建構專業虛擬社群內團隊之社會網絡分析系統

Developing Social Network Analysis System for Virtual Teams in a Professional Virtual Community

研究生：陳俊宏 撰

中華民國九十一年六月
Abstract

Social network analysis is used to find all relationships from the group, dig out the prominent patterns, and observe how information flows between dyads. By social network analysis approaches, users can know how information flows through network ties, how people acquire information and resources, and how cleavages and coalitions operate. In this research, we develop a useful social network analysis system to facilitate teams’ collaboration. The system can draw a social network in ego-centered or whole network layout, and provide information of social network attributes of all users. Both team leaders and general members can make use of it to understand relations and interaction patterns of their team. We also generalize social network attributes to analyze task-based teams at different team development stages for discovering the interaction patterns of different groups in groups’ life cycles. Interaction patterns of members in the team and roles that users play have high influence on a virtual team’s development. With these discoveries, team leaders can obtain concise information about their teams’ performance, and community managers can capture stereotypes of virtual teams in the community. From these evaluation results, we confirm that social network analysis is a useful means to analyze the knowledge activities conducted by virtual teams.

Keywords: Social network, social network analysis, virtual team, professional virtual community
中文摘要

社會網絡分析主要是研究群體內使用者之間彼此的關係，並能描繪出群體內顯著的結構及使用者間資訊流動的情形。透過社會網絡分析的方法，使用者可以了解群體內不同個體間關係的強度、追蹤資訊的流動及群體內小群體的分合情形。在本研究中，我們發展了一個線上的社會網絡分析系統來促進團隊成員彼此的協同合作，此系統可以針對某一個體或整個群體描繪其社會網絡圖，並提供相關的社會網絡特徵的資訊給使用者，讓使用者可以掌握團隊內的社會網絡關係。因此，不管是團隊的領導者或是一般的成員都可以使用此系統來了解團隊內成員間彼此的關係及互動情形，而團隊的領導者更可以針對不同的成員的特性採用不同的行銷或關懷策略，讓成員能更投入團隊的運作，進而促使團隊的發展。除此之外，我們也結合了社會網絡分析及資料礦採的技術分析部分工作導向型態的團隊，此部份的資料可以讓我們了解在不同發展時期團隊內成員的互動型態，團隊領導者也可以清楚了解他的團隊表現為何，並可讓社群的管理者可以了解社群內不同團隊的活動類型。從以上兩方面的結果，我們認為社會網絡分析對於虛擬團隊內知識管理的活動具有顯著的成效，而成員間的互動型態跟角色也會影響團隊的整體發展。

關鍵字：社會網絡、社會網絡分析、虛擬團隊、專業虛擬社群
Table of Contents

Abstract..................................................................................................................................... I
中文摘要...................................................................................................................................II
Table of Contents.....................................................................................................................II
Table of Contents................................................................................................................... III
List of Figure ..........................................................................................................................VI
List of Table ...........................................................................................................................VIII
Chapter 1 Introduction............................................................................................................1
  1.1 Research Background....................................................................................................1
  1.2 Research Motivation .................................................................................................2
  1.3 Research Objectives .................................................................................................3
  1.4 Research Scope ........................................................................................................3
  1.5 Research Limitations ...............................................................................................4
  1.6 Thesis Organization .................................................................................................4
Chapter 2 Literature Review ..................................................................................................5
  2.1 Social Network Analysis ..........................................................................................5
    2.1.1 Units of Analysis ...............................................................................................6
    2.1.2 Properties of Networks and Actors .................................................................7
    2.1.3 Views of Social Networks ...............................................................................9
    2.1.4 Centrality and Power ......................................................................................10
    2.1.5 Prestige ...........................................................................................................12
    2.1.6 Type of Nodes in a Network ..........................................................................13
    2.1.7 Cliques and Subgroups .................................................................................14
  2.2 Graph Theory ..........................................................................................................17
  2.3 Virtual Community ...................................................................................................18
    2.3.1 General Community ......................................................................................18
    2.3.2 Professional Community ...............................................................................19
  2.4 Virtual Team ............................................................................................................19
    2.4.1 Definitions of Virtual Teams ..........................................................................19
    2.4.2 Relations in the Virtual Teams ....................................................................20
  2.5 Knowledge Management .........................................................................................21
Chapter 3 Design of the Study .................................................................24
  3.1 Research Process ........................................................................24
  3.2 Research Framework ..................................................................25
    3.2.1 The development of Social Network Analysis System ..........26
    3.2.2 Discovering Interaction Patterns of Task-based Virtual Teams ..28
  3.3 Applying Social Network Analysis System to a Virtual Community ..29
  3.4 The Design of Evaluation ...........................................................29

