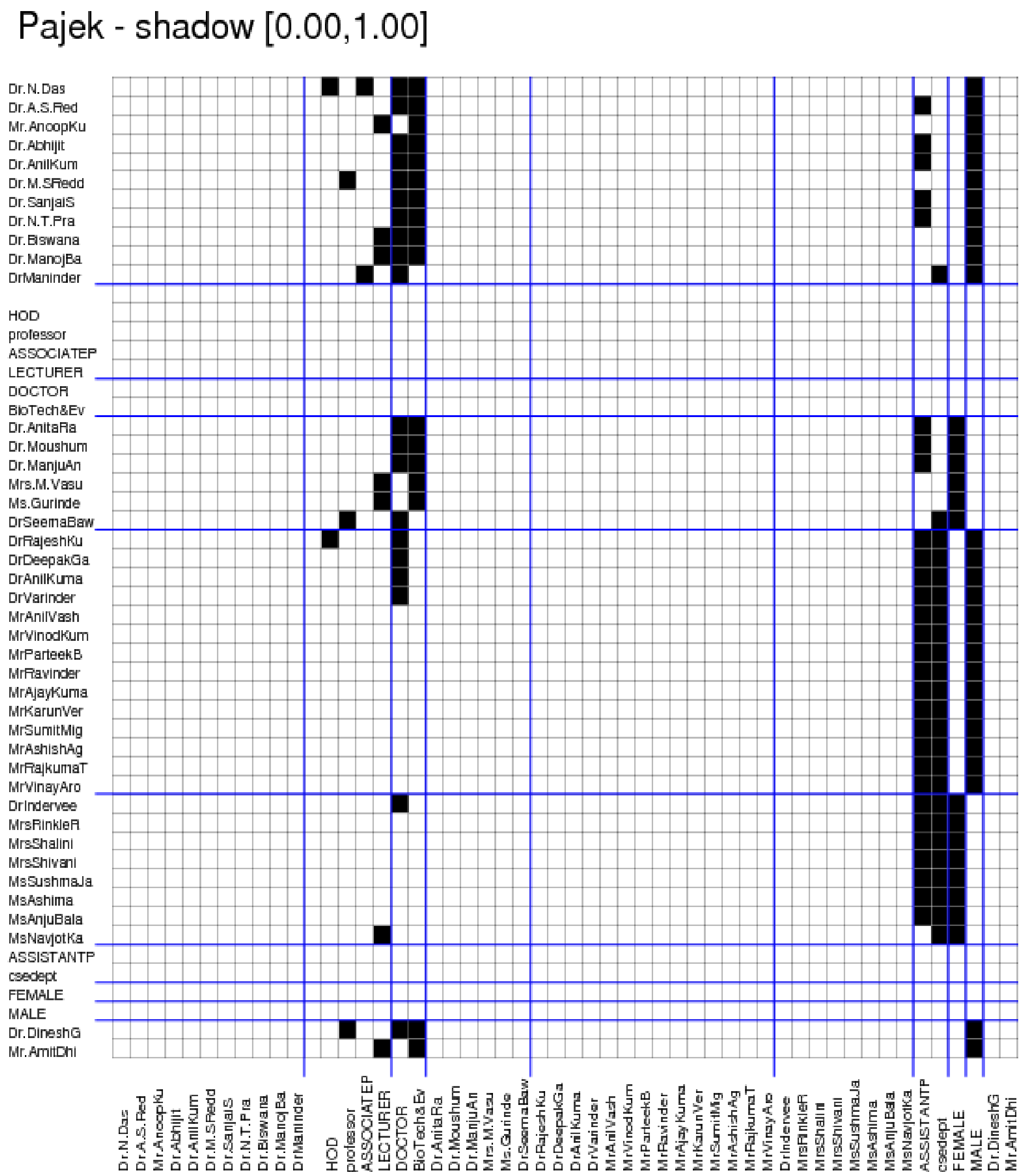


Figure 5.11 Blockmodel for CSED & Biotech Dept. using 10 Clusters

Figure 5.12, 5.13, 5.14, 5.15 and 5.16 on the next page display the Game1 data network after analysis by the PAJEK with using the blockmodeling.

