

EFFECTS OF HOMOPHILY ON ACADEMIC PERFORMANCE



By

Zia Ullah

2009-NUST-MS-PhD-IT-29

Supervisor

Dr. Aimal Tariq Rextin

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of
Information Technology

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Pakistan.

(September 2013)

APPROVAL

It is certified that the contents of thesis document titled, “Effects of Homophily on Academic Performance” submitted by Zia Ullah have been found satisfactory for the requirement of degree.

Advisor: Dr. Aimal Tariq Rextin

Signature: _____

Date: _____

Committee Member 1: Dr. Muhammad Mudassar Malik

Signature: _____

Date: _____

Committee Member 2: Dr. Muhammad Usman Ilyas

Signature: _____

Date: _____

Committee Member 3: Ms Mehwish Naism

Signature: _____

Date: _____

DEDICATION

To my family

CERTIFICATE OF ORIGINALITY

I declare that, the research work titled “**Effects of Homophily on Education**” is my own work, and, to the best of my knowledge, it contains no materials previously published or written by another person, nor material which to a substantial extent, has been accepted for the award of any degree or diploma at SEECS or any other educational institute, except where due, acknowledgment is made in the thesis. Any contribution made to the research by others, with whom I have worked at SEECS or elsewhere, is explicitly acknowledged in the thesis.

I, also, declare that, the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in design and conception or in style, presentation and linguistic is acknowledged. I, also, verified the originality of contents through plagiarism software.

Author Name: Zia Ullah

Signature: _____

ACKNOWLEDGMENT

Thanks to Almighty Allah for writing the final words of my thesis. Before finish it, I must oblige to Dr. Aimal Tariq Rextin whose continuous guide and time to time counseling abled me to finish this work. Dr Aimal always drive me in right direction whenever I felt difficulty regarding my thesis. I also express my greatest gratitude to all my committee members including Dr.Muhammad Usman Ilyas, Dr.Muhammad Mudassar Malik and Ms Mehwish Nasim because their time to time affectionate help was a landmark for me when conducting this research. I specially thanked to Ms Mehwish Nasim whose generous help regarding development phase of my thesis made me able to do this work. I am also cardinally thankful to my friends and all family members for their moral support to fulfill all the requirement of this work.

Table of Contents

CHAPTER 1	11
INTRODUCTION	11
1.1 Overview of Social Networks	11
1.2 Visualizing a social network	12
1.3 Behavioral study and social network	13
1.4 Motivation	14
1.5 Major Contribution.....	15
1.6 Thesis Organization	16
CHAPTER 2	17
RESEARCH REVIEW.....	17
2.1. Why Social Networks Are Important?	17
2.2. Qualitative Measure of Human Behavior and Social Networks	19
2.2.1. Graph theory, Game Theory and Social Networks:	19
2.3 Technological Advancement and Social Networks	23
2.4 Homophily:.....	25
2.5 Overview of Educational Psychology:	25
2.5.1 Homophily and Constructivism:	26
2.5.2 Homophily as a cause of community formation in a class	27
2.6 Social Networking and student’s Interaction:	28
2.6.1 Causes of friendships on Social Networking Sites:	29
2.7 Techniques for social network analysis.....	30
2.7.1 Community Detection and connected components.....	30
2.7.2 Degree distributions and stress distributions	33
2.7.3 Neighborhood of nodes and connectivity.....	33
CHAPTER 3	35
EFFECTS OF HOMOPHILY ON ACADEMIC PERFORMANCE.....	35
3.1 Methodology and results.....	37
3.1.1 Data collection through Social Network Site:	38
3.1.2 Social networks data analysis	39
3.1.3 Community and exam data	41
3.1.4 Hypothesis	41
3.1.5 Similarity between grades of students belonging to same cluster	42
3.2 Conclusions	43

CHAPTER 4	44
RESULTS AND DISCUSSIONS	44
4.1 Introduction	44
4.2 Friendship of each user and similarity between grades and literacy rates.....	45
4.3 Efficacy and accuracy.....	64
CHAPTER 5	66
CONCLUSION AND FUTURE WORK	66
5.1 Conclusion.....	66
5.2 Future Work	66
REFERENCES	68

LIST OF FIGURES

Figure 1: A network formed by different communities	13
Figure 2: Behavioral structure of students, faculty and exam branch.	14
Figure 3: Usage of social networking sites by different age groups during 2005-2012....	15
Figure 4: Strong ties and Weak ties.....	18
Figure 5: Graph representing some basic definitions of Graph theory.....	20
Figure 6: Friends in common and triadic Closure	22
Figure 7: Socio-gram of class data collected through Facebook Application	24
Figure 8: Friendship matrix graph of users and commentee in class	36
Figure 9: Block diagram of methodology.....	37
Figure 10: Components of class with their percentages	40
Figure 11: Components of class community and their interactions	41
Figure 12: Cluster-1 Users Degree, CGPS and literacy rate.....	46
Figure 13: Cluster-2 Users Degree, CGPS and Literacy rate	47
Figure 14: Cluster-3 Users Degree, CGPS and Literacy rate	48
Figure 15: Cluster-4 Users Degree, CGPS and Literacy rate	49
Figure 16: Cluster-5 Users Degree, CGPS and Literacy rate	50
Figure 17: Cluster-6 Users Degree, CGPS and Literacy rate	51
Figure 18: Cluster-7 Users Degree, CGPS and Literacy rate	52
Figure 19: Cluster-8 Users Degree, CGPS and Literacy rate	53
Figure 20: Cluster-9 Users Degree, CGPS and Literacy rate	54
Figure 21: Cluster-10 Users Degree, CGPS and Literacy rate	55
Figure 22: Cluster-11 Users Degree, CGPS and Literacy rate	56
Figure 23: Cluster-12 Users Degree, CGPS and Literacy rate	57
Figure 24: Cluster-13 Users Degree, CGPS and Literacy rate	58
Figure 25: Cluster-14 Users Degree, CGPS and Literacy rate	59
Figure 26: Cluster-15 Users Degree, CGPS and Literacy rate	60
Figure 27: Cluster-16 Users Degree, CGPS and Literacy rate	61
Figure 28: Cluster-17 Users Degree, CGPS and Literacy rate	62
Figure 29: Cluster-18 Users Degree, CGPS and Literacy rate	63

LIST OF TABLES

Table 1: Users of Cluster-1	38
Table 2: Users of Cluster-2	39
Table 3: Users of Cluster-3	40
Table 4: Users of Cluster-4	41
Table 5: Users of Cluster-5	42
Table 6: Users of Cluster-6	43
Table 7: Users of Cluster-7	44
Table 8: Users of Cluster-8	45
Table 9: Users of Cluster-9	46
Table 10: Users of Cluster-10	47
Table 11: Users of Cluster-11	48
Table 12: Users of Cluster-12	49
Table14: Users of Cluster-13	50
Table 14: Users of Cluster-14	51
Table 15: Users of Cluster-15	52
Table16: Users of Cluster-16	53
Table 17: Users of Cluster-17	54
Table 18: Users of Cluster-18	55

ABSTRACT

Social networking is one of the most important fields of sociology and with emergence of social networking sites such as Facebook; Twitter etc. importance of this field is even more. The possible sociological force behind formation of social networks is 'homophily', which means that people tend to form friendships links with people who are similar to them (social selection) or people modify their behavior in order to conform to the behavior of their friends (social influence). Homophily is an artifact of these two social phenomena. In field of educational psychology "Homophily" is considered as the major cause of community formation in a class. For analysis of student's behavior, it is important to study the effects of "Homophily". In this research, online data of a class from social networking site "Facebook" is collected by developing a Facebook application. By analyzing data using social networks analysis techniques student's communities are found. Interaction levels and friendship strengths are studied on the basis of pokes, messages, wall posts and comments. Then student's results are compared with their friends and peers and all those who are involved in community formation. As it is comparatively easier to visualize a complex structure of human behavior in social networks through socio-gram hence all results are visualized through different graphs. Different techniques are used for social network data analysis and then data is co-related with student's grades. Results showed homophily in class on the basis of gender, as 89.47% female students have friendship with females and 86.63% male students are friends with male. Similarly, students belonging to well-off regions have more interactions with each other. Interactive patterns also showed an interesting result about student relationship i.e. student's belong to same cluster or group have similar grades and their belonging regions have similar literacy rates.

CHAPTER 1

INTRODUCTION

Social media such as Facebook, Twitter etc. are very important source of interaction with others and are used abundantly in our daily life. About 97% of college student are actively engaged on Facebook or other social media (Smith, Caruso, 2010). In this research Facebook data is analyzed for finding the friendship patterns of a class; and then effects of those friendships on student's grades are found. This is a cross-sectional research that belongs to the field of educational psychology and social networks analysis.

1.1 Overview of Social Networks

In social networks, the term “*networks*” refers to the connection among people and the term “*social*” refers to the structure of society. A “social network” is a *social structure* made up of individuals or groups of individuals that are connected to each other. The reasons of formation of groups and communities are interactions among the individuals and groups. These interactions are on the basis of some relations such as friendship, kinship, common interest, likes, dislike, religion etc. (Easley, Kleinberg, 2010)

Main focus of social networks is to find and visualize ties among people. Structure consists of “*ties*” or “*interpersonal ties*”. The interpersonal ties show the behavior of an individual and overall behavior of society. Due to the importance of behavioral study, social analyst gave importance to social ties for studying the behaviors of individual and groups because these relationships form communities. (Joyner, K. & Kao, G., 2001) For instance, students having same age and belong to same class form a community. A community may be formed due to common property for instance; same language, religion and belong to one nation,

same ethnic group. All these properties provide reasons for community formation. These common properties are called homophily and it is defined as;

“Common property that is the cause of community formation and makes interpersonal ties among individuals is called homophily”. (McPherson, et al. 2001)

Homophily is any similarity that cause or tends to causes connection among individuals and group of individuals. These individuals and their groups are bonded by “*strong ties*” if they belong to same community. People in one community may also link with other communities due to common behavior form “*weak ties*”. (Easley, Kleinberg, 2010)

1.2 Visualizing a social network

A *Social network* can be visualized and analyzed through graphical and behavioral structures respectively. The graphical structure is called socio-gram and behavioral structure is called interaction diagram.

“A graphical structure of a network provides us a clear view of complex relationships of data of a network.” (Easley, Kleinberg, 2010)

A graphical structure of a social network or socio-gram of a simple social network is shown in fig.1. Nodes from P1 to P8 represent people and edges between them represent interactions or connections.

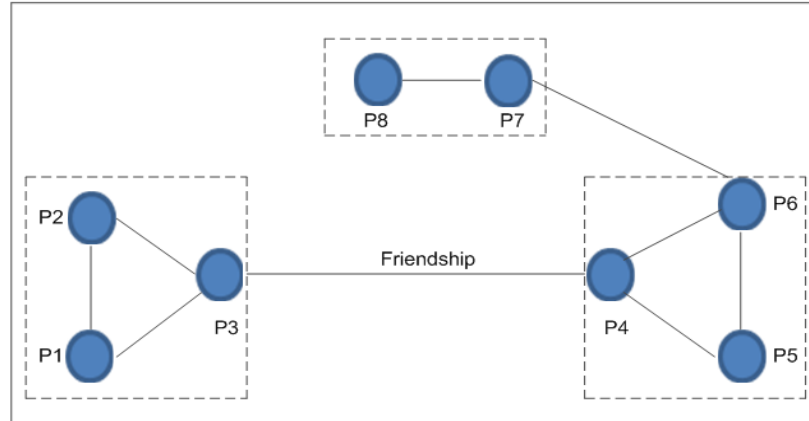


Figure 1: A network formed by different communities

Nodes P1, P2 and P3 form one community and nodes P4, P5, P8 form another community. A community formation depicts that there must be a tie between these nodes.

1.3 Behavioral study and social network

Studying the behavioral structure gives information about interpersonal relation and comes under the field of study known as “*socio-metry*”. (Berger, Peter L., Luckmann T., 1966)

“Socio-metry is branch of social sciences that deals with the study of interpersonal relations”. (Moreno, Jacob L., 1953)

In this study “socio-metric techniques” are applied for the analysis of social network data. This data is gathered from social media sites. Socio-metric techniques are discussed in details in chapter 2, section 2 .7.

People having common properties are grouped together and form a social network. For example, college students are more likely to be friends among themselves. Similarly faculty members have more interaction among them. Behavioral structure or interaction diagram of students, faculty and employees of exam branch is shown in fig. 2.

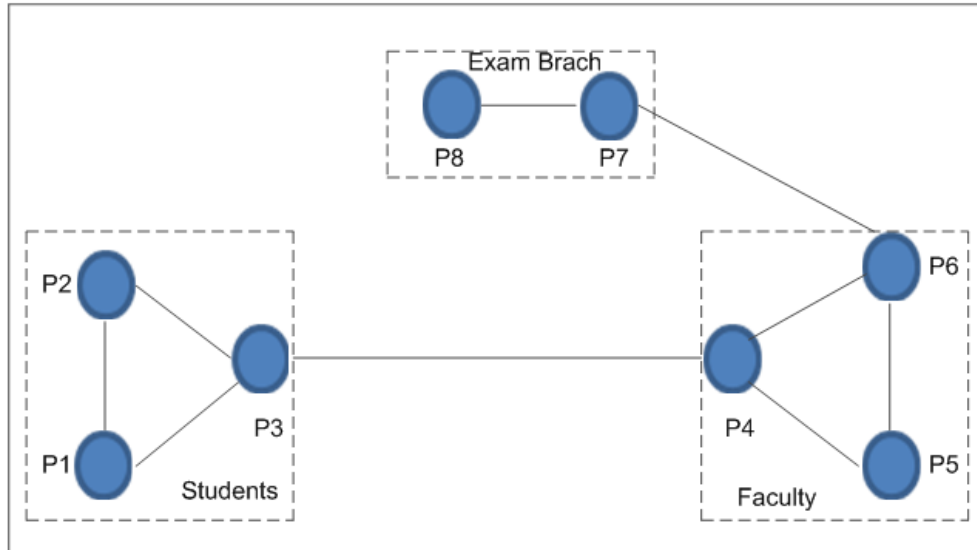


Figure 2: Behavioral structure of students, faculty and exam branch.

1.4 Motivation

Social networks deal with visualization of human behavior and relationships. There are many types of networks such as; networks of players, students, terrorists etc. all the members belong to these networks have some connective property called “homophily”. These networks are assembled on the basis of one or more common agenda. Studying the common agenda and impacts of that agenda is an interesting but an uphill task.

In this technological age “Social networking sites” have played an important role in growth of social networks. These sites act as an important tool for social networks. Demographics of social media users with respect to internet users in percentage from 2005 to 2012 are shown in Fig. 3 (Maeve, D., Joanna. B., 2013). The importance of SNS is clear from rapid growth of usage of SNS by internet users from 2005 to 2012.