Chapter 4 The Development and Evaluation of Social Network Analysis System.....31
  4.1 Functions of Social Network Analysis System ............................32
    4.1.1 Visualization of Social Network ............................................32
    4.1.2 The Computation of Social Network Evolution Trend ............39
  4.2 The Design of Evaluating Social Network System .......................41
    4.2.1 The Evaluation Design .......................................................42
    4.2.2 Evaluation Criteria of the System Performance ....................43
  4.3 Results of Performance Evaluation .............................................45
    4.3.1 Information of Subjects ......................................................45
    4.3.2 Evaluation Results with Four Criteria .................................47
  4.4 Discussions .............................................................................54

Chapter 5 The Discovery of Group Interaction Patterns ...............................56
  5.1 Training Project and Task-based Teams .....................................57
  5.2 Clustering Teams Using Social Network Attributes ....................59
  5.3 Main Communication Channels of Tasked-based Teams ...............63
  5.4 Role Difference in Social Network Attributes ............................63
  5.5 Classification and Rules Discovery ............................................65
  5.6 Managerial Implications ............................................................68

Chapter 6 Conclusion and Future Research .............................................70

References .........................................................................................73

Appendix ............................................................................................77
  Appendix A.1 Announce of Evaluation ............................................77
  Appendix A.2 Step 1 of Evaluation ...................................................77
  Appendix A.3 Step 2 of Evaluation ...................................................78
Appendix A.4 Step 3 of Evaluation.................................................................80
Appendix A.5 Step 4 of Evaluation.................................................................80
List of Figure

Figure 2.1. Hubs and Authorities (Kleinberg 1999) ................................................................. 14
Figure 2.2 A Virtual Organizational Learning Model (Lin and Lin, 2001) ................................. 22
Figure 3.1 Research Process .................................................................................................... 24
Figure 3.2 Social Network of Virtual Teams in a Virtual Community ........................................ 26
Figure 3.3 The Development Process of Social Network Analysis System ................................. 26
Figure 4.1 The Architecture of Social Network Analysis System ............................................. 31
Figure 4.2 Information Presented by Social Network Analysis System ..................................... 33
Figure 4.3 The User Interface of the Ego-centered Network Function ....................................... 33
Figure 4.4 The “Dot” Layout of an Ego-centered Network ...................................................... 34
Figure 4.5 The “Neato” Layout of an Ego-centered Network .................................................... 34
Figure 4.6 Result of an Ego-centered Network ....................................................................... 35
Figure 4.7 The User Interface of Whole Network Function ....................................................... 36
Figure 4.8 Whole Network Display with Sub-groups .............................................................. 37
Figure 4.9 Whole Network Display with Detailed Sub-group Information ................................ 37
Figure 4.10 Whole Network Display without Sub-group Information ....................................... 38
Figure 4.11 An Example of a Whole Network ......................................................................... 38
Figure 4.12 The Operation Interface of Calculating Statistics of Individual Social Network
Attributes .................................................................................................................................. 39
Figure 4.13 An Individual’s Outdegree Evolution Trend .............................................................. 40
Figure 4.14 An Individual’s Relative Outdegree and Transitivity Evolution Trend .................... 40
Figure 4.15 The Operation Interface of Calculating Statistics of Team’s Social Network
Attributes .................................................................................................................................. 40
Figure 4.16 A Team’s Indegree Evolution Trend ....................................................................... 41
Figure 4.17 Evaluation Criteria of the System Performance ..................................................... 44
Figure 4.18 Statistics of Subjects’ Energy Score .................................................................46
Figure 4.19 Statistics of Subjects’ Caring Score ..................................................................46
Figure 5.1 The Results of Clustering Six Teams across Different Time Periods...............60
Figure 5.2 The Result of Clustering Ordered by Density ......................................................61
Figure 5.3 Decision Tree Obtained from Classification .........................................................67
**List of Table**

Table 4.1 Results of Users’ Cognition of Social Networks ................................................. 48

Table 4.2 Results of Evaluation of Social Network Analysis System ................................. 49