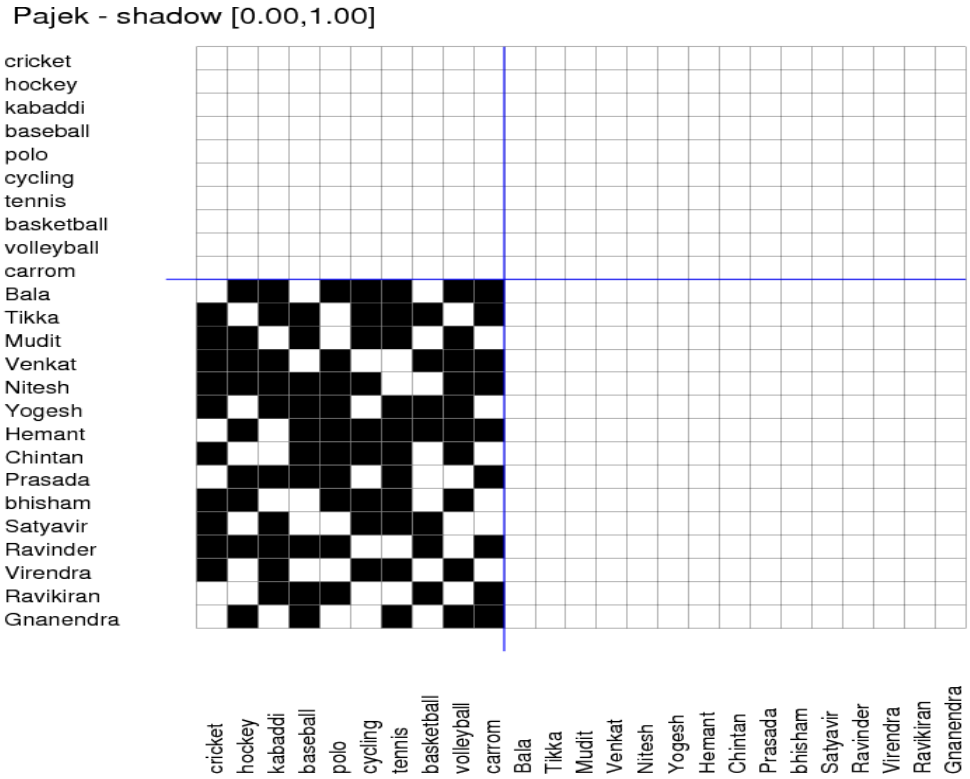


Figure 5.12 Blockmodel for Game1 using 2 Clusters

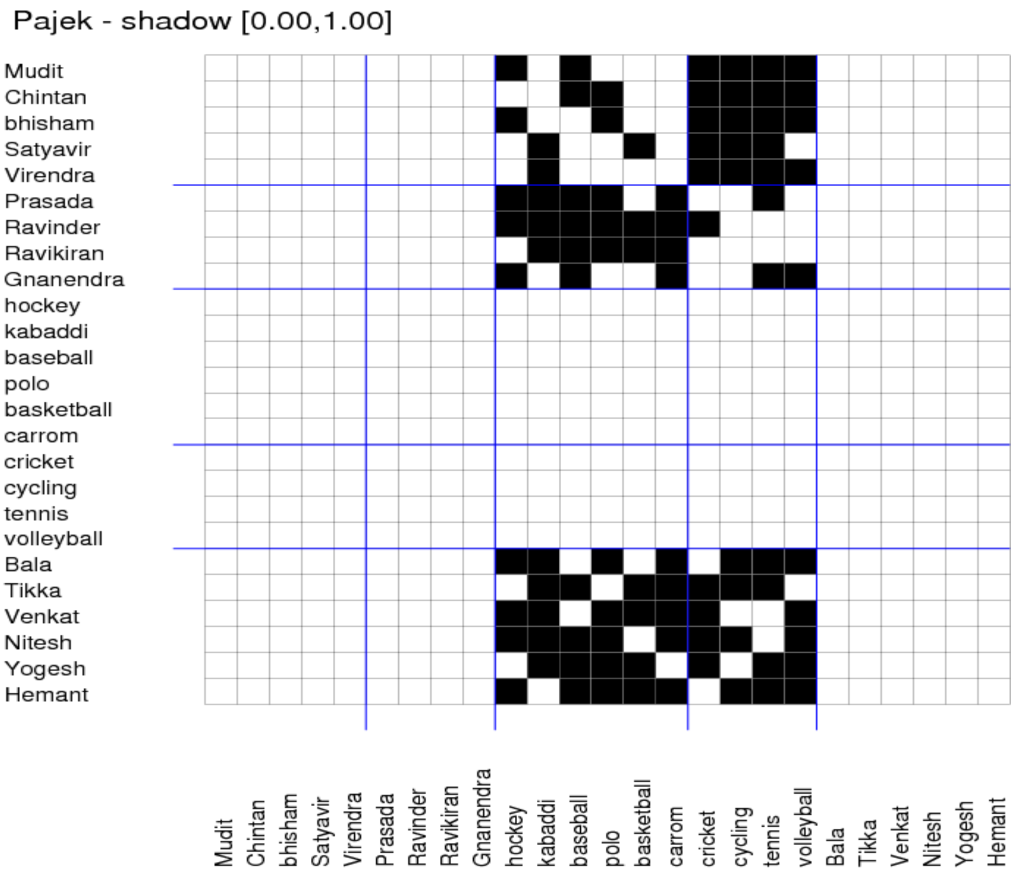


Figure 5.13 Blockmodel for Game1 using 5 Clusters