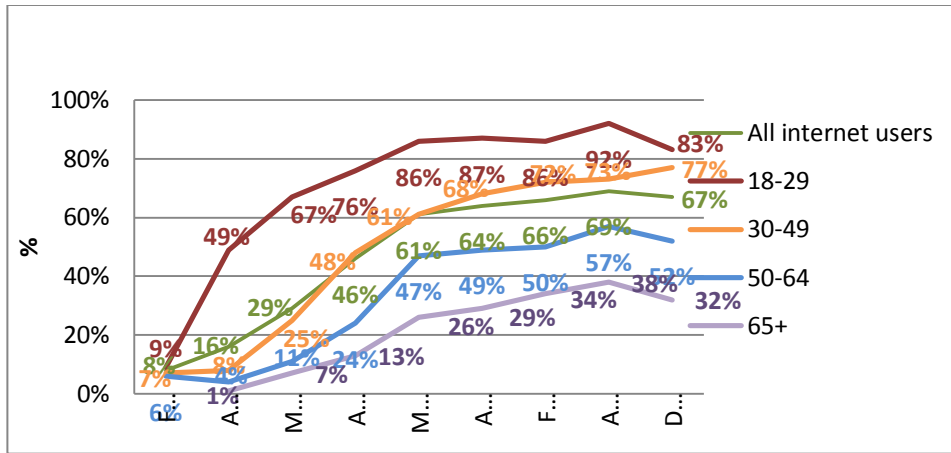


Figure 3: Usage of social networking sites by different age groups during 2005-2012

Motive behind this study is to use such important media; like Facebook for analysis of friendship patterns of students of same class. The other important motive is to study impacts of student's friendship on academic performance.

1.5 Major Contribution

As friendships occur due to some homophilic characteristics (McPherson, Miller, L., Smith, L., Cook, J. 2001) and these friendships must have some effects on the student's results and on life. (Ladd, G. W., Kochenderfer, B. J., Coleman, C. C. 1997) The primary objective of this study is to prove that student's friendships on social networking sites are also due to any homophilic characteristic, and then, finding the effects of friendships on one's result. Also, this analysis can give us the fact that whether student's academic performance is a cause of homophily.

To analyze friendships we have collected student data through Facebook application. For finding the relationships between friendships and grades, other type of data is collected from exam branch of School of Electrical Engineering and Computer Science (SEECs), NUST, Pakistan.

This study is one of the most important researches in the field of educational psychology and social networking. This study helps to gather real time, un-intentional data of targeted users which improve data gathering. Generally, user feel it a hectic task to fill manual forms and give interviews. So this study has provided us clear picture that we may use data from social networking sites for the purpose of behavior analysis. Contributions from this research can also help the teacher to better understand the psychology of students. Results are also helpful for students to improve their academic performance by managing their social life and making new ties in such interactive patterns that can be helpful in their studies.

1.6 Thesis Organization

Rest of this thesis is organized as follows:

In chapter 2 all important points related to this thesis and previous contributions, in the field of social networking, educational psychology and Facebook applications are discussed. In 3rd chapter, all important aspects related to methodology of this research and some results gathered from this research are discussed. Chapter 4 is all about results gathered from this study. Each result is discussed in details. Chapter 5 is end note and conclusion of this research.

CHAPTER 2

RESEARCH REVIEW

2.1. Why Social Networks Are Important?

Brief overview of social networks and homophily has been given in chapter 1. *Social networks* are formed due to linkage between people of groups or among groups. Social networks are important because they deal with connections or interactions due to different factors involved in our daily life. Human beings interact together within a community, for example, interaction with neighborhood.

Relationships and linkages are formed between two persons due to a same agenda or purpose. These linkages are called “*social ties*”. Relationship due to common agenda, purpose motive or habit is called a “*tie*”. Interactions of communities are further classified as “strong and weak ties”. (Easley, D., Kleinberg, J., 2010)

Strong ties” are interactive patterns within the community or group and “weak ties” are interactive patterns between two or more groups or communities.

One can say that strong ties show “*cohesion*” and weak ties depict “*coherence*” in relationships. (Frank K, A., 1995)

In simple words intra community relationship cause a “*strong tie*” and tie due to inter community relationship is “*weak tie*”. Concept of strong and weak tie is shown in fig. 4.

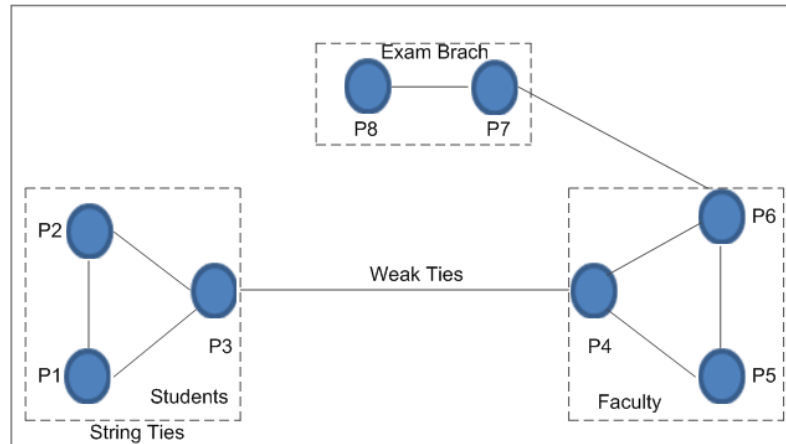


Figure 4: Strong ties and Weak ties

In sociology while studying social networks both strong and weak ties are important. In “strong ties” people belong to the same community are focused on some common agenda so they are strongly connected to each other. As strong interactions have influence on each other so individuals belong to a cohesive sub group can have some common properties and show similar behavior. (Cooley, 1909; Freud, 1952) Strong ties have interaction within the community. The weak ties bridge gap between two communities. One can learn the dominant properties of other community through weak ties. In general weak ties bring different communities close to each other. (Easley,D., Kleinberg,J., 2010) Due to social ties, subgroups can be mapped on the basis of cohesiveness of subgroups. In a social network cohesive subgroups can be measured by stochastic criterion that allowed interactive patterns in strong and weak ties. (Frank K,A.,1994)

Due to social interactions one can find a contact or known person in any part of world by passing only six hops, this theory is known as “*Six degree of Separation*”. It is an important phenomenon because it shown the importance of social interactions because interactions enables us to find contact in every area of world (Easley, D., Kleinberg, J., 2010)

Generally *Social networks* provide us information about, “Structure of community” and “Behavior of community”. *Structural study* of social networks can be done through the *Graph Theory*, and, “*behavioral study*” of social networks can be done by using the concepts of “*Game Theory*”. It is important to mention here that scope of this thesis covers both aspects; structure and behavior of social network i.e. “linkage and interactive patterns” of students of class.

2.2. Qualitative Measure of Human Behavior and Social Networks

“*People like to be friends with them who have similar properties*” is one of the most important studies to understand human behavior and qualities. Social networks can be analyzed for studying the similar people as they make strong ties with each other. (McPherson, et al.2001). This property of social networks is called “*Homophily*”, a reason of the formation of communities. For example, in any gathering smokers will try to sit in the separate place with other smokers. This forms a community, and smoking is the common property that brings them together.

For two people belonging to the same organization, living in same neighbor and having common hobby or interest can interact with other due to these shared activities and this increase the likelihood of a *strong tie* and finally turns into friendship.

2.2.1. Graph theory, Game Theory and Social Networks:

Structure and behavior of the community can be studied using Graph theory and Game theory respectively. Brief overview of graph theory and game theory is discussed below.

2.2.1.1. Graph theory:

Graph theory deals with modeling and visualization of any network in the form of graph or socio-gram. A “*graph*” is a combination of nodes and edges. Nodes of a graph represent participants of the network and edges represent the behavior or relation between those participants. Graph of a social network is also called a socio-gram.

Definitions about graph theory:

A “graph” is structural representation of object and their relation. Graph consists of set of nodes which represent objects and set of edges between nodes represent relationship between nodes.

A “path” is sequence of nodes in which each consecutive pair connected by an edge. For instance in fig.5 a path from node to node5 can be “node4, node5 and node6”.

A graph is connected if all the elements are connected, i.e. for every pair of nodes; there is a path between them. “Connected entities” of a graph form a component. If the components of graph have large number of edges and nodes and give a complex structure we called it as a giant component. Examples of some basic definition of graph theory are shown in fig. 5.

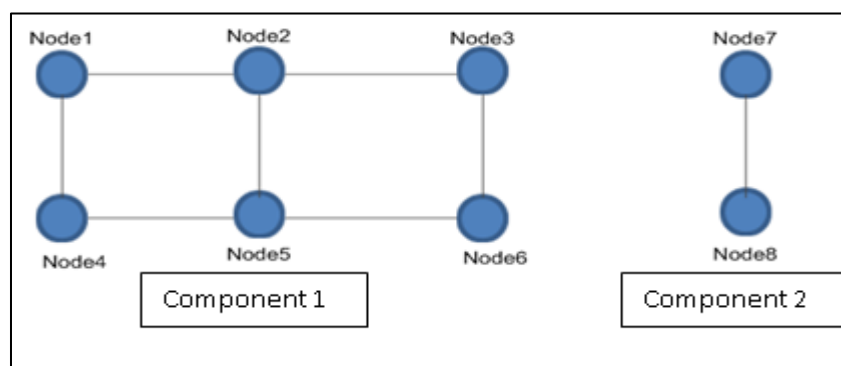


Figure 5: Graph representing some basic definitions of Graph theory

Why structural analysis?

Structural analysis is used to generate graphs of data. It is used to provide a graphical view of network. Structural view gives a clear representation of interaction between participants of any network. The different types of graphs generated during structural analysis of a data are “*collaboration graphs*”, “*who talks to whom graphs*”, “*Information linkage graphs*” etc.

“Who talks to whom graph” gives information about peer or friends. It shows that greater the friendships, stronger the links. Similarly; “information linkage graph” provides information about network, or about data. ‘Collaboration graphs’ give information about the quantitative measure of interactive behaviors of an individual in social network data.

2.2.1.2 Game theory:

“*Game Theory*” deals with the behavior of individual and groups while they interact with one another. One can conclude some results using game theory about a situation and can take decisions on the basis of such results.

Definitions about game theory

As discussed in previous section that Structural analysis gives us information about “*strong ties*” and “*weak ties*” of a social network. Concept of strong ties and weak ties arises from “*triadic closure*” principle and it is stated as:

“If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in future.” (Easley, D., Kleinberg, J., 2010)

Friend in common mean two people have friendship with same person. If two persons are friends of same person then there is high possibility that they both will become friend. Similarly;

a common enemy can cause relationships between two persons. The concept of triadic closure is explained in fig.6.

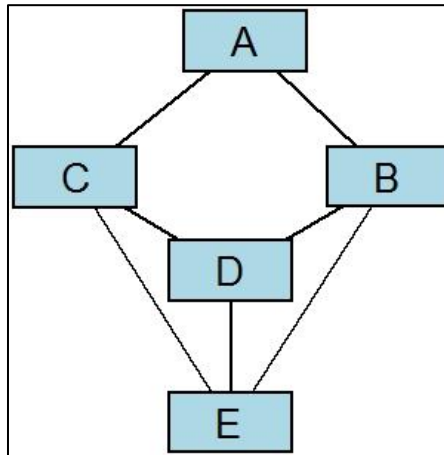


Figure 6: Friends in common and triadic Closure

Node C and node E are friends in fig.6. It is clear that node D is friend in common with node C and E. From behavioral perspective there is a high possibility that node C and node E may become friends in future. The basic role of triadic closure in social networks is to motivate the formulation of simple social network measures to capture its prevalence. One of these is “*clustering co-efficient*”.

“The clustering co-efficient of any node is the probability that two randomly selected friends of that node are friend with each other”.

Clustering Co-efficient ranges from 0 to 1. For example the clustering co-efficient of node D in above figure is $\frac{1}{2}$. Remember that triadic closure is very important property for the analysis of network structure and it has following reasons:

- It provide *degree of opportunity* to be a friend, if value of clustering co-efficient of two nodes is closer to one, it means there is high probability of friendships between them

- *Trust* is built due to common friend and interaction, greater the value of clustering coefficient more is the trust between friends
- *Incentive*, because common friend bring close to each other the non-friends, hence greater value of incentive means more chance to be close to each other.(Easley, D., Kleinberg, J., 2010)

Why behavioral Analysis?

Behavioral study in social networking has unique importance, because, it explains the specific behavior of an individual or group of individuals.

In any network neighborhood overlap depends upon its strength. The strength can provide an important perspective about online activities of any network. We can study the strength of ties on the social networking sites where people maintain the hundreds of friendships. We can find the strong ties on the basis of involvement of frequent contact and weak ties on the basis of rare engagements of a user with the other user in common activity.

2.3 Technological Advancement and Social Networks

In last two decades, due to rapid advancement in web technologies, social networking sites have great contribution in the area of social networks. In this era, a new world of social sites for example Facebook, Twitter, Hi5 etc. has emerged. Due to these websites, sources of interaction have been improved. These social networking sites have users of all ages and areas. Facebook is the most common and well-known social networking website. Among users of internet 97% use SNSs and 90% of them use Facebook. (Smith, S, D. et al. 2009)

In the Facebook (Cameron, M. et al. 2009) data strength of ties can be defined by three categories of links:

- A link represent mutual communication
- A link which show one way communication
- A link for maintaining the relationship but have passive engagement

A graphical overview of complex structure of social network formed by the data of this thesis is shown in fig.7.

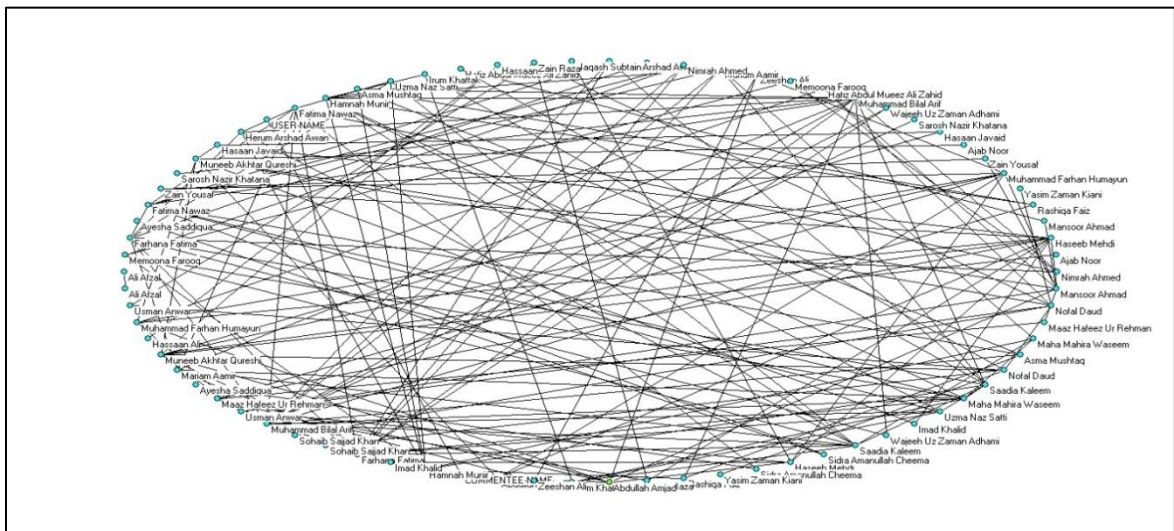


Figure 7: Socio-gram of class data collected through Facebook Application

Complex structure shown in fig.7 consists of all friendships, their maintained relationships and their mutual relationships. As strengths of ties depend upon the engagement of user; more the level of engagement of user with others, the more strength exists.