Table 4.3 Simplicity Measured in the Questionnaire ......................................................... 50

Table 4.4 Utility Measured in the Questionnaire ............................................................... 51

Table 4.5 Utility Measured in Different Aspects ............................................................... 52

Table 4.6 Certainty Measured in the Questionnaire ......................................................... 53

Table 4.7 Novelty Measured in the Questionnaire ........................................................... 54

Table 5.1 The Agenda of Nine-week On-job Training Project (Huang, 2001) .................... 58

Table 5.2 Aggregated Data of Six Teams in Nine Weeks .................................................. 59

Table 5.3 The Time and Proportion of Density Value in Nine Weeks ................................. 62

Table 5.4 Statistics of Interactions among Team Members ............................................... 64

Table 5.5 Rules of Classification Tree ............................................................................. 68
Chapter 1 Introduction

1.1 Research Background

The wide spread of Internet technology, especially World Wide Web (WWW), facilitates individuals to communicate synchronously or asynchronously. When people interact (for example, communicate through e-mail, online talk or chat), they build their relations directly at the same time. In addition to these direct communications, when people access the resources other individuals provide or any information they contribute, they also build relations indirectly. These indirect relations become significant as the quantity and frequency of interactions increase. More and more virtual teams facilitated by Internet technology are formed to execute tasks. The management of relations has become a key success factor for strengthening the cohesion of a virtual team.

In this research, we investigate social network structures from virtual teams formed in a teachers’ professional cyber community, called SCTNet (http://sctnet.edu.tw). SCTNet, denoted as Smart and Creative Teachers Network, was established in 2000 to provide a cyber space for teachers in Taiwan area to share their educational experience, design of teaching, and research results. While they upload documents, other members of the community can download these files if they are interested in them, and they can comment on them with their professional expertise. Besides the mechanism of resources sharing, the web site also provides discussion boards in curriculum and issues of compulsory education for teachers, parents, and people who are interested in education to exchange viewpoints. They can discuss issues in education policies, teaching methods, educational reform, curriculum, and even emotional support. In the community, there is a groupware system for teachers with similar interests to form special interest groups (SIGs) as virtual teams. The facilitation functions of a SIG include the discussion board, guest book, calendar, chatting room, file sharing, websites recommendation, news announcement, e-mail, and activity status of team
members. Users can use these functions to communicate, share knowledge, and collaborate at works. There are more than two hundred SIGs in the community at this point, and they exhibit different progress patterns. It is imperative to understand how groups evolve in the teachers’ professional community.

1.2 Research Motivation

As striding forward the knowledge economy era, knowledge management becomes a key issue for organizations. It is more difficult to manage knowledge of a virtual team in a virtual community than in a real organization. Members from different professions may be hard to transcend organizational barriers to exchange knowledge. If a team wants to have good performance and produce high quality knowledge, members of the team must have close and strong relations each other. Paulleen and Yoong (2001a) consider that the important social process that can be successfully addressed by a virtual team facilitator is the building of relationships with team members. The link between team effectiveness and team member relationships is an important area of study in virtual teams. In other words, the performance of virtual teams depends on the relationships of the team.

There are more than two thousand SIGs (virtual teams) on the SCTNet. Most SIGs are formed automatically because of similar interest and some of them are built because of the training project. Because members of virtual teams rarely meet together in person, it is important for them to keep informed members’ interaction status. Knowledge of an organization is created by individuals, not by the organization itself, so that the organization should build an appropriate environment and provide useful tools to facilitate teams’ development. We believe that group knowledge will be much enhanced when team members have cooperative relationships.
1.3 Research Objectives

In a professional virtual community, individuals transcend their organizational boundaries to exchange domain expertise by the group support mechanisms, such as SIGs in SCTNet. The interaction among members is important in the virtual teams of a virtual community. The acquisition of members’ relationship and the analysis of interaction patterns are essential for facilitating knowledge sharing and creation in the community. In this research, we will develop a social network analysis system to facilitate teams’ collaboration. The system can draw a social network graph in ego-centered or whole network layout and provides information of social network attributes of all users. No matter team leaders or general members can make use of it to understand relations and interaction patterns of their team. For example, from the interdependence among members, team leaders can identify roles members play and adopt different strategies for different types of people.