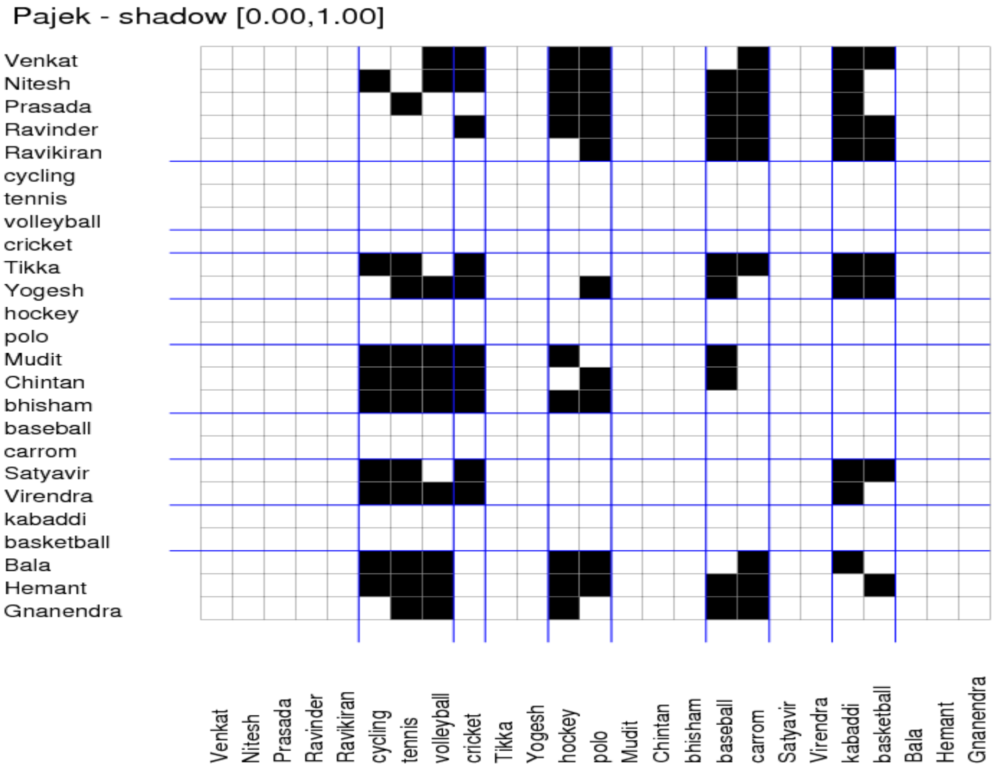


Figure 5.14 Blockmodel for Game1 using 10 Clusters

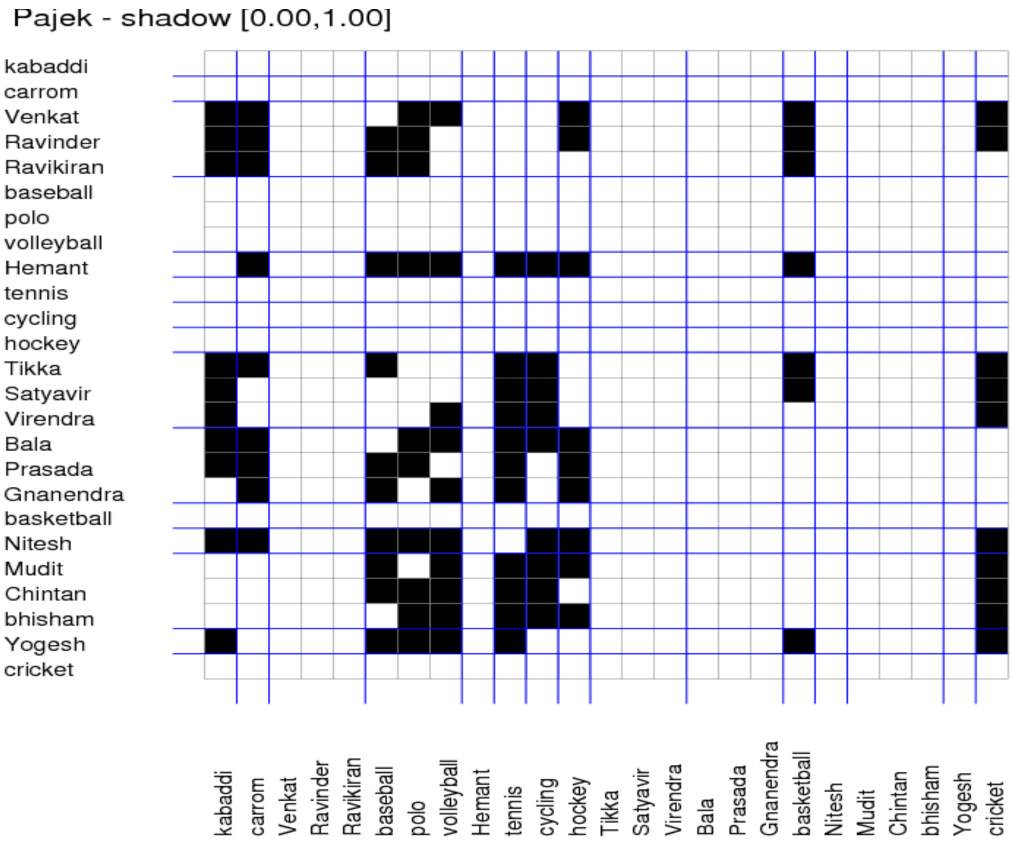


Figure 5.15 Blockmodel for Game1 data using 15 Clusters