“Level of engagement” is how much a friend is busy on other’s wall by sending messages or pokes.

In social networking, “social capital” plays very important role and can be defined as:

“Social capital is an ability of actors to secure benefits by virtue of membership in social network”.

Social capital shows the role of different types of capitals for example:

- Physical Capital implements the technologies that are helpful to do work
- Human capital shows the skill and talent of people.
- Cultural capital shows the accumulated resources of culture

All above types of capitals represent the amount of energy and efforts that somebody put on the social network sites.

2.4 Homophily:

Homophily and Homogeneity are common and important phenomenon in human life. Homogeneity attracts persons who have similar properties, agenda, goal, motive or similar likings and disliking. This “foci” is called “*homophily*” and it is major cause of community formation. A study about homogeneity (Verbrugge, L.M.,1977) showed the amount of homogeneity that exists, in the different areas, this study was conducted to analyze the common thing that exist in the good friends and vice versa. Results show that strongly connected communities have something common such as age, gender, marital status, job, life motives, political views, ethnic background etc. It also discusses the reasons of the friendships and one important reason is “*opportunity to meet*”.

2.5 Overview of Educational Psychology:

“*Educational psychology*” is branch of psychology that deals with education, teachings, issues related to education and concerns of students. It deals with the study about the learning process, behavior of students, interaction between students and response of students towards the teacher. Educational psychology tells that, in a class, how friendships are established and how

they expand. It also discusses the consequences of friendships in class on overall academic performance and student's results. It enables us to understand the learning process. It also explains adolescences friendships and effects of such friendships on one's progress in class.

It is worth mentioning here that, in Psychology, learning from the environment is very important. This idea came from "*nurture theory*". A child learns the social behavior and norms from its environment in early childhood and in adolescence. Behaviors learned from the environment are generally social and ethical. Educational Psychology claims that these social behaviors have great impact on student's result.

One of the most important social interactions of a child during school age is with classmates. This interaction impacts student's result and extra-curricular activities. For instance, students who play cricket during leisure are more interactive toward the cricket player of their class. Hence, in this research, analysis of patterns of friendships and their impacts on student's results is lies most likely in the area of educational psychology.

2.5.1 Homophily and Constructivism:

As discussed earlier according to nurture theory, environment has impact on human's behaviors. In childhood the learning and cognitive process is very efficient. Youngers are more tend to learn from their environment. They perceive things more quickly as compared to the older. Generally, at young age, a man is more impressed from other's feeling and deeds without thinking of consequences.

As children's personality is more affected by surroundings and they are affiliated more with those who fascinate them.

In every class, students have their own ideology and way of understanding; as they belong to different social and economic backgrounds. This concept is called *constructivism*

(Brooks, J. G., Brooks, M. G., 1993). There can be similarities in the learning behavior of students due to some common homophilic characteristics, for example, goals, affiliations, belongingness to same place or speaking same language etc. Homophily is very important for understanding the types of relation within a network. In Educational Psychology, this property explains student's relation with others and provides the reasons of these relationships.

2.5.2 Homophily as a cause of community formation in a class

Age-mates have friendships with each other hence "age" is homophilic characteristic causing a community in class. (Gray, M. R., Steinberg, L., 1999)

Within class there can be a number of students belonging to different region, religion, gender, backgrounds and ethnic groups because they have different backgrounds so in a single class there can be different communities, it is important to note that according to age, class is a community, but according to the different social and economic background, there can be different communities within a class. Hence, community formation is a relative phenomenon depends upon the homophilic characteristic used for analysis. For example one can choose age, gender, religion, language, economic background or geographical background etc.

Friendship in a class has impact on grades of students, and popular students have good grades. The difference in grades is due to the different factors, such as, location and backgrounds. (Mihaly, K., 2009) In adolescence, influence of friends is important to build attitude and behavior; it also has *positive or negative* impact on physical and mental development. This process leads to "*intimacy*".

"Intimacy is affiliation of somebody with others and it arises due to homophily". (Berndt, T. J., 1992)

Trend, that, the students have one or two mutual friends tend to convert into the group, and, school performance of the students in peers have effect on academic progress, have been found. Students who are accepted by more peers or groups have more sense of belongingness and hold more friends, performed well in the scholastic tasks. The similar pattern exists in class contribution, that student having more relatedness participate actively in class activities. (Ladd, G. W. et al.1997)

2.6 Social Networking and student's Interaction:

Social networking gives information about relationships, their strengths and weaknesses. One can find this information from traditional surveying or by using social networking websites. Most of student use social networking sites to be in contact with their class fellows. Many students spent substantial amount of time on Facebook. According to a survey students spent 101.09 minute (Mean time) on the site per day actively or passively engaged on Facebook. (Junco, R.,2011) The most recent data collected by the “SDUCAUSE center for Applied Research (ECAR)” from a sample of 36950 students from 126 US universities and one Canadian university showed that the 90% of students who use social networking websites, 97% said that they use Facebook and 97% reported actively engaging on the site daily.(Smith S. D., et al. 2009)

Another traditional survey about Facebook reveals that students spent 27.93 min per days and 28.44 min at weekends. Friendship on Facebook depends upon motive of a person to use Facebook, and about 68.48% people have friendships with the people they know. (Pempek, T.,et al. 2009)

2.6.1 Causes of friendships on Social Networking Sites:

Friendships on social media such as Facebook can be due to friendships in real life, common interests, joining the same groups, visit the same pages, common likes and dislikes etc. One must keep in mind that adding a person as friend in profile friend list does not mean, a real friendships. Friendships depend upon the level of interactions and links. There can be more than hundred and even thousands of friends but one can determine about good friendship by analyzing messaging, poking, wall comments, and activities on one another's walls etc.

Facebook and all other social media are handy tools to analyze students' activities on the daily, weekly and seasonal basis. The activities of student show that the friendship pattern on the Facebook support both distant and geographically proximate relationships. The strong and weak temporal patterns can be found about student relationships. Groups between the students are formed on basis of similar temporal and seasonal variation and this can be analyzed by messages sent or pokes etc. Facebook users tend to connect with the geographically bounded and same class community. (Golder, S.D. Wilkinson , Huberman, B.A., 2007).

On a single campus, several kinds of capital can foster by the use of Facebook. Similarly, within the school the value of this social capital is high. About 41.6% messages are sent to friends on different schools while others on the same school. (Ellison, N.B., et al. 2007)

People suffering with the similar circumstances have more interaction and relationship among them. The similar circumstances; may be due to geographical relations of community, having common goal, or attending the same class etc. and this may be the primary reason that people have friendship belong to the similar sort of community. (Feld, Scott, L., 1981).

Above all, discussion shows that homophily is a cause of community formation whether it is on social networking site, in school or on any other forum. One can found distinct patterns in

social network data about community formation by using different techniques. It is also clear from previous discussion that different behavior patterns are formed, due to, belongingness from different communities or multiple communities are formed that are composed of such individuals who have same behavior. One can easily found community through social networking analysis techniques as discussed in next section.

2.7 Techniques for social network analysis

Several approaches for modeling and analyzing an individual behavior in social sciences exist. In psychological research, researchers must keep in mind about collection of right and useful data because collecting right data is the basic step which leads toward the correct and informative results. It is very difficult task, because users do not share the secrets of their life. Data collections through social networking sites give us information about the daily activities of users. From these activities we can deduce results and conclude some useful information about behavior of user. Several social networking techniques are used for the deduction of useful information from raw data set. Following are major techniques or different ways to analyze social networks data.

2.7.1 Community Detection and connected components

In social networking community detection is one of the most important tasks. All other analysis is based upon the formation of community. Community is defined as:

“*Group of nodes that have interactions among them*”. (Girvan, M., Newman, M. 2002; Radicchi, F., et al. 2004)

Community finding in social networking data can be done on the basis of following:

- 1) Link based
- 2) Link and content based
- 3) Interaction based

Link based community detection is based on graph theory. Component of graphs and graph partitioning are treated as a community in link based techniques. (Evans, T., Lambiotte, R.2009)

A “*link or line graph*” is constructed from the large original graph. In line graph every vertex is connected to an original edge. Link in graph represents the adjacency between two edges in the original graph. As it is a memory inefficient technique hence it is not suitable for large social networks but in our research data set is limited and it is a suitable technique for our study. A link or line graph technique is also called Edge-Clustering (Tang, L., Liu, H. 2009) in this case edges are taken as center of the graph. Edges are considered as instances and nodes as features. One can find highly overlapping communities. Other methods to obtain communities include soft clustering (Newman, M. E., Leicht, E. 2007) and probabilistic models (Erosheva, E., et al.2004).

In content based analysis (Blei, D.M., et al.2003) content of the messages, shares etc. is analyzed for community detection This technique is known as generative modeling. In this method fixed types of communities are predefined and data is analyzed on the basis of predefined set of communities. These techniques analyze the content and put a user into a particular community on the basis of words. But problem with these models are that in these models irrelevant key words are also analyzed and can cause community in network model. Other methods used to find the community on the basis of content is probabilistic models (Yang,T., et

al.2009) This is an improved form of generative models and these models first build a conditional model for estimation of the probability of connection of different node. After finding probability of connection, membership of nodes of a community is modeled on the basis of content. Then both two models are unified by community memberships. Scope of this research does not cover the content based community detection because we have only counted the number of comments, likes and number of posts on other walls etc.

All users on the social media sites have different types of interactions. As more interaction between the users mean more closeness or strong ties, hence interaction based methods for community detection are very important for finding the community in social networks data. Interaction Based methods for community detection extend further to co-clustering framework.(Wang, X. et al.2010) Co-clustering framework finds the community on the basis of user activities with other especially tagging behavior. This model shows the more fine and accurate community on the basis of tagging information.

“*Meta-Graph Factorization*” (Meta-Fac) is also based on the interaction based community detection algorithms (Lin,J., et al.2009) It finds community structures on the basis of various interactions.

Heterogeneous interactions are also used for the community detection by integrating the information community. (Tang, L., et al. 2009].

In this research we have found communities on the basis of interactions. We have simply assumed that more the interaction stronger the tie, closer the relation and lead to friendship. More the interactions in community shows strong community in the structural model of interaction based community detection algorithm.

A joint optimization framework for the integration of links and multiple data sources for community detection explain the multi-source community detection. (J. Tang, X. Wang, and H. Liu, 2012)

As discussed in section 2.2.1 that set of nodes connected by a path within a network, called component and if there is no path or edge between nodes then a separate component is formed. Connectivity of the network can be predicted by finding the number of connected components; smaller the number of connected components stronger the connectivity of network.

2.7.2 Degree distributions and stress distributions

In social networks data, the degree of a node is the number of edges connected to the node and self-loop of any node has double the degree value (Diestel, R, 2005). The degree distribution of any node is the number of nodes having degree k where $k = 0, 1, \dots$

In directed graph, the “*in-degree*” is defined as the total number of incoming edges and the “*out-degree*” is defined as the total number of edges from a node. Random and scale free network topologies are distinguished by using the node degree distribution. (Erdős, P., Renyi, A .1960; Bollobas, B.2001)

Stress distribution

Path or edge lengths in social network socio-gram indicate the closeness and level of interaction among users. Shorter the path more close the users are. The stress of a node is measure of the number of shortest paths passing through node. More the number of short distant edges implies the higher the stress on node. (Brandes, U , 2001 ; Shimmel, A. 1953)

2.7.3 Neighborhood of nodes and connectivity

The neighborhood connectivity of any node is defined as, the set of nodes connected directly with the node as neighbor. Average connectivity of nodes is found by average number of

neighbors indicate that higher the average value higher the network density. The density is a value between 0 and 1. It also explains the dense populated networks with edges. In this case self-loops are ignored. A network with isolated and without any edges have density of 0 and a clique or combination of dense connected nodes has density 1.

2.7.3.1 Network centralization

In this case central node has assigned the value 1 Networks whose topologies resemble a star assigned a centralization value close to 1, and decentralized networks area assigned a centralization value close to 0. (Dong, J., Horvath, S, 2007) In this thesis network centralization is applied on each node and discussed with details in chapter 4.

2.7.3.2 Neighborhood connectivity

Number of nodes connected with neighbor nodes defined the connectivity of nodes. Average connectivity of nodes with other nodes is called neighborhood connectivity. (Maslov, S., Sneppen, K , 2002) It gives the average of the neighborhood connectivity of all nodes. Neighborhood connectivity of nodes have following types:

only in – The average out-connectivity of all the nodes that are attached with it.

only out - The average in-connectivity of all the nodes that are attached with it

in and out - The average connectivity of all node that are in neighborhood of the node.

2.8 Conclusion

This chapter includes the overall background of this thesis, introduction and related details of social networks, important features of educational psychology and sociometric techniques that are used community detection analysis.

CHAPTER 3

EFFECTS OF HOMOPHILY ON ACADEMIC PERFORMANCE

This chapter includes the descriptions of methodologies use in this research. As discueed earlier that social networking sites provide an interfaces through which one can collect user data by permission for the analysis of their activities. This data can predict the behavior of user in a community. For the purpose of data collection, the social networking sites such as Facebook, provide application development kits and app engines which helps in development of applications and interfaces.

More specifically regarding our study, we have collected data through social networking site and used it for the analysis of student behavior in class. This research targeted the students of same class for finding communities within the class and friendships patterns, through Facebook data. The most important feature of this study is to explain the impacts of class friendships on student's grades. Friendship matrix of users and commentee is shown in fig.8. It provides the information about total number of comments of any user in the class on the others.