Besides developing the social network analysis system, we also use social network analysis system to analyze some task-based SIGs at different periods. The results enable the discovery of the interaction patterns of different groups in groups’ life cycles. With these discoveries, team leaders can obtain concise information about their teams’ performance, and community managers can capture stereotypes of virtual teams in the community. The obtained knowledge will benefit group development in such professional community. Furthermore, community managers can make use of these rules that computed from the classification technique to predict similar groups’ performance in the future.

1.4 Research Scope

The data we use to build the social network analysis system of virtual teams on the SCTNet is from January 28, 2000 to May 01, 2002. During the time period, there are 240 SIGs formed, and we compute social network attributes values from these groups. For
evaluating the system, we choose the SIGs that have less than 40 members and set up before March 31, 2002. If there are too many members in a team, the social network is uneasy to read, and this will prohibit subjects to answer questions when doing evaluation. In the stage of discovering group interaction patterns, we only choose six SIGs built for an on-job training project from April 11, 2001 to June 06, 2001.

1.5 Research Limitations

In this research, we use the frequency of communication in a team as the strength of the relations, but we do not distinguish whether these relations are positive or negative. Besides, we ignore the strength of ties when computing some social network attributes, such as centrality, prestige and strongly connected component (SCC) because of the limitations of these algorithms.

1.6 Thesis Organization

This thesis includes six chapters:

Chapter 1. Introduction,
Chapter 2. Literature Review,
Chapter 3. Design of Study,
Chapter 4. The Development and Evaluation of Social Network Analysis System
Chapter 5. The Discovery of Group Interaction Patterns, and
Chapter 6. Conclusion and Future Research
Chapter 2 Literature Review

2.1 Social Network Analysis

When a computer network connects people or organizations, it links to a social network too. Just as a computer network is a set of machines connected by a set of cables, a social network is a set of people (or organizations or other social entities) connected by a set of socially-meaningful relationships, such as friend, co-working or information exchange (Wellman, 1996). The notation of social networks and the methods of social network analysis have attracted considerable interest from the social science research community (http://www.heinz.cmu.edu/project/INSNA). Social network analysts are interested in finding all relationships from the network, digging out the prominent patterns, and observing how information flows between dyads. They examine the information of social networks to explain the behavior and attitudes of social members. By social network approaches, analysts can study how information flows through network ties, how people acquire information and resources, and how cleavages and coalitions operate (Garton et al., 1997). Therefore, in addition to studying on whole networks, analysts are also interest in discovering densely-knit clusters or cliques and looking for similar relations. In the past three decades, social network analysis has developed a range of concepts and methods for detecting structural patterns, identifying patterns of different types of relationship interrelate, analyzing the implications that structural patterns for the behavior of network members, and studying the impact on social structures of the characteristics of network members and their social relationships (Berkowitz, 1982; Wellman, 1988; Scott, 1991; Wasserman and Faust, 1994).

Graphs have been widely used in social network analysis as a means of formally representing social relations and quantifying important social structural properties. Graph theory has been useful in social network analysis for many reasons. First, graph theory provides a vocabulary which can be used to label and denote many social structural properties.
Second, graph theory gives us mathematical operations and ideas with which many of these properties can be quantified and measured (Freeman, 1984; Seidman and Foster, 1978). When sociologists use graphs to display social networks, they rename graphics as “sociograms”.

2.1.1 Units of Analysis

Social network analysts look beyond the specific attributes of individuals to consider relations among social actors. Garton et al. (1997) synthesize literatures in past years and list four units of analysis:

**Relations.** Relations are characterized by content, direction and strength (Garton et al., 1997). The content refers to the resource that is exchanged. In social networks, pairs may exchange different kinds of information, such as social matters or work-related. The relations between peoples may have direction; in other words, a relation is directed or undirected. When one of dyad gives information and the other only receives it, the relation is directed. Besides content and direction, relations also differ in strength. We can decide the degree of strength with frequency of communication, the amount of information exchanged and the importance of information. The strength of relations that dyad communicates once a day is greater than communicate weekly or yearly.

**Ties.** A tie connects a pair of actors by one or more relations. Pairs may maintain a tie based on one relation only or a multiplex tie based on many relations. Thus, ties also have characteristics like content, direction and strength, but they are often referred to as weak or strong. Actors that are connected at short lengths or distances may have stronger connections; actors that are connected many times may have stronger ties. Ties that are weak are generally non-intimate connections by infrequently maintained. Strong ties include combinations of intimacy, self-disclosure, provision of reciprocal services, frequent contact, and kinship (Garton et al., 1997).
Multiplexity. The more relations in a tie, the more multiplex the tie. Social network analysts have found that multiplex ties are more intimate, voluntary, supportive and durable (Wellman and Wortley, 1990).