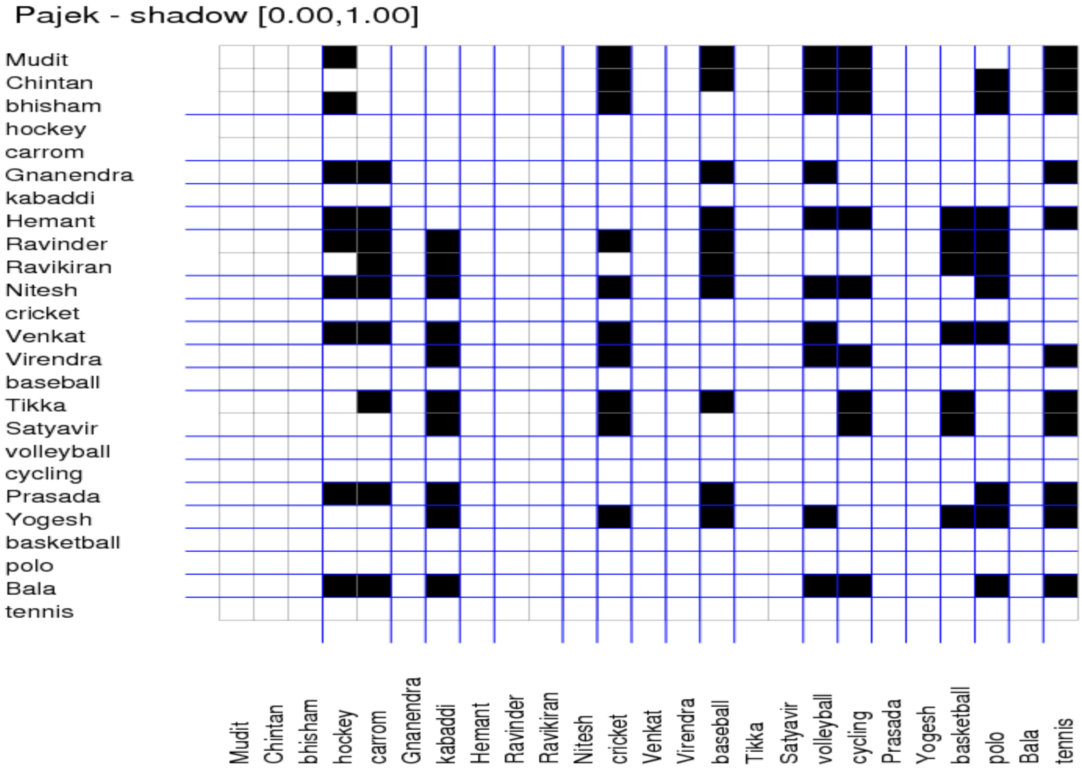


Figure 5.16 Blockmodel for Game1 Data using 20 Clusters

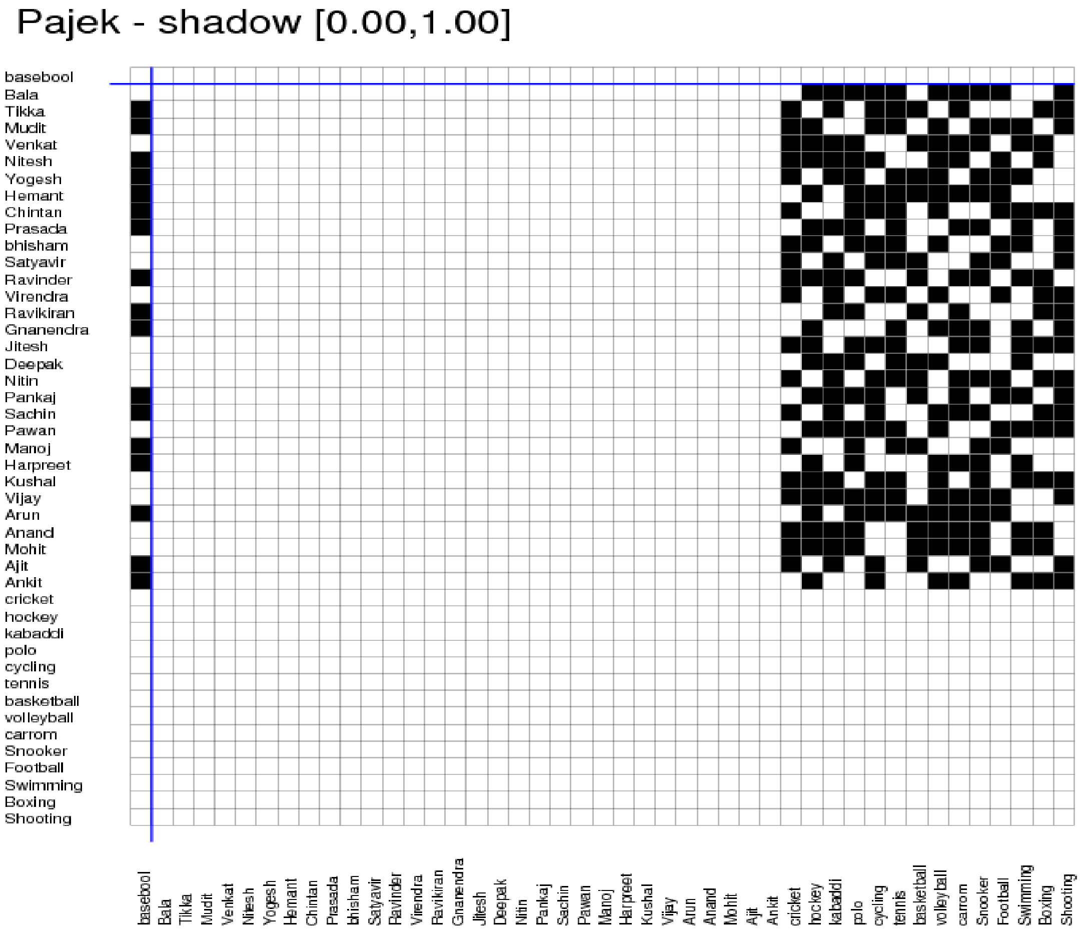


Figure 5.17 Blockmodel for Game2 Data using 2 Clusters