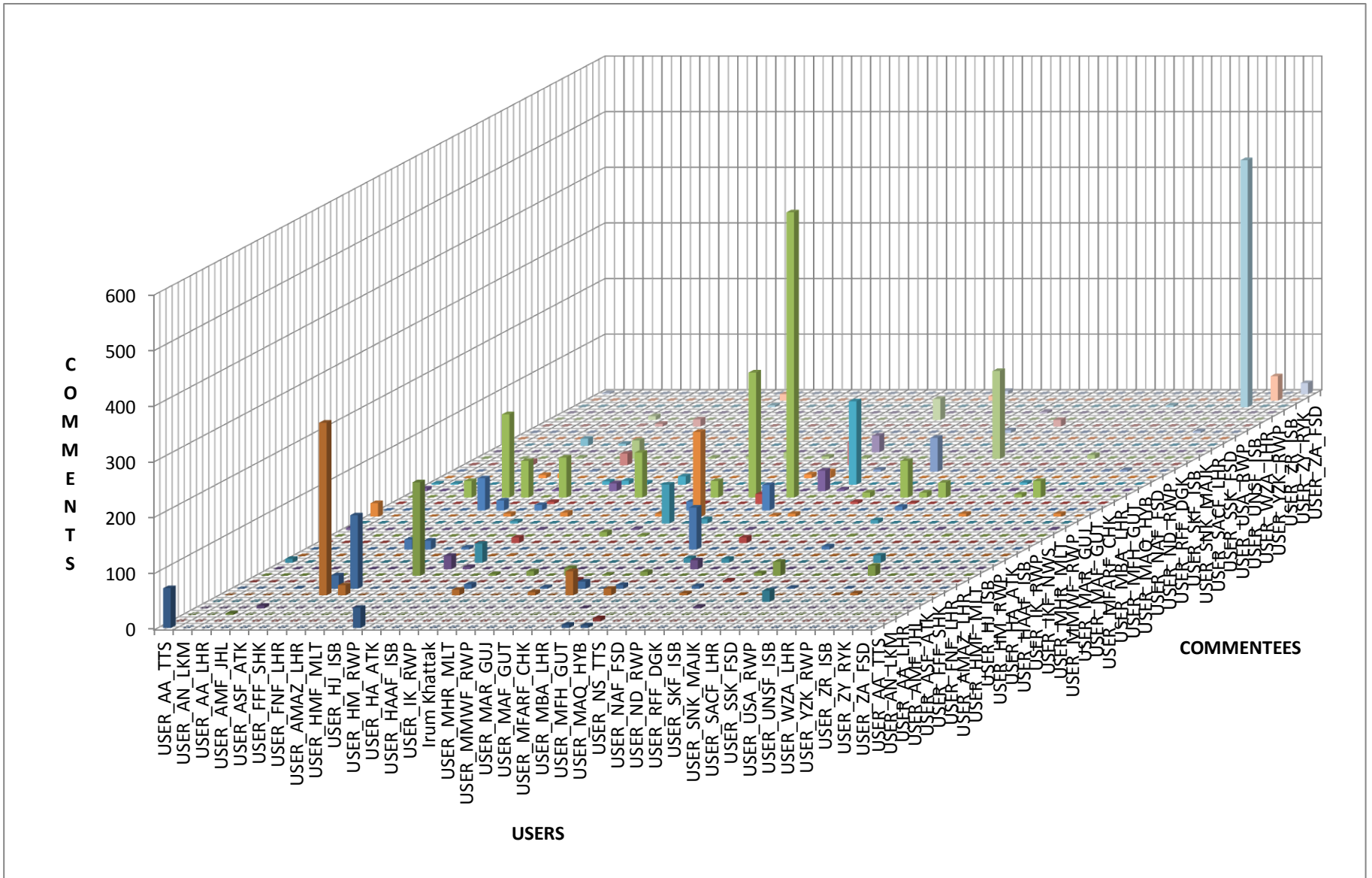


Figure 8: Friendship matrix graph of users and commentee in class

3.1 Methodology and results

Data analysis techniques, used for social networks data have already been discussed in chapter 2 section 2.7. We have found the homophily in class using such techniques. For this purpose we found the clusters in class on the basis of interactions between students.

All the steps taken for this research is shown in fig.9.

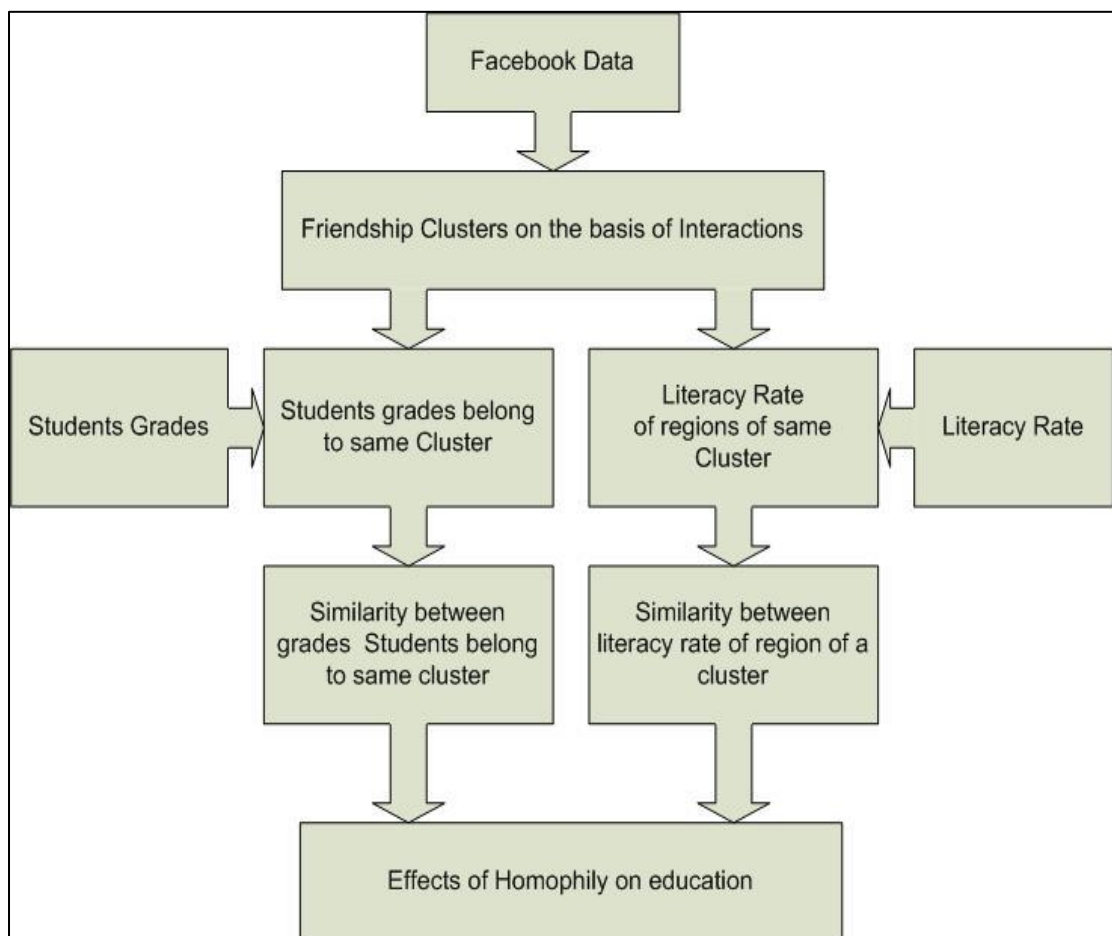


Figure 9: Block diagram of methodology

For finding the effects of homophily on education, first of all community and friendships patterns are found from social network data. We have found friendship patterns and named them as clusters. After finding the clusters educational results and literacy rates of each cluster are matched using chi- square tests.

As we have compared the literacy rate of the regions with student's grades. Literacy rate of each region is taken from the "Federal ministry of education, Pakistan" and Provincial ministries of education". The similarity between the literacy rates of student's belonging regions are found of each cluster. We have discussed the literacy rate of regions and student grades and their friendship level separately one by one.

As we have categorical data in both cases so we have applied chi-square test for the co-relation or finding the similarities between the grades of students belongs to same cluster. Similarly, as having similar literacy rate is also a hemophilic characteristic. So we have hypothesized the students belong to same cluster must have similar or near to similar literacy rate.

3.1.1 Data collection through Social Network Site:

Several traditional methods such as form filling by interviewee or conducting oral interviews are used to gather data for surveying of any focused area. These methods may not be effective because it is difficult to motivate a person or community to fill the forms and provide proper information. Also, there is possibility to report fake data by users during oral interview.

Methods for Social networking data collection are different from the traditional techniques. Social networking sites have facilitated the researchers for data collection using some applications, by the permission of users.

For friendship analysis, we have collected student's data from Facebook. It is because all the targeted students are regular Facebook user. For the purpose of data collection we have developed a Facebook application. The collected data consist of User-ID, User-Name, User-Friend-list, number of likes, comments and posts on wall of other users. When targeted students click on Facebook application, their data was stored in text file under the heading of User_ID. All the users have unique value of User_ID which differentiate a user from others.

3.1.2 Social networks data analysis

For data analysis, different network parameters and techniques are used as already discussed in chapter 2. In this section we shall briefly discuss the results gathered by using these techniques.

3.1.2.1 Components and community Detection:

Collected data through Facebook application is analyzed for finding interactions of students with one another. By using community detection techniques as discussed in section 2.7, we have found the communities within the class. In this study the foremost information given by social networks is about the number of components. Structural or graphical representation of class data shows that there are three components. First component consists of 48.71% students of class and all of them are male. Second component consists of 48.71% students of class. This component consists of 15.78% male students and all others 89.47% are female. One node among these is totally isolated from others and has no interaction with others. The isolated node has shown an anomalous behavior while interacting with others. All three components along with the percentage of male and female in each component are shown in fig.10.

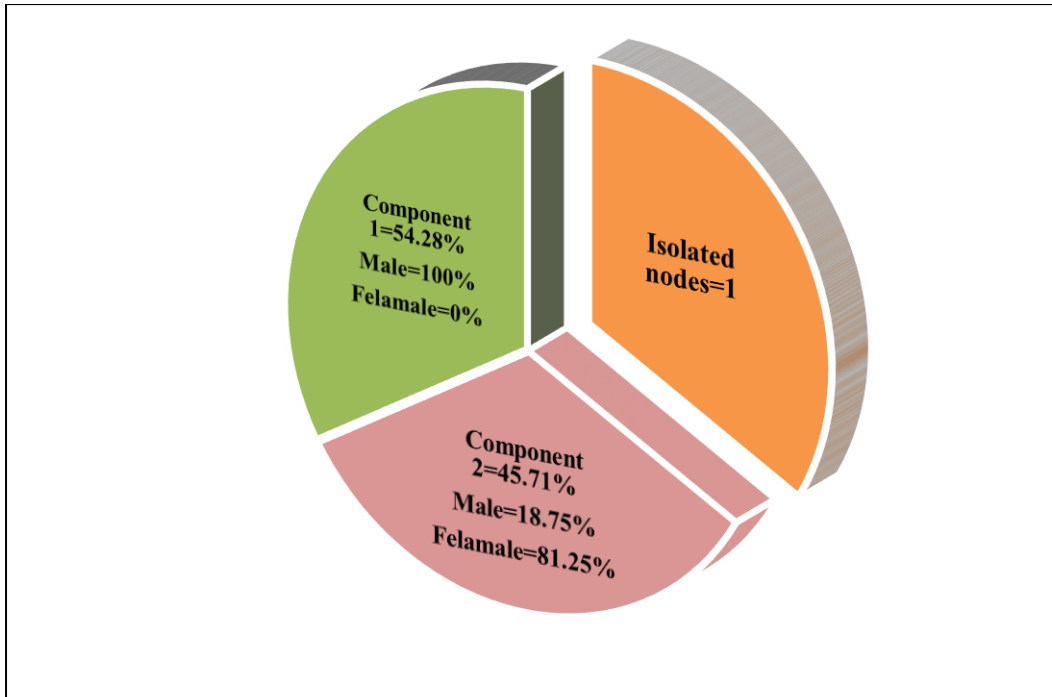


Figure 10: Components of class with their percentages

It is crystal clear from all above discussion that first and second component are formed in class on the basis of gender, implies that gender is a homophilic characteristic that devise a class into components. Proceeding further, among these components communities within the community are found by community detection approach based on hypothesis i.e.

“More interactive users with each other form strong ties and less interactive users are considered as weak ties”.

“Socio-gram” or graphical view of social networks gathered through Facebook application has been shown in fig.11. This figure shows the graphical view of three components of class and interactions among users. More interactions imply strongly connected users; hence this socio-gram shows relationships among users and strengths of relationships. The nodes of socio-gram show the users and the edges depict relationships among them. Numbers of edges between nodes correspond to number of comments, likes and shares that are exchanged between two users. Detailed analysis of each user of these components will be discussed in chapter 4.

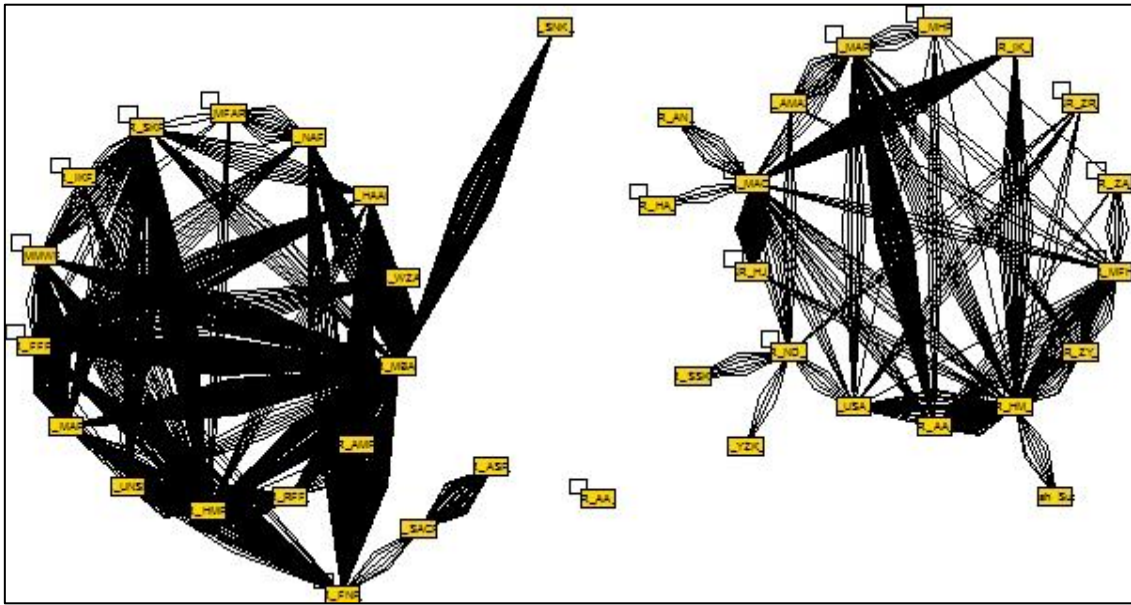


Figure 11: Components of class community and their interactions

3.1.3 Community and exam data

Social networking data has provided us the interaction level of students on the basis of number of comments. We have divided the interactive users in separate components. Grades of all these users are compared with each other. We have found the similarity between the comments and grades.

3.1.4 Hypothesis

This research study based on the basis of following hypothesis.

H0: Students having hemophilic characteristic having similar grades.

H1: The region from which student's belongs to have similar literacy rates.

In our study either we accept the hypothesis or reject the hypothesis on the basis of results of Chi square tests.