Composition. The composition of a relation or a tie is derived from the social attributes of both participants. For example, a tie is between different or same sex dyads, between a supervisor and an underlying or between two peers.

2.1.2 Properties of Networks and Actors

In this section, we will examine some of the most obvious ideas of formal network analysis methods. There are two main properties of networks and actors. One is connection, and the other is distance.

Connection. Difference among individuals in how much they are connected can be extremely consequential for understanding their attributes and behavior. More connections often mean that individuals are exposed to more diverse information. Individuals highly connected may be more influential, and may be more influenced by others. Disease, rumors and useful information spread more quickly where there are high rates of connection. More connected populations may be able to better mobilize their resources, and may be able to better bring multiple and diverse perspectives to bear to solve problems.

(1) Size, density and degree. Size is critical for the structure of social relations because of the limited resources and capacities that each actor has for building and maintaining ties. Usually the size of a network is indexed simply by counting the number of nodes. Larger social networks have more heterogeneity in the social characteristics of network members and more complexity in the structure of these networks (Wellman and Potter, 1997). The density of network is defined as the proportion of all ties that could be presented that actually are. The degree of an actor is defined as the sum of the connections between the actor and others. In a directed graph, the sum of connections from the actor to the others is called the
out-degree of the point. The degree on the contrary is called in-degree. Hanneman (1999) found that actors with ties to almost everyone else, or with ties to almost no-one else are more “predictable” in their behavior toward any given other actor than those with intermediate numbers of ties. Therefore, actors with many ties (at the center of a network) and actors at the periphery of a network (few ties) have patterns of behavior that are more constrained and predictable. Actors with only some tie can vary more in their behavior depending on to whom they are connected.

(2) Reachability. An actor is reachable by another if there exists any set of connections by which we can trace from the source to the target actor, regardless of how many others fall between them (Hanneman, 1999). If some actors in a network cannot reach others, there is the potential division of the network.

(3) Centrality. Social network analysis has developed measure of centrality which can be used to identify network members who have the most connections to others (high degree) or those whose departure would cause the network to fall apart (cut point) (Wasserman and Faust, 1994; Bonacich, 1987).

Distance. Because most individuals are not usually connected directly to most other individuals in a population, it can be quite important to go beyond simply examining the immediate connections of actors, and the overall density of direct connections in populations. The idea of the distance between actors represents how close they are to one another. It can help us to understand diffusion, homogeneity, solidarity, and other differences in macro properties of social groups. Where distances are great, it may take a long time for information to diffuse across a population. It may also be that some actors are quite unaware of, and influenced by others, even if they are technically reachable, the costs may be too high to conduct exchanges. Walk, trail and path are basic concepts to develop more powerful ways of describing various aspects of the distances among actors in a network.

(1) Walk. A walk is a sequence of actors and relations that begins and ends with actors.
It can involve the same actor or the same relation multiple times. A closed walk is one where the beginning and end points of the walk is the same actor. A cycle is a closed walk of three or more actors, all of who are distinct, except for the origin/destination actor.

(2) Trail. A trail is a walk in which all of the lines are distinct, though some nodes may be included more than once. If the trail begins and ends with the same actor, it is called a closed trail. All trails are walks, but not all walks are trails.

(3) Path. A path is a walk in all nodes and all lines are distinct. A closed path begins and ends with the same actor. All paths are trails and walks, but not all walks and all trails are path.

(4) Geodesic distance. The idea of geodesic distance is widely used in network analysis. The geodesic distance is the number of relations in the shortest possible walk from one actor to another. If there are more than one walk between two actors, we can use “geodesic path” to find the best connection. The geodesic path is often the optimal or most efficient connection between two actors.

(5) Diameter. The diameter of a network is the largest geodesic distance in the network. The value of diameter can tell us how big the network is; that is, calculate how many steps are necessary to get from one side to the other.