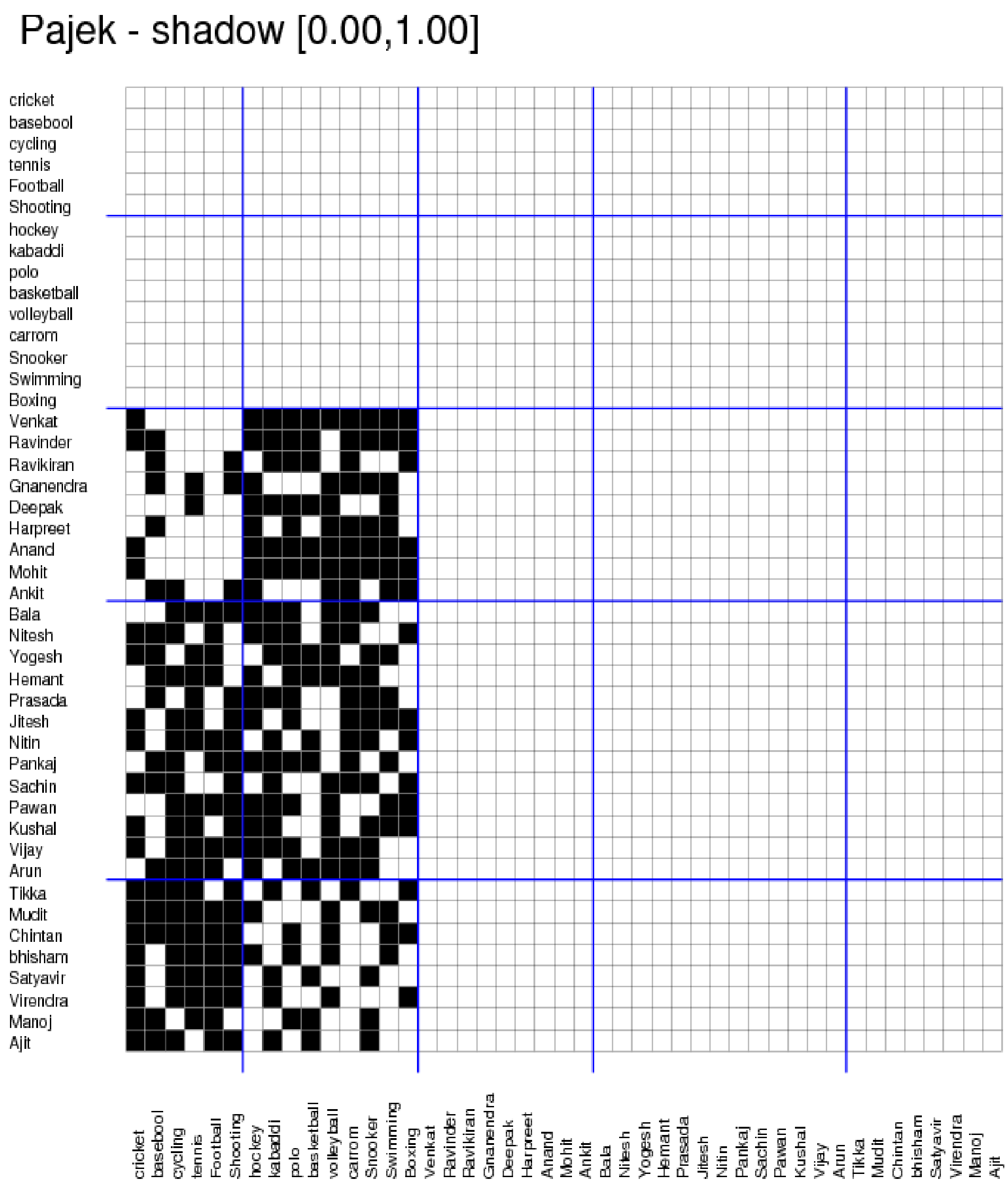


Figure 5.18 Blockmodel for Game2 Data using 5 Clusters

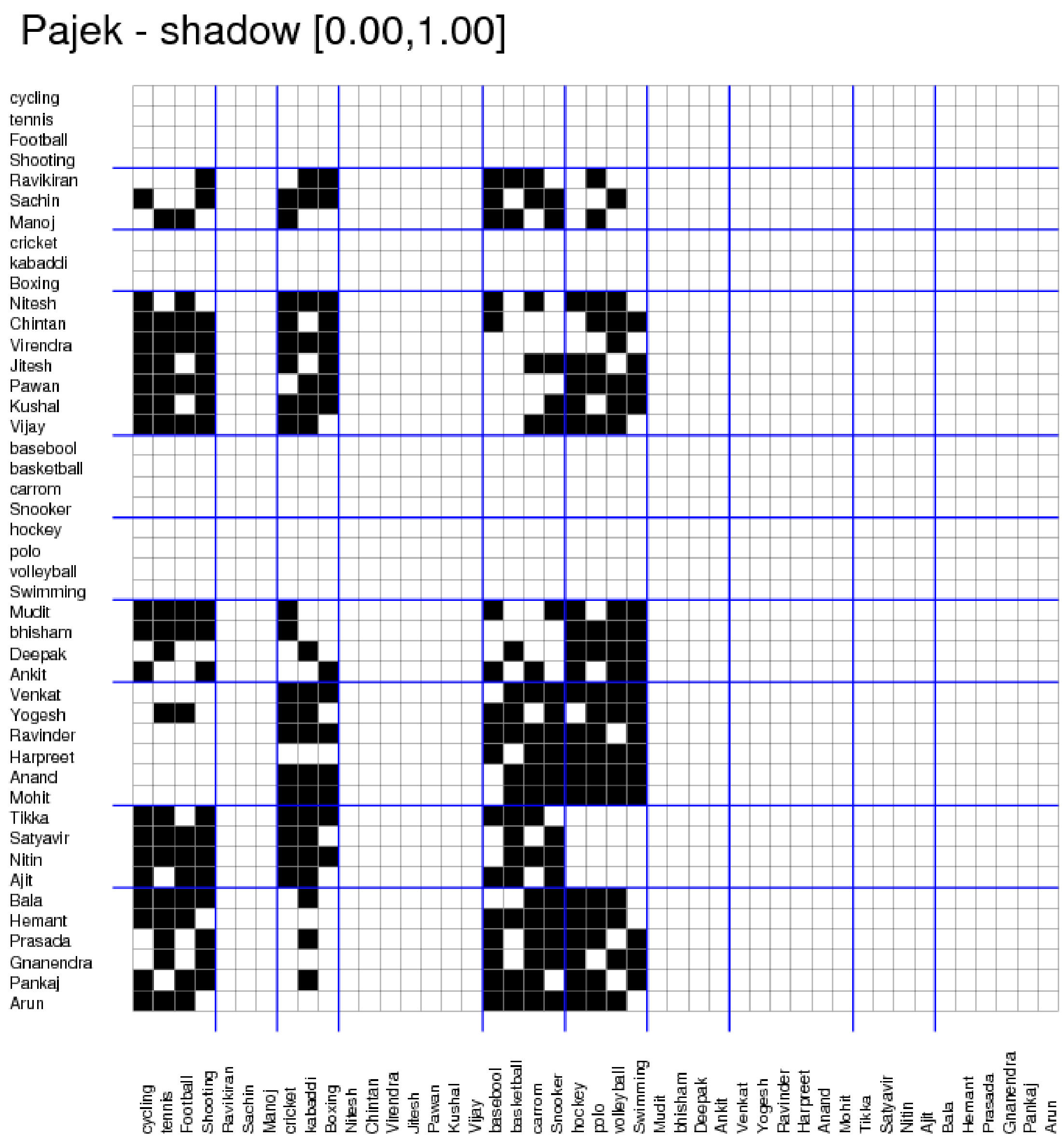


Figure 5.19 Blockmodel by using 10 Clusters' Game2 Data

Figure 5.17, 5.18, 5.19, 520 and 5.21 the Game 2 data network of student and game play by the student in matrix form after analysis by the PAJEK using the blockmodeling.