We have defined the primary and secondary variables. In our case the primary variable is "hemophilic property" i.e. interactions between the students. The secondary variables are "student's grades". In the second case we have used the "literacy rate" of

student's belonging district as a primary variable and interaction as a secondary. This is because the literacy rate may be one of the causes for the friendships among students and the similarity or dis similarity of grades is dependent upon the interactions among the students.

3.1.5 Similarity between grades of students belonging to same cluster

Social network data and exam data is analyzed on the basis of grades and comment. Relationships of student grades who are good friends with each other is measured and discussed separately.

Because of dealt with the categorical data from the Facebook and from the exam branch in the form of CGPA we have applied the non-parametric test for the co-relation between the categorical data. For this purpose we have applied the Chi Square test for the goodness of fit. The selection of Chi Square is due the fact that it analyzes the relationships between the two categorical data and chi square test found the co-relation or similarity between the categorical values on the basis of similarity and dissimilarity.

For chi square analysis we have devise the hypothesis H_0, H_1 , and then critical significance level "alpha". In our case we have selected $\alpha=.05$ which is 5 percent.

After this we have applied the chi- square test and then degree of freedom is found by total number of categories minus 1.

3.1.5.1 Friendships of each student with the literacy rate of regions

As literacy rate can also a hemophilic characteristic so we can hypothesized that belonging to the same city or region that have similar literacy rate can cause the friendships among students of same class and can affect the grades of students. So we have taken the literacy rate of student's region from the website of ministry of education and have co-related the grades and literacy rate of students who belongs to same region.

3.2 Conclusions

In all above discussion we have defined the methodology for this study. We have found the clusters of the friends of same class students. After that we have defined that these clusters would be considered as a group of friends. We shall find the similarity between the grades of students and literacy rates of the student's belonging regions.

We have of conclude that homophily exist in class as male students are friends with male and vice versa. In the regions of South Asia, friendships among the Males and Females are generally not in practice. In an Islamic country like Pakistan it is more significant because the inter mingling of the both gender is not in practice due to the religious norms. Hence we can claim this behavior as normal and also we can claim that the output from Facebook data is valid and aligned with the social norms of Pakistan. The similarity between the grades of students and literacy rates would be discussed in chapter 4.

CHAPTER 4

RESULTS AND DISCUSSIONS

We have already seen a brief overview of effects of homophily on education in chapter 3. In this chapter we shall discuss the findings of this research in details. As we have already discussed that “Homophily” can exist in different forms in a community due to different reasons for example students of same class can be divided into different groups on the basis of region, gender or some other hemophilic characteristics. In this study, homophily is found in dataset of students of one class. After finding homophily and friendship patterns, student’s results are analyzed and an effect of homophily on grades has been found. This chapter focuses on all the finding of this research.

4.1 Introduction

Scope of our study covers the three major types of data. First kind of data is directly collected through Facebook application by requesting the students of focused class to like the Facebook application. Second kind of data was taken from exam branch of School of electrical engineering and computer science and the third kind of data is taken from the website of ministry of education Pakistan. All these types of data are very concrete and are without any anomaly.

Some initial results of this research have already been discussed in chapter 3. In this chapter we shall analyze each and every individual one by one. We shall see the interactions of every user in class; then grades of all students that interact with each other has been analyzed.

Using the Facebook data we have found friendships on the basis of interactions or “Who Talks to Whom”. We assumed that, more number of interactions on social media sites closer the relationships are. In this chapter we have found belongingness and level of

belongingness of every student of class. The homophilic characteristics among users are considered as:

1. Students belong to same class.
2. Belong to same area or community.
3. Same gender.

These classifications produce some pattern of friendship by forming clusters. The students grades belong to one clusters are matched with each other.

Another hemophilic characteristic is defined on the basis of area or region from where an individual belongs. It is observed that the students belong to same cluster have matching with the literacy rate of their belonging regions

Now we discuss all results in details in next section. First of all we will discuss all the comments on a particular user.

4.2 Friendship of each user and similarity between grades and literacy rates

This section covers the analysis of student data of same class. We have calculated all comments of users on other users. In our data set total number of values is“4435” after data cleaning.

Now we have analyzed all the clusters of class one by one. The cluster is made on the basis of network centrality. Each user use is dealt separately by selecting him arbitrarily and finding his friends. Selected users and all other users who are connected with chosen user are considered as a cluster. Total number of comments of each user in any cluster is found by counting the degree of the vertex in sociogram of the cluster. For instance in CLUSTER-1 shown in table, USER_AA_TTS has degree 40 and USER_HM_RWP has 36 and so on. We have found the matching value between the grades of students belong to cluster-1 by using chi square test. Also we have found the matching value of literacy rate of the belonging regions of the cluster-1 by using chi square test.

Users of cluster-1 are shown in table 1.

NODES	DEGREE	CGPA	CITY	LITERACY RATE
USER_AA_TTS	40	2.68	TOBATEK SINGH	46.97
USER_HM_RWP	36	2.89	CITY_2	67.50
USER_MAQ_HYB	5	3.13	CITY_7	42.15
USER_MFH_GUT	6	2.86	CITY_9	57.87
USER_MHR_MLT	4	2.91	CITY_4	41.66
USER_ZA_FSD	2	2.12	CITY_5	51.94
USER_MAR_GUJ	25	2.66	CITY_8	53.67

Table 1: Users of Cluster-1

The bar graph of users of cluster-1 is shown in figure-12.

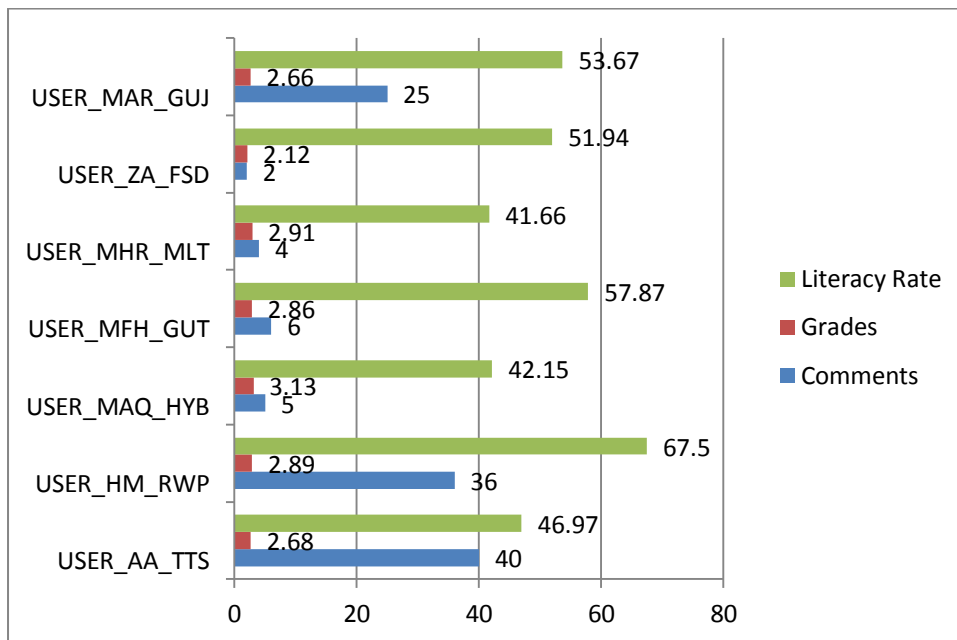


Figure 12: Cluster-1 Users Degree, CGPS and literacy rate

The participants showed the significant preferences of grades among seven users are $X^2(6,n=7)= 0.482, p<.05$. Hence H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among seven users are $X^2(6,n=7)= 0.136, p<.05$. Hence H_1 is valid and not to be rejected.

The students that belong to cluster 2 are shown in table 2.

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_AMF_JHL	12	3.27	CITY_10	60.00
USER_RFF_DGK	11	2.23	CITY_6	28.71
USER_MBA_LHR	31	3.06	CITY_3	62.98
USER_HMF_MLT	2	3.23	CITY_4	41.66
USER_FNF_LHR	1	2.83	CITY_3	62.98

Table 2: Users of Cluster-2

The bar graph of users of cluster-2 is shown in figure-13.

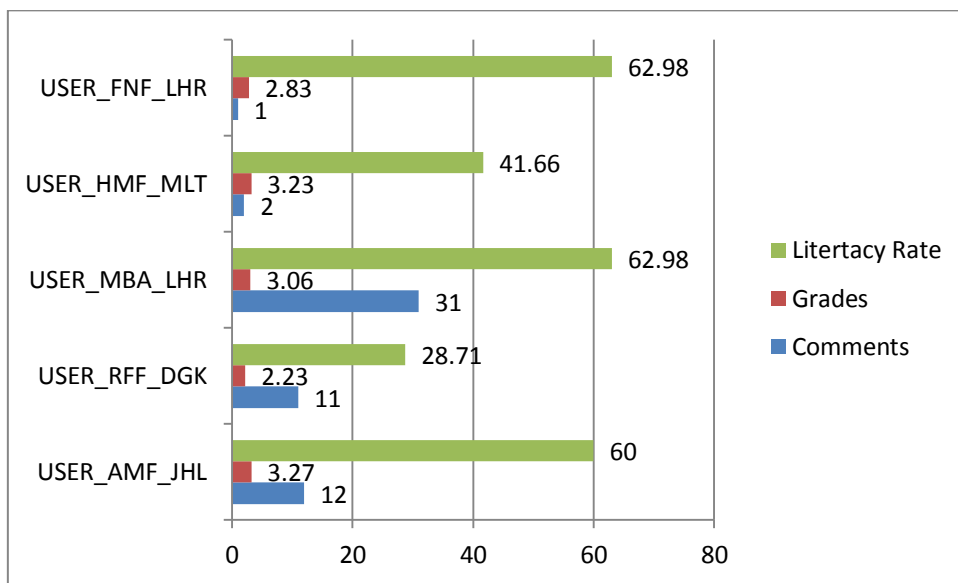


Figure 13: Cluster-2 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among five users is $X^2(4,n=5)= 0.190, p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among five users is $X^2(4,n=5)= 0.000913, p<.05$. Hence H_1 is valid and not to be rejected.

The students that belong to cluster 3 are shown in table 3

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_FFF_SHK	309	2.45	CITY_11	40.72
USER_NAF_FSD	10	2.27	CITY_5	51.94
USER_MBA_LHR	164	3.06	CITY_3	62.98
USER_WZA_LHR	11	2.68	CITY_5	51.94
USER_HAAF_ISB	28	2.27	CITY_1	70.20
USER_MAF_GUT	101	3.18	CITY_9	57.87
USER_SKF_ISB	3	3.29	CITY_1	70.20
USER_FNF_LHR	42	2.83	CITY_3	62.98
USER_MMWF_RWP	6	2.95	CITY_2	67.50
USER_UNSF_ISB	2	3.04	CITY_1	70.20

Table 3: Users of Cluster-3

The bar graph of users of cluster-3 is shown in figure-14.

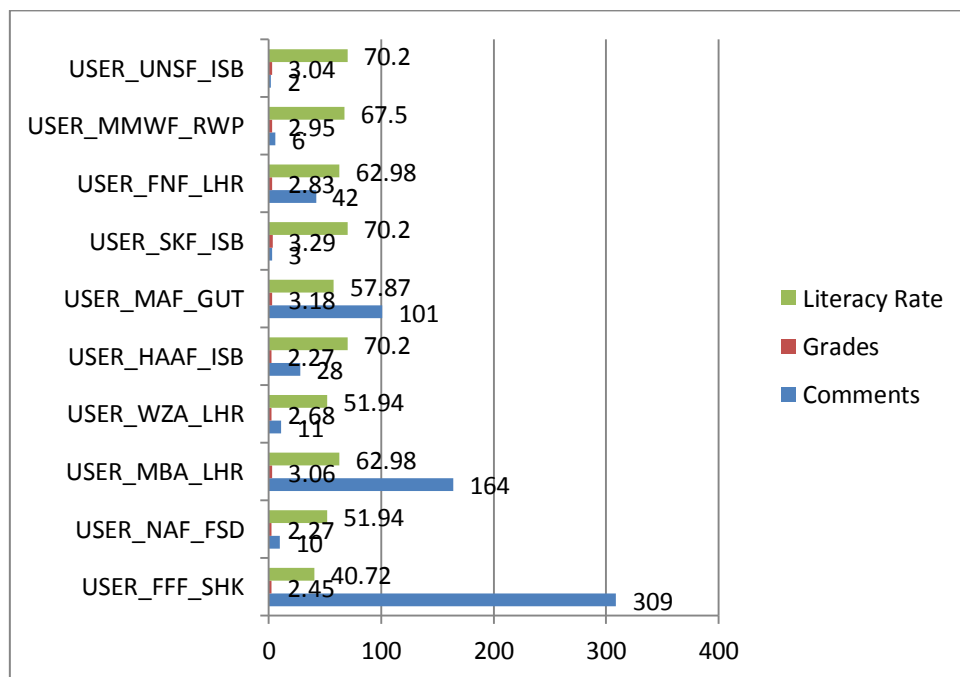


Figure 14: Cluster-3 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(9, n=10) = 0.304126$, $p < .05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(9, n=10) = 0.103201$, $p < .05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 4 are shown in table 4

NIDE	DEGREE	CGPA	CITY	LITERRACY RATE
USER_FNF_LHR	131	2.83	CITY_3	62.98
USER_MMWF_RWP	6	2.95	CITY_2	67.50
USER_UNSF_ISB	6	3.04	CITY_1	70.20
USER_MAF_GUT	35	3.18	CITY_9	57.87
USER_HAAF_ISB	24	2.27	CITY_1	55.72
USER_SACF_LHR	6	2.07	CITY_1	70.20
USER_NAF_FSD	13	2.27	CITY_5	51.94
USER_MBA_LHR	73	3.06	CITY_3	62.98

Table 5: Users of Cluster-4

The bar graph of users of cluster-4 is shown in figure-15.