2.1.3 Views of Social Networks

A social network has a set of relations of ties, which can be viewed in two different ways. One approach focus on an individual, called ego-centered network, and put it at the centers of the network. Members of the network are defined by the relations with the ego. By this approach, analysts can count the number of relations, the diversity of the relations, and the links between actors named in the network. The ego-centered approach is useful when the population is large, or the boundaries of the population are hard to define (Laumann, 1983; Wellman, 1982).
The second approach considers the whole network based on some specific criterion of population boundaries such as a formal organization, department, club or kinship group. A whole network describes the ties that all members of a population maintain with all others in that group. Ideally, this approach requires responses from all members on their relations with all others in the same environment, such as the extent of email and video communication in a workgroup (Haythornthwaite et al., 1995).

Ego-centered network analysis can show the range and breadth of connectivity for individuals and identify those who have access to diverse pools of information and resources. Whole network analysis can identify those members of the network who emerge as central figures or who act as bridges between different groups.

2.1.4 Centrality and Power

One of the primary uses of graph theory in social network analysis is the identification of the most important actors in a social network. Actors who are the most important or the most prominent are usually located in strategic locations within the network. An individual has power if he can dominate others; that is, ego’s power is alter’s dependence, and vice versa. There are three measures that can calculate the importance of an actor. They are degree, betweenness, and closeness.

**Degree centrality.** The simplest definition of actor centrality is that central actors must be the most active in the sense that they have the most ties to other actors in the network or graph. The more ties an actor has, the more power they have. We define $C_D(n_i)$ as an actor’s degree centrality and $d(n_i)$ as the total number of links of $n_i$.

$$C_D(n_i) = \frac{d(n_i)}{g-1}$$

An actor has large degree if he is adjacent to many other actors and can contact with them directly. This actor with high degree centrality should be recognized by others as a major channel of relational information, and indeed play a crucial role in the network.
**Betweenness centrality.** Interactions between two nonadjacent actors might depend on the other actors in the set of actors, especially the actors who lie on the paths between the two. These middle actors potentially might have some control over the two nonadjacent actors because these nonadjacent actors depend on them. Let \( g_{jk} \) be the number of geodesics linking actor \( j \) to actor \( k \). The probability of the communication using any one of these geodesics is \( 1 / g_{jk} \). \( g_{jk}(n_i) \) is the number of geodesics of all geodesics from \( j \) to \( k \) that contain actor \( i \). Therefore, the probability of communication from actor \( j \) to \( k \) that across actor \( i \) is \( g_{jk}(n_i) / g_{jk} \). The actor betweenness index for \( n_i \) is simply the sum of these estimated probabilities over all pairs of actors not including the \( i \)th actor for \( i \) distinct from \( j \) and \( k \):

\[
C_B(n_i) = \sum_{i<j<k} \frac{g_{jk}(n_i)}{g_{jk}} \frac{1}{(g-1)(g-2)}
\]

Brandes (2001) proposed a faster algorithm for betweenness centrality. General algorithms of betweenness centrality require \( O(n^3) \) time, but the faster algorithm only require \( O(nm) \) time on unweighted networks, where \( n \) is the number of actors in the network and \( m \) is the number of links.

**Closeness centrality.** The measure focuses on how close an actor is to all the other actors in the set of actors. The idea is that an actor is central if it can quickly interact with all others; that is, central nodes in a network have minimum steps when relating to all other nodes. Sabidussi (1966) proposed that actor closeness should be measured as a function of geodesic distances. When geodesics increase in length, the centrality of the actors involved should decrease. We denote \( d(n_i, n_j) \) as the number of lines in the geodesic linking actors \( i \) and \( j \). The closeness centrality is defined as

\[
C_c(n_i) = (g - 1) \left( \sum_{j=1}^{g} d(n_i, n_j) \right).
\]
Freeman (1979) defines general group centralization index as

\[ C_A = \frac{\sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}{\max \sum_{i=1}^{g} [C_A(n^*) - C_A(n_i)]}, \]

where \( C_A(n^*) \) is the largest value of selected actor centrality measure \( C_A(n_i) \) in the set of units of a network. Network centrality value is a number between 0 and 1. \( C_A \) equals zero when all actors have exactly the same centrality index (cycle), and equals 1 if one actor completely dominates all other actors (star). Therefore, group centralities of degree, closeness, and betweenness are defined as follows,

- **Group degree centrality**: \( C_D = \frac{\sum_{i=1}^{g} [C_D(n^*) - C_D(n_i)]}{[(g-1)(g-2)]} \),
- **Group closeness centrality**: \( C_C = \frac{\sum_{i=1}^{g} [C_C(n^*) - C_C(n_i)]}{[(g-1)(g-2)]/(2g-3)} \), and
- **Group betweenness centrality**: \( C_B = \frac{\sum_{i=1}^{g} [C_B(n^*) - C_B(n_i)]}{(g-1)} \).