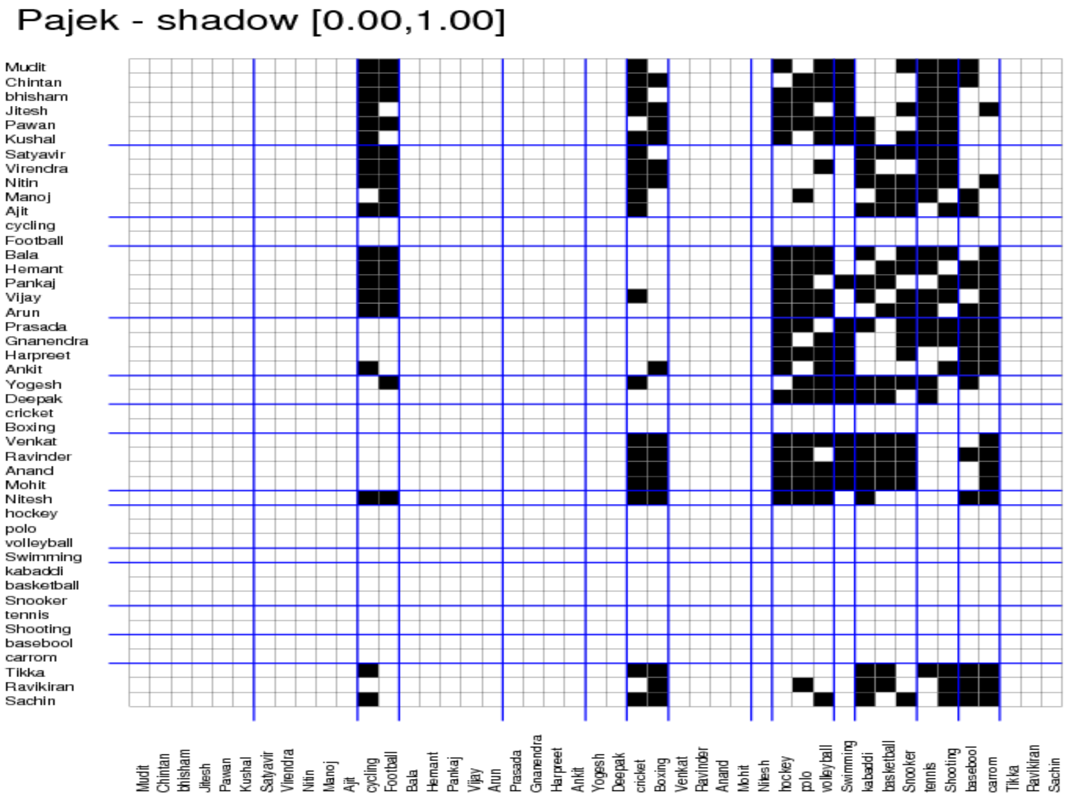


Figure 20: Blockmodel for Game2 Data by using 15 Clusters

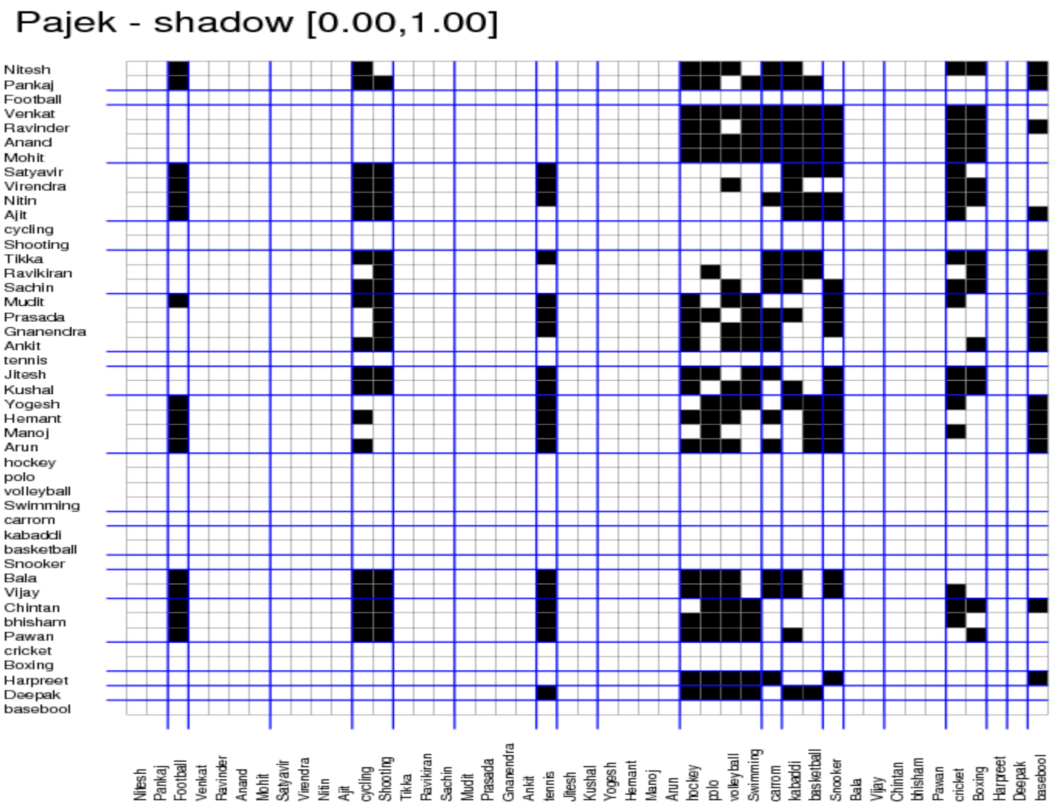


Figure 21: Blockmodel for Game2 Data using 20 Clusters

Interpretation of Matrices:

1) In figure 5.2 to 5.5 it is observed that all the data is clustered at one side of the matrix and hence no clear interpretation is possible
2) When 2 clusters are used in figure 5.6 interpretation is much better compared to 5.2
3) The clarity of the matrix goes on increasing as number of clusters goes on increasing, seen in figure 5.7 and 5.8. Same observation is made in figure 5.9, 5.10 and 5.11.
4) Although figure 5.12 gives a good idea about groups but the clarity increases as clusters size increases, evident in figures 5.13, 5.14 and 5.15.
5) As observed in previous points (1-4) clarity should increase in figure 5.16 since the clusters size has increased to 20 but just by visualizing the matrix grouping on the basis of student is not quite evident. Same observation is made in figure 5.17, 5.18 and 5.19.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

After analyzing the results achieved by applying blockmodeling techniques on the tables using different type of data sets and varying cluster sizes, it can be concluded that if the original matrix is considered without applying any modeling technique, network analysis becomes more difficult as the number of data sets increase. In case network is sparse, entire data tends to be concentrated on one side of the matrix in the original matrix but as the number of clusters increase network groups become more visible and can be easily identified. In case of dense networks, although the initial matrix gives a clear picture of the relationship between entities, the grouping and clarity increases manifold up to a certain limit on increasing the cluster size but when cluster size is increased to the level of 20 clusters, the analysis becomes more complex. So, it can be concluded that blockmodeling can analyze the data only when data set is small and increasing the cluster size beyond a certain limit increases the complexity of the analysis.

6.2 Future Scope

In our work the data was limited to a small group and no real time example of social network was considered but in future this can be extended to actual social networks. Secondly, we have considered only structural equivalence in blockmodeling but data analysis can also be done by using other equivalence techniques.

References

[1] S. Wasserman and K. Faust. Social Network Analysis: Methods and Applications Cambridge University Press, 1994. (page no 89). Cambridge University press.

[2] Scott, 2000. Social Network Analysis. A Handbook. Sage.