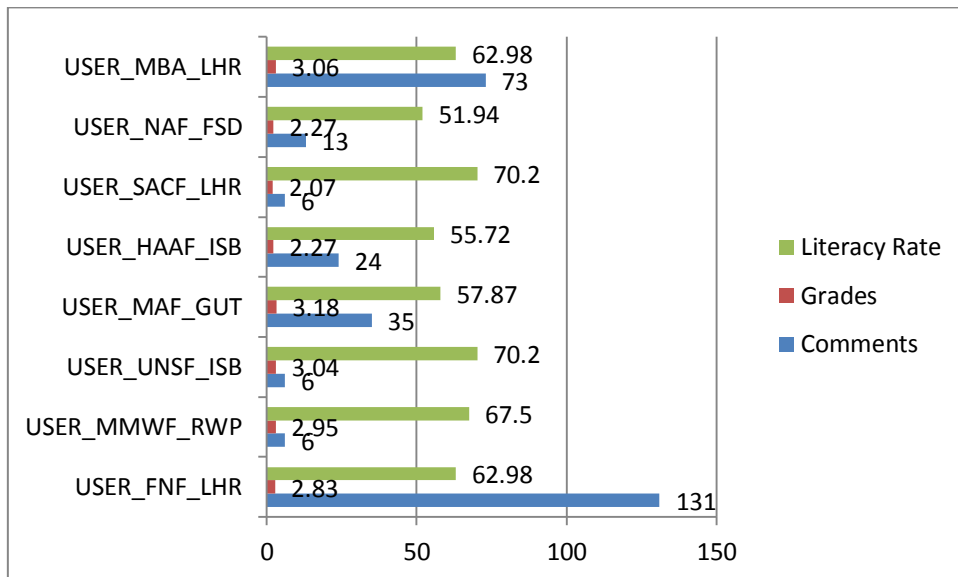


Figure 15: Cluster-4 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(7,n=8)=0.1115$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(7,n=8)=0.6312$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 5 are shown in table 5

NAME	COMMENTS	CGPA	CITY	LITERACY RATE
USER_AMAZ_LHR	5	2.34	CITY_3	62.98
USER_ND_RWP	4	3.56	CITY_2	67.50
USER_MAR_GUJ	11	2.66	CITY_8	53.67
USER_MFH_GUT	2	2.86	CITY_9	57.87

Table 6: Users of Cluster-5

The bar graph of users of cluster-5 is shown in figure-16.

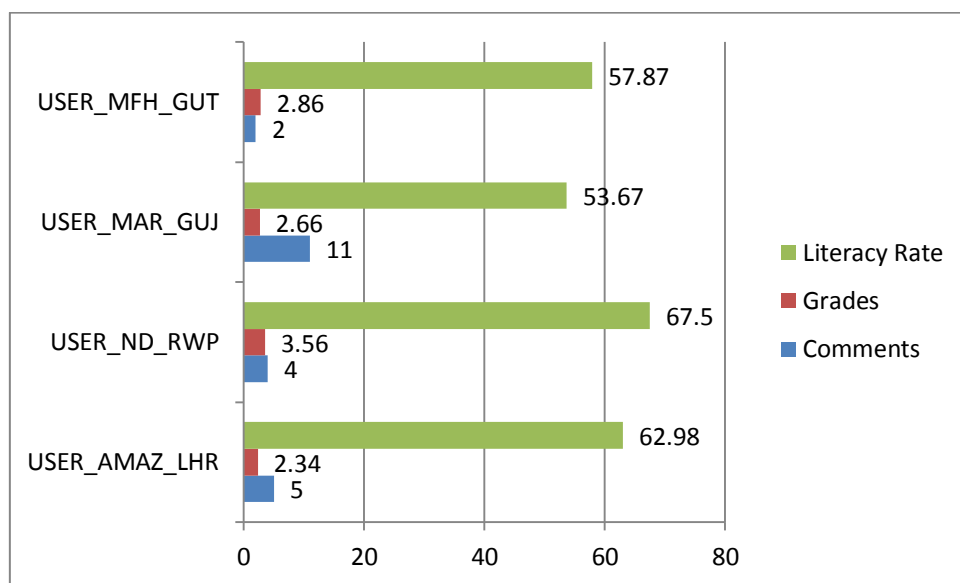


Figure 16: Cluster-5 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(3,n=4)=0.071647$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(3,n=4)=0.615596$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 6 are shown in table 6

NAME	COMMENTS	CGPA	CITY	LITERACY RATE
USER_HMF_MLT	167	3.23	CITY_4	41.66
USER_UNSF_ISB	31	3.04	CITY_1	70.20
USER_MMWF_MLT	18	2.95	CITY_2	67.50
USER_IKF_NWS	8	3.19	CITY_13	39.32
USER_SKF_ISB	63	2.29	CITY_1	70.20
USER_MFARF_CHK	8	3.71	CITY_12	56.72
USER_HAAF_ISB	7	2.27	CITY_1	70.20
USER_MAF_GUT	12	3.18	CITY_9	57.87
USER_RFF_DGK	26	2.23	CITY_6	28.71
USER_MBA_LHR	79	3.06	CITY_3	62.98

Table 7: Users of Cluster-6

The bar graph of users of cluster-6 is shown in figure-16.

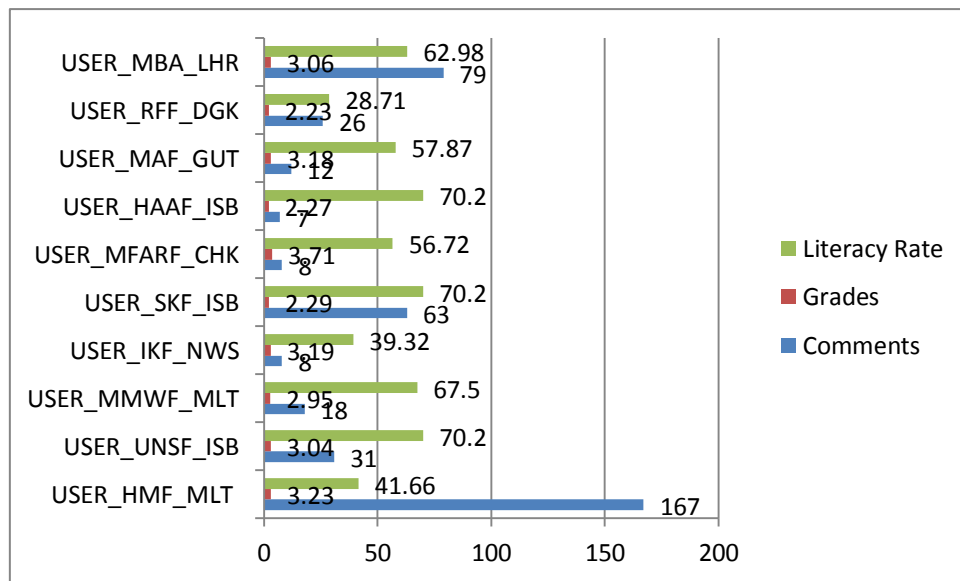


Figure 17: Cluster-6 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(9,n=10)=0.027753$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(9,n=10)=0.00004569$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 7 are shown in table-7

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_HJ_ISB	24	2.65	CITY_1	70.20
USER_HM_RWP	4	2.89	CITY_2	67.50
USER_MAQ_HYB	22	3.13	CITY_7	42.15

Table 8: Users of Cluster-7

The bar graph of users of cluster-7 is shown in figure-18.

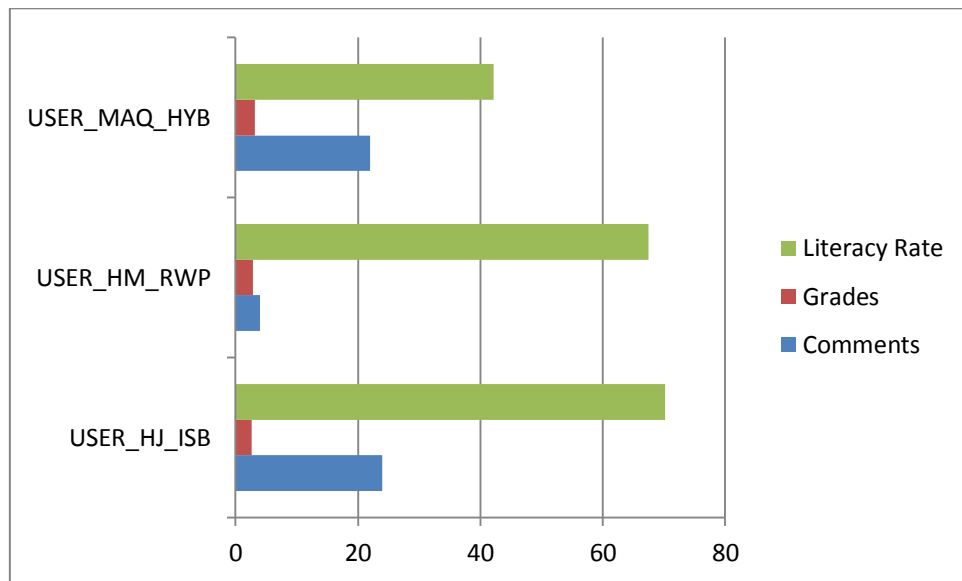


Figure 18: Cluster-7 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(2,n=3)= 0.60758, p<.05$. Shows that the H0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(2,n=3)=0.018422, p<.05$. Shows that H1 is valid and not to be rejected.

The students that belong to cluster 8 are shown in table 7

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_HM_RWP	34	2.89	CITY_2	67.50
USER_IK_RWP	11	2.72	CITY_2	67.50
USER_NS_TTS	7	2.16	CITY_14	46.95
USER_USA_RWP	13	2.88	CITY_2	67.50
USER_MFH_GUT	22	2.86	CITY_9	57.87
USER_MHR_MLT	10	2.91	CITY_4	41.66
USER_MAQ_HYB	7	3.13	CITY_7	42.15
USER_ZR_ISB	13	1.78	CITY_1	70.20
USER_ZY_RYK	12	3.31	CITY_15	30.93
USER_MAR_GUJ	8	1.78	CITY_8	53.67

Table 9: Users of Cluster-8

The bar graph of users of cluster-8 is shown in figure-19.

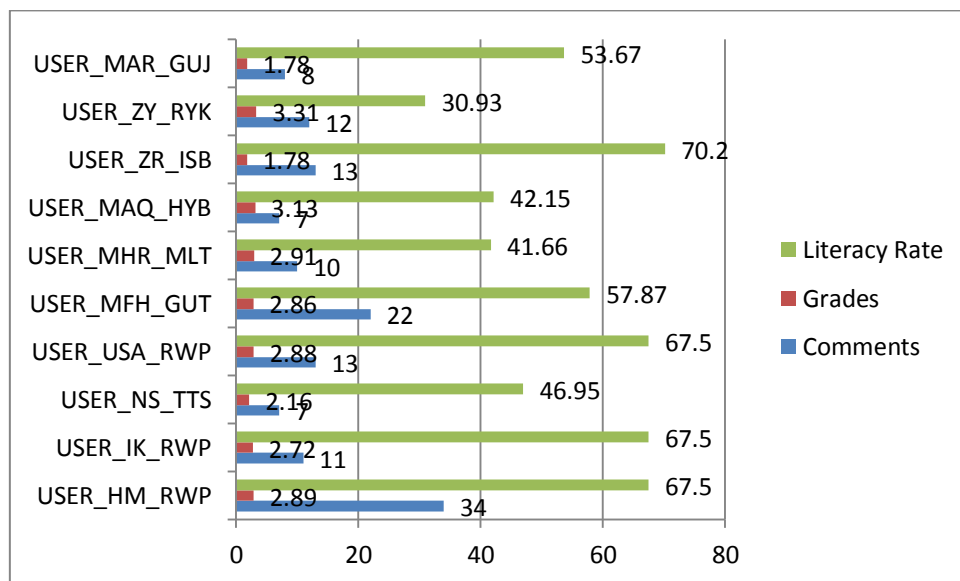


Figure 19: Cluster-8 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(9,n=10)=0.001564$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(9,n=10)=0.00029$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 9 are shown in table-9

NAME	COMMENTS	CGPA	CITY	LITERACY RATE
USER_IKF_NWS	8	3.19	CITY_13	39.32
USER_UNSF_ISB	1	3.04	CITY_1	70.20
USER_MMWF_RWP	2	2.95	CITY_2	67.50
USER_SKF_ISB	5	3.29	CITY_1	70.20

Table 10: Users of Cluster-7

The bar graph of users of cluster-9 is shown in figure-20.

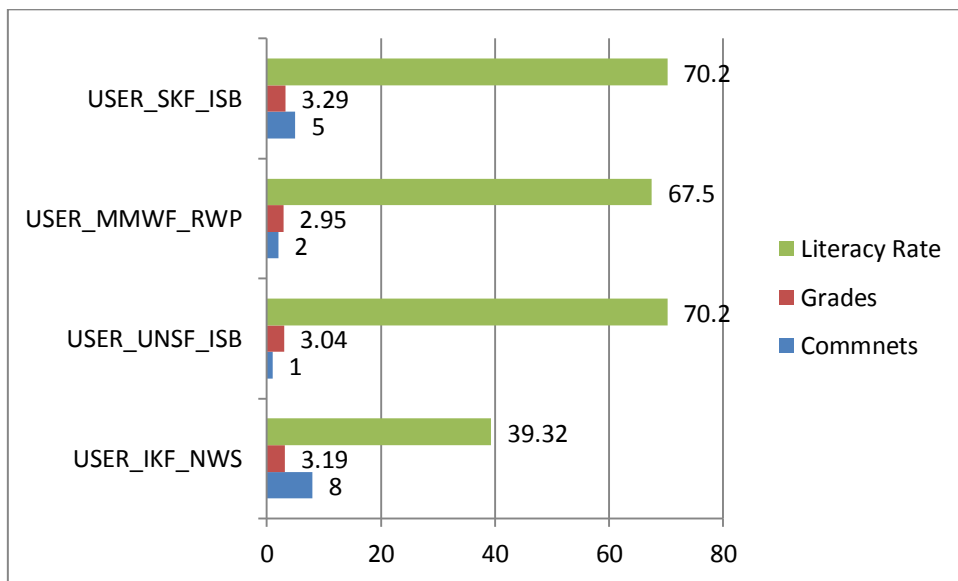


Figure 20: Cluster-9 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(3,n=4)=0.909922$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(3,n=4)=0.011804$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 10 are shown in table-10

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_MHR_MLT	2	2.91	CITY_4	41.66
USER_MFH_GUT	1	2.86	CITY_9	57.87
USER_USA_RWP	1	2.88	CITY_2	67.50
USER_ZA_FSD	1	2.12	CITY_5	51.94
USER_MAR_GUJ	6	2.66	CITY_8	53.67

Table 11: Users of Cluster-10

The bar graph of users of cluster-10 is shown in figure-21.