### 2.1.5 Prestige

Prestige measures are usually computed for directed networks only. Social prestige is connected to social power and the privilege not to reciprocate choices.

**Degree prestige.** The simplest measure of prestige is the indegree of each user, which we denoted before. This method only takes into account direct relations, but not indirect ones.

**Proximity prestige.** We define \( I_i \) as the number of users in the influence domain of user \( i \). *Influence domain, \( I_i \) equals the number of users who can reach user \( i \).* And the *proximity* is closeness that focuses on distances *to* rather than *from* each actor. In other words, what matters is how close all the actors are to \( n_i \). The average distance,
\[ \sum d(n_j, n_i)/I_i, \] is a crude measure of proximity. The proximity prestige is

\[ P_p(n_i) = \frac{I_i/(g-1)}{\sum d(n_j, n_i)/I_i}. \]

### 2.1.6 Type of Nodes in a Network

The indegrees and outdegrees of the nodes in a directed graph can be used to distinguish four different kinds of nodes based on the possible ways that arcs can be incident with the node (Wasserman and Faust, 1994). The indegree of node \( n_i \) is denoted by \( d_i(n_i) \) and outdegree is denoted by \( d_o(n_i) \). Nodes can be classified into

- **isolate** if \( d_i(n_i) = d_o(n_i) = 0 \),
- **transmitter** if \( d_i(n_i) = 0 \) and \( d_o(n_i) > 0 \),
- **receiver** if \( d_i(n_i) > 0 \) and \( d_o(n_i) = 0 \), and
- **carrier or ordinary** if \( d_i(n_i) > 0 \) and \( d_o(n_i) > 0 \).

Kleinberg (1999) proposed the concepts of *authorities* and *hubs*. An authority not only has large indegree, but also the nodes that connect to it have large outdegree. A hub is a node that has large outdegree and the nodes it connects have large indegree. In other words, a node is a good hub, if it points to many good authorities, and is a good authority, if it is pointed by many good hubs.
2.1.7 Cliques and Subgroups

One of the most common interests of structural analysts is the sub-structures that may be present in a network. Subgroups are subsets of actors among whom there are relatively strong, direct, intense, frequent, or positive ties. From the idea of subgroups or cliques within a network, we can understand social structure and the embeddedness of individuals. For example, some people may act as “bridges” between groups. Some actors may be part of tightly connected and closed elite, while others are completely isolated from this group. Such differences in the ways that individuals are embedded in the structure of groups within in a network can have profound consequences for the ways that actors see the work, and the behaviors that they are likely to practice.

2.1.7.1 Bottom-up approaches

The approach begins with basic groups, and seeks to see how far this kind of close relationship can be extended. The notion is to build outward from single ties to construct the network. It emphasizes how the macro might emerge out of the micro. Bottom-up
approaches include cliques, n-cliques, n-clans, k-plexes and k-cores.

**cliques.** A clique in a graph is a maximal complete sub-graph of three or more nodes. It consists of a subset of nodes, all of which are adjacent to each other, and there are no other nodes that are also adjacent to all of the members of the clique.

**n-cliques.** A n-clique is a maximal sub-graph in which the largest distance between any two nodes is no greater than n. Formally, a n-clique is a sub-graph with node set $N_s$, such that $d(i, j) \leq n$ for all $n_i, n_j \in N_s$, and there are no additional nodes that are also distance n or less from all nodes in the sub-graph. The n-clique approach allows an actor to be a member of a clique even if it does not have ties to all other clique members; just so long as they do have ties to some member, and are no further away than n steps from all members of the clique. Because of these characteristics, an n-clique has two problems: first, an n-clique, as a sub-graph may have a diameter greater than n, and second, an n-clique might be disconnected.

**n-clans.** The n-clan starts with the n-cliques that are identified in a network and excludes those n-cliques that have a diameter greater than n. An n-clan is an n-clique in which the geodesic distance, $d(i, j)$, between all nodes in the sub-graph is no greater than n for paths within the sub-graph. The n-clans in a graph can be found by examining all n-cliques and excluding those that have diameter greater than n.