[3] Moreno, Jacob L. (1934): Who Shall Survive?: Foundations of Sociometry Group Psychotherapy and Sociodrama. Nervous and Mental Disease Publishing Washington DC.

[4] V. Krebs, 2006, How to do Social Network Analysis, [Online], Available: \ http://www.orgnet.com/sna.html

[5] Marin, Alexandra and Barry Wellman. Forthcoming. “Social Network Analysis: An Introduction.”in Peter Carrington and John Scott (eds.). Handbook of Social Network Analysis. Thousand Oaks:sage.

[6] Effective and Efficient Methodologies for SocialNetwork Analysis (phd thesis) Eunice E. Santos, Chair Elisa D. Sotelino Yang Cao Ezra Brown Eugene Santos

[7] Carrasco JA, Hogan B, Wellman B, Miller EJ (2008) Collecting social network data to study activity-travel behavior: egocentric approach. J Environ Plan B: Plan Des 35(6):961–980

[8] W. Nooy, A. Mrvar, and V. Batagelj, 2005, Exploratory Social Network Analysis with Pajek, Cambridge University Press.

[9] A. L. Barabasi, H. Jeong, Z. Neda, and E. Ravasz, A. Schubert, T. Vicsek, 2002, “Evolution of the Social Network of Scientific Collaborations”, Physica A, vol. 331, (3-4) pp. 590.

[10] M. E. J. Newman, 2001, “The Structure of Scientific Collaboration Networks”, Proceedings of the National Academy of Sciences of the United States of America, vol. 98, p. 404

[11] P. J. Carrington, J. Scott, and S. Wasserman, 2005, Models and Methods in Social Network Analysis, Cambridge University Press.

[12] S. Ressler, 2006, “Social Network Analysis as an Approach to Combat Terrorism: Past, Present and Future Research”, Homeland Security Affairs, vol. II, Issue 2.

[13] Network Analysis of an Exchange Program (Phd thesis) by Jaime Antonio Rivero Ostoic.

[14] Borgatti, S. P., & Everett, M. G. 1997. Network analysis of 2-mode data. Social Networks, 19(3): 243-269.

[15] Models and Methods in Social Network Analysis Series(book): Structural Analysis in the Social Sciences (No. 28) Edited by Peter J. Carrington.

[16] Generalized Blockmodeling with Pajek Vladimir Batagelj1, Andrej Mrvar(2004).

[17] Doreian P., Batagelj V., Ferligoj A.: *Generalized blockmodeling of two-mode network data.* Social Networks **26**(2004), 29-53.

[18] Blockmodels with maximum concentration European Journal of Operational Research 148 (2003) 56–64.

[19] Borgatti, Stephen P. and Martin G. Everett (1992b): Regular blockmodels of multiway multimode matrices. Social Networks 14, 91–120.

[20] Everett, Martin. G. and Stephen P. Borgatti (1994): Regular equivalence: General theory. Journal of Mathematical Sociology, 19 (1), 29-52.

[21] Winship, Christopher and Michael J. Mandel (1983): Roles and positions: A critique and extension of the blockmodeling approach. In: Samuel Leinhardt (ed.): Sociological Methodology 1983-84. Jossey-Bass, San Francisco.

[22] Doreian, Patrick, Vladimir Batagelj and Anuška Ferligoj (1994): Partitioning Networks on Generalized Concepts of Equivalence. Journal of Mathematical Sociology, 19 (1), 1–27.

[23] Lorrain, Francoise P. and Harrison C. White (1971): Structural equivalence of individuals in social networks. Journal of Mathematical Sociology, 1, 49–80.

[24] Edoardo Airoldi, David M. Blei, Stephen E. Fienberg, and Eric P. Xing. Mixed membership stochastic blockmodels. In NIPS, pages 33–40. MIT Press, 2008.

[25] Radoslaw Brendel, Henryk Krawczyk: Extended Generalized Blockmodeling for Compound Communities and External Actors. CASoN 2009: 32-39.

[26] DOREIAN, P., BATAGELJ, V. and FERLIGOJ, A. (2005): Generalized Blockmodeling. University of Cambridge, New York.

[27] BORGATTI, S. P. and EVERETT, M. (1989) The class of regular equivalences: Algebraic structure and computation. *Social Networks, 11, 65-88.*

[28] White, Douglas R. and Karl P. Reitz (1983): Graph and semigroup homomorphisms on networks of relations. Social Networks, 5, 193-234.

[29] Faust, K., & Wasserman, S. (1992a). Blockmodels: Interpretation and evaluation. Social Networks (Invited Paper), 14, 5–61.

[30] Batagelj, Vladimir, Patrick Doreian and Anuška Ferligoj (1992a): An optimizational approach to regular equivalence. Social Networks 14, 121–135.

[31] Breiger, Ronald L. and John W. Mohr (2004): Institutional Logics from the Aggregation of Organizational Networks: Operational Procedures for the Analysis of Counted Data. Computational & Mathematical Organization Theory, 10, 17–43.

[32] Bonacich, Phillip. 1991. Simultaneous group and individual centralities. Social Networks 13, 155-168.

[33] http://www.kedit.com

[34] http://www.analytictech.com

[35] http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm

[36] GSview http://www.cs.wisc.edu/∼ghost/

`

List of Publications

Communicated

[1] Pawan Kumar, Shalini Batra, “Analyzing the Social Networks using Block Modeling Technique”, communicated to International Journal of Computer Theory and Engineering.