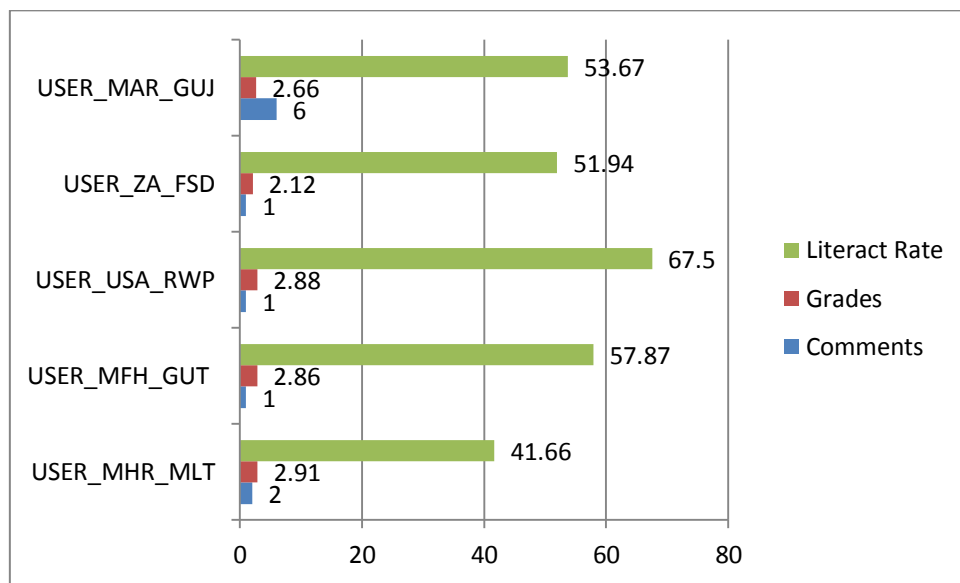


Figure 21: Cluster-10 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(4,n=5)= 0.394344$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(4,n=5)= 0.167082$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 11 are shown in table-11

NAME	COMMENTS	CGPA	CITY	LITERACY RATE
USER_MMWF_RWP	70	2.95	CITY_9	57.87
USER_SKF_ISB	18	3.29	CITY_1	70.20
USER_MAF_GUT	24	3.18	CITY_9	57.87
USER_MFARF_CHK	2	3.71	CITY_12	56.72
USER_MBA_LHR	30	3.06	CITY_3	62.98

Table 12: Users of Cluster-11

The bar graph of users of cluster-11 is shown in figure-122.

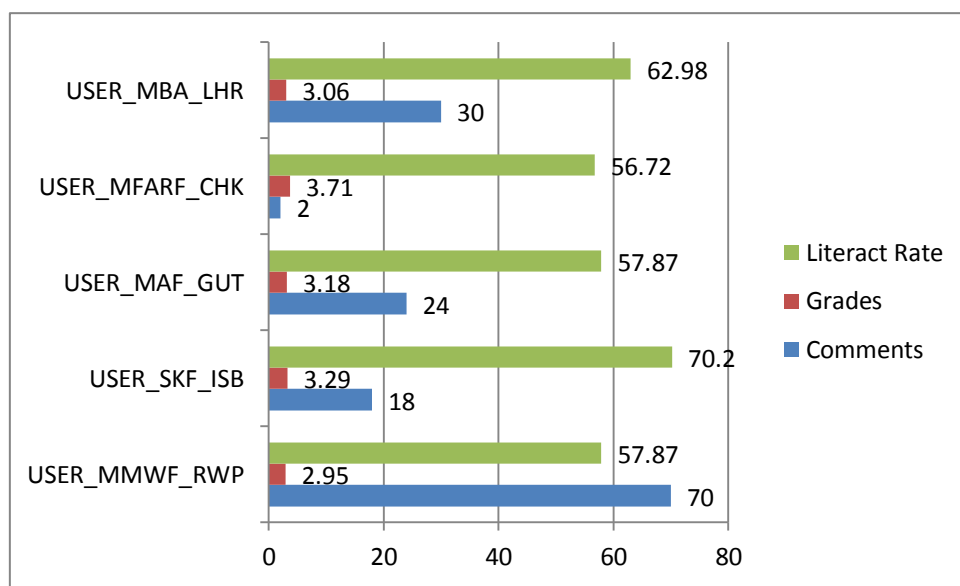


Figure 22: Cluster-11 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(4,n=5)=0.587041$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(4,n=5)=0.72332$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 12 are shown in table-12

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_MAR_GUJ	10	2.66	CITY_8	53.67
USER_MFH_GUT	3	2.86	CITY_9	57.87
USER_MAQ_HYB	6	3.13	CITY_7	42.15
USER_USA_RWP	6	2.88	CITY_2	67.50
USER_ZY_RYK	6	3.31	CITY_15	30.93

Table 13: Users of Cluster-12

The bar graph of users of cluster-12 is shown in figure-23.

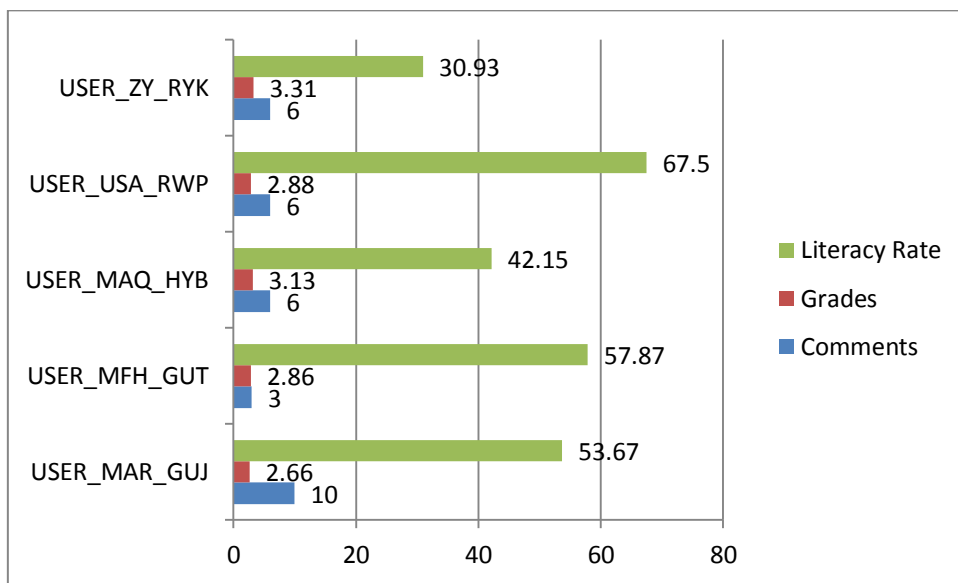


Figure 23: Cluster-12 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(4,n=5)=0.0849971$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(4,n=5)=0.003039$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 13 are shown in table-13

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_MFARF_CHK	18	3.71	CITY_12	56.72
USER_MMWF_RWP	2	2.95	CITY_2	67.50
USER_SKF_ISB	2	2.29	CITY_1	70.20
USER_NAF_FSD	11	2.27	CITY_5	51.94

Table 14: Users of Cluster-13

The bar graph of users of cluster-13 is shown in figure-24.

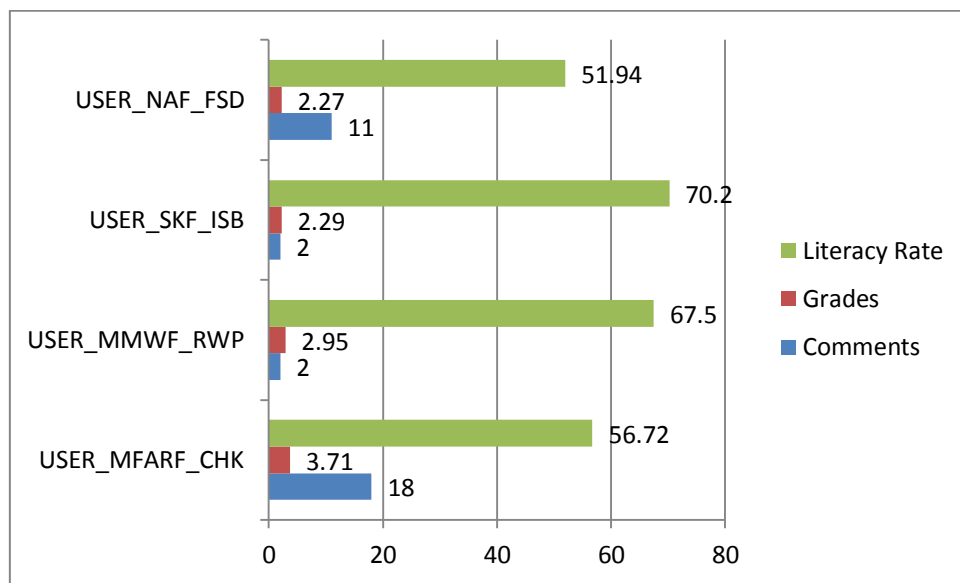


Figure 24: Cluster-13 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(3,n=4)=0.006126$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(3,n=4)=0.299643$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 14 are shown in table-14

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_MBA_LHR	513	3.06	CITY_3	62.98
USER_HAAF_ISB	161	2.27	CITY_1	70.20
USER_MAF_GUT	271	3.18	CITY_9	57.87
USER_NAF_FSD	22	2.27	CITY_5	51.94
USER_RFF_DGK	108	2.23	CITY_1	70.20
USER_SKF_ISB	122	3.29	CITY_1	70.20
USER_SNK_MAJK	57	2.69	CITY_16	60
USER_WZA_LHR	68	2.68	CITY_3	62.98
USER_UNSF_ISB	6	3.04	CITY_1	70.20
USER_MMWF_RWP	30	2.95	CITY_2	67.50

Table 15: Users of Cluster-14

The bar graph of users of cluster-14 is shown in figure-25.

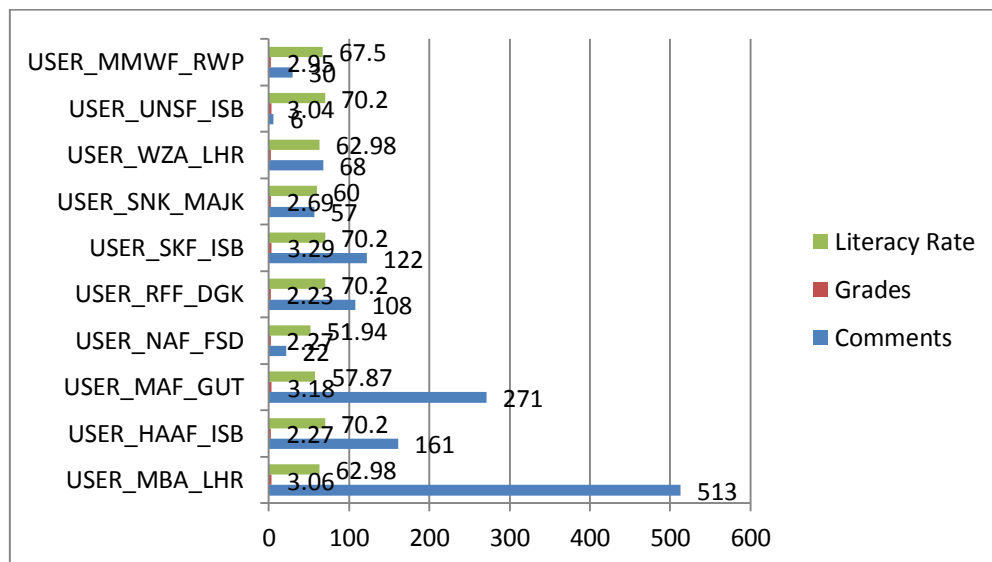


Figure 25: Cluster-14 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(9,n=10)=0.164447$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(9,n=10)=0.772036$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 15 are shown in table-15

NODE	DEGREE	CGPA	CITY	LITERACY FRATE
USER_MFH_GUT	37	2.86	CITY_9	57.87
USER_MAQ_HYB	3	3.13	CITY_7	42.15
USER_USA_RWP	1	2.88	CITY_2	67.50
USER_ZY_RYK	8	3.13	CITY_15	30.93
USER_ZA_FSD	6	2.12	CITY_5	51.94

Table 16: Users of Cluster-15

The bar graph of users of cluster-15 is shown in figure-26.

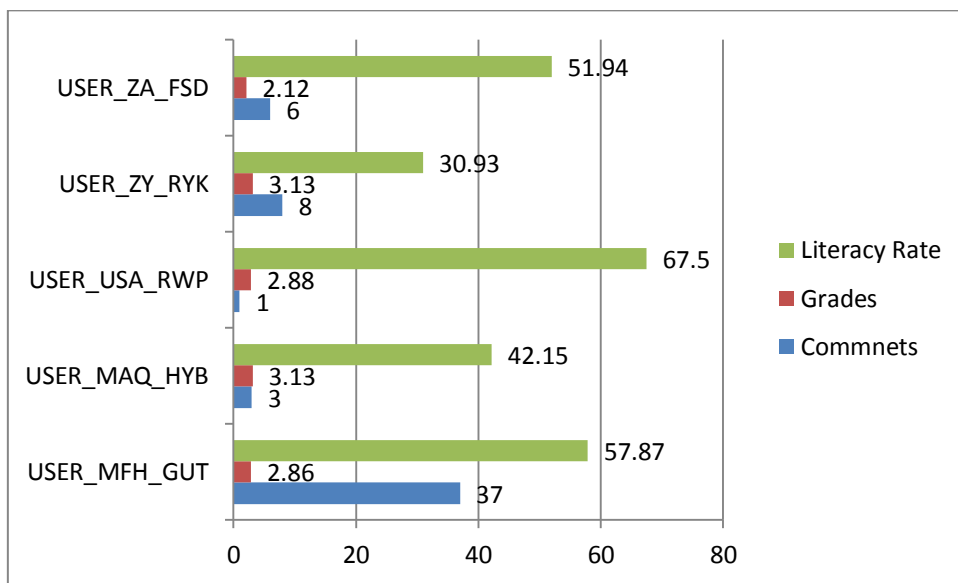


Figure 26: Cluster-15 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(4,n=5)=0.192916$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(4,n=5)=0.003129$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 15 are shown in table-16

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_MAQ_HYB	25	3.13	CITY_7	42.15
USER_IK_RWP	10	2.72	CITY_2	67.50
USER_ND_RWP	15	3.56	CITY_2	67.50
USER_HA_ATK	4	3.24	ATTOCK	44.65
USER_USA_RWP	4	2.88	CITY_2	67.50
USER_AN_LKM	8	1.95	CITY_17	29.29
USER_MAR_GUJ	6	2.66	CITY_8	53.67

Table 17: Users of Cluster-16

The bar graph of users of cluster-15 is shown in figure-27.

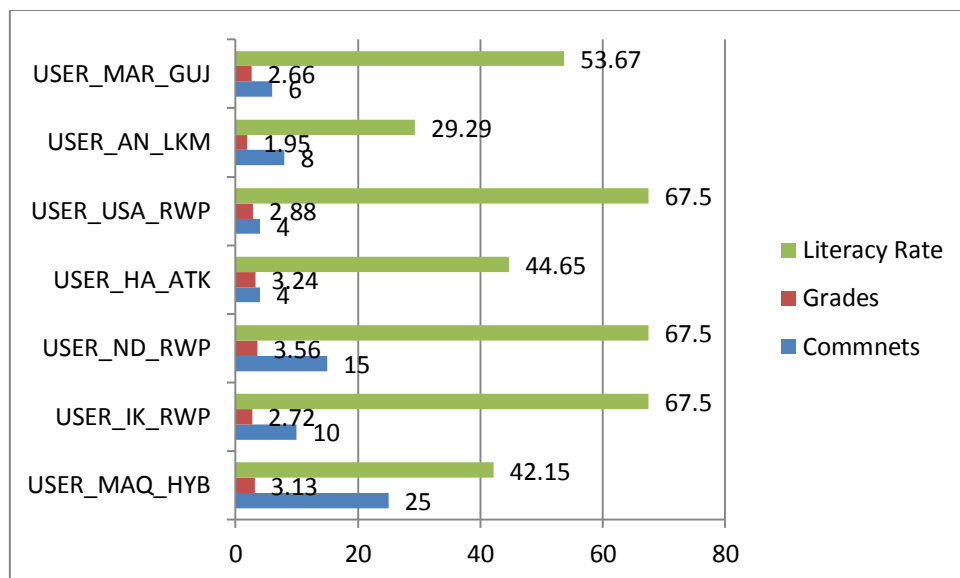


Figure 27: Cluster-16 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(6,n=7)=0.031417$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(6,n=7)=0.00226$, $p<.05$. Shows that H_1 is valid and not to be rejected.

The students that belong to cluster 17 are shown in table-17

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_ND_RWP	10	3.56	CITY_2	67.50
USER_MAQ_HYB	5	3.13	CITY_7	42.15
USER_SSK_FSD	6	2.56	CITY_5	51.94
USER_USA_RWP	6	2.88	CITY_2	67.50
USER_YZK_RWP	4	3.39	CITY_2	67.50
USER_ZR_ISB	3	1.78	CITY_1	70.20

Table 18: Users of Cluster-17

The bar graph of users of cluster-7 is shown in figure-28.

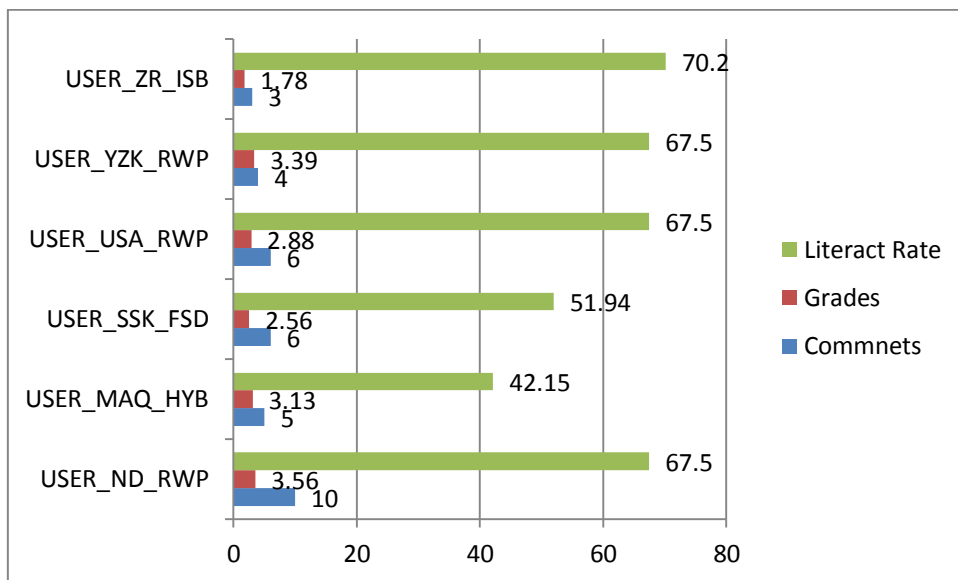


Figure 28: Cluster-17 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(5, n=6) = 0.02722$, $p < .05$. Shows that the H_0 is valid and not to be rejected.

Similarly; the participants showed the significant preferences of literacy rate among users is $X^2(5, n=6) = 0.059651$, $p < .05$. Hence H_1 is valid and not to be rejected.

The students that belong to cluster 18 are shown in table-18

NODE	DEGREE	CGPA	CITY	LITERACY RATE
USER_SKF_ISB	20	3.29	CITY_1	70.20
USER_MAF_GUT	4	3.18	CITY_9	57.87
USER_HAAF_ISB	8	2.27	CITY_1	70.20
USER_UNSF_ISB	19	3.04	CITY_1	70.20

Table 19: Users of Cluster-18

The bar graph of users of cluster-18 is shown in figure-29.

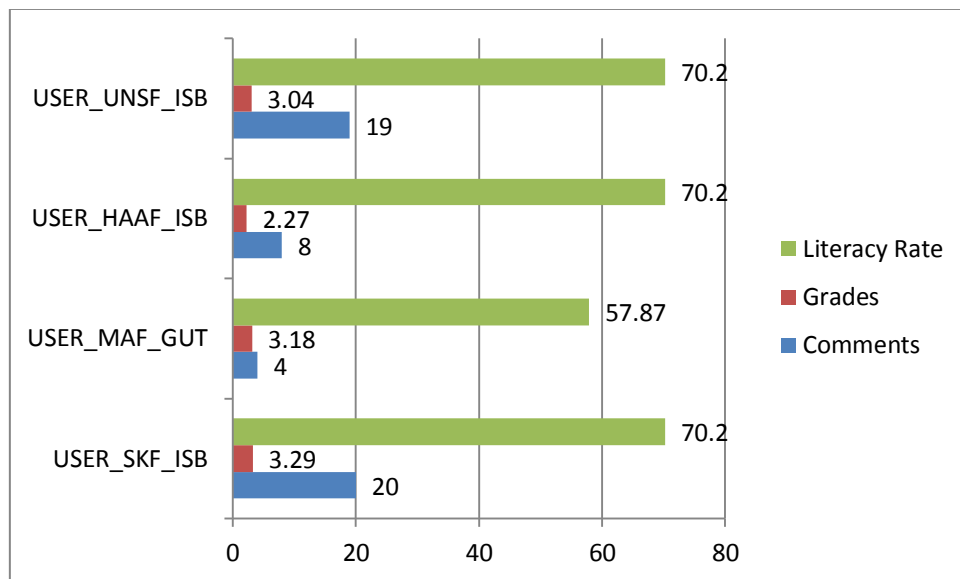


Figure 29: Cluster-18 Users Degree, CGPS and Literacy rate

The participants showed the significant preferences of grades among users is $X^2(3,n=4)=0.142406$, $p<.05$. Shows that the H_0 is valid and not to be rejected.

Similarly; The participants showed the significant preferences of literacy rate among users is $X^2(33,n=4)=0.63714$, $p<.05$. Hence H_1 is valid and not to be rejected.

From above, we have seen that in all clusters hypothesis H0 and H1 are accepted which concludes that homophily affect the student's performance and vice versa. In all clusters the similarity is closer and is less than alpha which indicates that all friendships have great impacts on the grades. As the hypothesis H0 is bi directional so that we can say that student's friendships are formed due to their grades. In both cases we can say that there is clear co-relation or similarity between the student's grades who are friends of each other.

4.3 Efficacy and accuracy

Results obtained by this research are more accurate and trustworthy because of the analysis of real life and un-intentional interactive data of users. In this study we have adopted standard social network data analytical techniques for community detection. We have collected the Facebook data representing daily life friendships. This study has shown community formation and interactions of male with male and females with females which is a "folkways" of Pakistani society, where females feel comfort to be friends with female and male feel comfort to have friendships with male. In our collected data only 8.5 % users have deviated from this social norm of Pakistani society. Our results also approve the norms of society and previously manual conducted sociological and psychological researches in different societies of world.

This research has shown the effects of homophily on academic performance and we have found our hypothesis true. At the end we have concluded that friendships affect the student's daily life activities. Our results have proven that the well-known saying "birds of feather flock together" is true.

It is crystal clear from all above discussion that results found in our study using IT methods are accurate and are well synchronized with the humanistic behavior in our society. This study is a great landmark towards the use of Facebook and other social media data for the behavioral analysis of students in the field of educational psychology and other

behavioral and social sciences. The use of social media data for behavioral study can ease the hectic task of manual data collection.

The other important aspect of this study is to specify the students to make friends of such students who are good in academic performances. We can recommend those students who are poor in academics to make buddy those students who are good in academics.

CHAPTER 5

CONCLUSION AND FUTURE WORK

This chapter will conclude the final results of this study. As discussed in Chapter 3 and Chapter 4, different social network analysis techniques are applied for friendships detection. These results concluded that good friends in real life are more engage with each other on the Facebook. This chapter is divided into two parts. In first part we shall discuss the overall conclusion drawn from this research work and in second portion we shall discuss how we extend research in future.

5.1 Conclusion

The real life friendships are based upon some homophilic characteristic. We have concluded that the friendships on the social media websites (Facebook in our case) are also based upon some homophilic characteristic. First of all the results of this thesis have concluded that friendships are based upon some homophilic characteristics secondly, the social media friendships also affects student's academic performance. This research also indicates us that by using social networking sites data we can study human behavior more efficiently and accurately. One can use social media and social networking sites as a tool for data collection to analyze the different features of society. The involvement of IT techniques for data collection in social and psychological researched can give us a clear picture of the feature of society and can provide us accurate and well managed research solutions.

5.2 Future Work

This research can be modified as a tool development that can analyzed the student friendships on timeline basis and then at the end of each semester students results can be analyzed automatically. This will help the academic branch of universities to improve the results of students by propagating the facts about friendships and results. Also this study can

be further extended on the other homophilic characteristics such as common language and ethnicity.

At the end in simple words we can conclude that homophily affects the academic performance of students and social networks such as Facebook, Twitter etc. can be used for finding the real life friendships of students and for measurement the effects of social life.

REFERENCES

- Blei, D.M., Ng, A.Y., Jordan M. I. (2003)** Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.
- Bollobas, B. (2001)** Random graphs. Cambridge University Press, Cambridge , ISBN: 0-521-80920-7
- Brandes, U. (2001)** A faster algorithm for betweenness centrality. *J Math Sociology* **25**, 163-177
- Brooks, J. G., Brooks, M. G. (1993)** In search of understanding: the case for constructivist classrooms, Alexandria, VA: American Society for Curriculum Development.
- Berndt, T. J. (1992)** Friendship and friends' influence in adolescence. *Current Directions in Psychological Science*, 1, 156–159.
- Berger, Peter L., Luckmann, T. (1966)** *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Doubleday.
- Cameron, M., Lee, B., Tom, L., Itamar, R. (2009)** Maintained relationships on Facebook, On-line at <http://overstated.net/2009/03/09/maintainedrelationships-on-facebook>
- Dong, J., Horvath, S. (2007)** Understanding network concepts in modules. *BMC Syst Biol* 24.
- Diestel, R. (2005)** Graph theory. Springer-Verlag, Heidelberg, ISBN 3-540-26182-6
- Easley, D., Kleinberg, J. (2010)** *Networks, Crowds, and Markets: Reasoning About a Highly Connected World (Large Jacket Version)*. New York, Cambridge University Press
- Evans, T., Lambiotte, R. (2009)** Line graphs, link partitions, and overlapping communities. *Physical Review E*, 80(1):16105.

- Ellison, N.B., Steinfield, C., Lampe, C. (2007)** The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of CMC*, 12, 4 , 1143–1168.
- Erosheva,E., Fienberg,S., Lafferty.J. (2004)** Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5220.
- Erdős, P., Renyi, A. (1960)** *Publ Math Inst Hung Acad Sci* 5- 17
- Feld, Scott, L. (1981)** The Focused Organization of Social Ties. *American Journal of Sociology* 86:1015–35
- Golder,S. D., Wilkinson, Huberman,B.A. (2007)** Rhythms of social interaction: Messaging within a massive online network, C. Steinfield, B. Pentland, M. Ackerman, N. Contractor, Editors , *Proceedings of the third international conference on communities and technologies*, Michigan State University, Springer, London , pp. 41–66
- Girvan,M. Newman,M. (2002)** Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826.
- Gray, M. R., Steinberg, L. (1999)** Adolescent romance and the parent–child relationship: a contextual perspective. In *The Development of Romantic Relationships in Adolescence*, W. Furman et al. (Eds). New York: Cambridge University Press.
- Junco, R. (2011)** The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. *Computers & Education*,
- Joyner, K. & Kao, G. (2001)** School racial composition and adolescent racial homophily, *Social Science Quarterly* 81(3), 810–825.

- Lin, Y.R., Sun, J., Castro, P., Konuru, R., Sundaram, H., Kelliher, A. (2009)** Metafac: community discovery via relational hypergraph factorization. In KDD, pages 527– 536. ACM.
- Ladd, G. W., Kochenderfer, B. J., Coleman, C. C. (1997)** Classroom peer acceptance, friendship, and victimization: Distinct relational systems that contribute uniquely to children’s school adjustment? *Child Development*, 68, 1181– 1197.
- Maeve, D., Joanna, B. (2013)** The Demographics of Social Media Users — 2012
- Mihaly, K. (2009)** Do more friends mean better grades? Student popularity and academic achievement. RAND Working Paper Series WR-678.
- McPherson, Miller, L., Smith, L., Cook,J. (2001)** “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27: 415-44.
- Maslov, S., Sneppen, K. (2002)** Specificity and stability in topology of protein networks. *Science* **296**- 910-913.
- Moreno, Jacob L. (1953)** Who Shall Survive? Foundations of Sociometry, Group Psychotherapy, and Sociodrama. Student ed. Roanoke, VA: Royal.
- Newman, M.E., Leicht,E. (2009)** Mixture models and exploratory analysis in net-works. *Proceedings of the National Academy of Sciences*, 104(23):9564, 2007.
- Pempek,T., Yermolayeva,Y.A., Calvert, S.L. (2009)** College students’ social networking experiences on Facebook. *Journal of Applied Developmental Psychology*, 30, p. 227-238.
- Radicchi,F., Castellano,C., Cecconi,F., Loreto,V., Parisi.D. (2004)** Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663.

- Smith, S.D., Caruso, J.B. (2010)** The ECAR study of undergraduate students and information technology. Vol.6 Boulder ,CO:EDUCAUSE center for applied research.
- Smith, S. D., Salaway, G., Caruso, J. B. (2009)** The ECAR Study of Undergraduate Students and Information Technology, (Boulder, CO:Educause Centre for Applied Research.)
- Shimbel, A. (1953)** Structural parameters of communication networks. Bull Math Biophys 15 -501-507.
- Tang,T., Liu,H. (2009)** Scalable learning of collective behavior based on sparse social dimensions. In CIKM, pages 1107–1116. ACM.
- Tang, L., Wang, X., Liu, H. (2009)** Uncovering groups via heterogeneous interaction analysis. In ICDM, Miami, FL, USA.
- Tang, J., Wang, X., Liu, H. (2012)** Integrating Social Media Data for Community Detection. In Modeling and Mining Ubiquitous Social Media.
- Verbrugge ,L.M. (1977)** The structure of adult friendship choices. Soc. Forces 56:576–97
- Wang, X., Tang, L., Gao, H., Liu, H. (2010)** Discovering overlapping groups in social media. In ICDM, Sydney, Australia.
- Yang, T., Jin, R., Chi, Y., Zhu, S. (2009)** Combining link and content for community detection: a discriminative approach. In KDD, pages 927–936. ACM.