**k-plexes.** A k-plex is a maximal sub-graph containing n nodes in which each node is adjacent to no fewer than n-k nodes in the sub-graph. In other words, each node in the sub-graph may lack ties to no more than k sub-graph members. We denote the degree of a node $i$ in sub-graph $g_s$ by $d_s(i)$. A k-plex as a sub-graph in which $d_s(i) \geq (g_s - k)$ for all $n_i \in N_s$ and there are no other nodes in the sub-graph that also have $d_s(i) \geq (g_s - k)$. Since nodes within a k-plex will be adjacent to many other members, a k-plex is more robust than an n-clique, and removal of a single is less likely to leave the sub-graph disconnected.
Rather than the larger and stringy groups sometimes produced by \(n\)-clique analysis, \(k\)-plex analysis tends to find relatively large numbers of smaller groups.

**\(k\)-cores.** Another approach to cohesive subgroups based on nodal degree is the \(k\)-core (Seidman, 1983). A \(k\)-core is a sub-graph in which each node is adjacent to at least a minimum number, \(k\), of the other nodes in the sub-graph. As before, we denote the degree of a node \(i\) in sub-graph \(g\) by \(d_g(i)\). A sub-graph is a \(k\)-core if \(d_g(i) \geq k\) for all \(n_i \in N_g\).

A \(k\)-core is defined in terms of the minimum degree within a sub-graph, or the minimum number of adjacencies that must be present. In contrast to the \(k\)-plex, which specifies the acceptable number of lines that can be absent from each node, the \(k\)-core specifies the required number of lines that must be present from each node to others within the sub-graph.

### 2.1.7.2 Top-down approaches

**Component.** Components of a graph are parts that are connected within, but disconnected between sub-graphs. A component is a sub-graph in which there is a path between all pairs of nodes in the sub-graph (all pairs of nodes in a component are reachable) and there is no path between a node in the component and any node not in the component. A SCC is a strongly connected maximal subgraph in the graph. In other words, all nodes in a SCC have at least a path to other nodes in the same component.

**Block.** If a node of a graph is removed, the structure is divided into un-connected system. The node has the attribute is called “cutpoint”. The divisions into which cutpoints divide a graph are called blocks. In other words, a node \(n_i\) is a cutpoint if the number of blocks in the graph that contains \(n_i\) is fewer than the number of blocks in the sub-graph that results from deleting \(n_i\) from the graph.

**Lambda Set.** Borgatti et al. (1990) define the method. Consider pairs of nodes in the sub-graph with node set \(N_i\). The set of nodes, \(N_i\), is a lambda set if any pair of nodes in the lambda set has larger line connectivity than any pairs of nodes consisting of one node.
from within the lambda set and a second node from outside the lambda set. Formally, the lambda set is a subset of nodes, $N_s \subseteq N$, such that for all $i, j, k \in N_s$, and $l \in N - N_s$, $\lambda(i, j) > \lambda(k, l)$. Where the valued, $\lambda$, is the minimum number of lines that must be removed to disconnect the graph.

2.2 Graph Theory

Social network analysis is highly relative with graph theory. Graph theory gives us mathematical operations and ideas with which many of these properties can be quantified and measured (Freeman, 1984; Seidman and Foster, 1978). In graph theory, we only discuss the shortest paths because it is more related to social network analysis. The social network attributes, such as closeness prestige, betweenness prestige, prestige, and transitivity, will use the concept of shortest paths.

The property of shortest paths in graphs is elementary and important. A path from vertex $V_i$ to vertex $V_j$ in a graph $G$ is a sequence of vertices.

Single-Source Shortest Paths (SSSP). It is the descriptive name for the problem of finding the shortest paths to all the nodes in a graph from a single designated source. This algorithm we use is Bellman-Ford algorithm. The steps of the algorithm are shown as follows.

```
BellmanFord(graph (G, w), vertex s)
    Initialize Single Source (G, s)
    for $i \in [1, |V| - 1]$ do
        for $(u, v) \in E[G]$ do
            Relax $(u,v,w)$
        for $(u, v) \in E[G]$ do
            if $d[v] > d[u] + w(u, v)$ then
                return false
            return true
```

All-Pairs Shortest Paths (APSP). Instead of finding a single source, this algorithm
finds all pairs of shortest paths. The algorithm we choose for finding APSP is Floyd-Warshall algorithm. The algorithm is shown as follows.

\[
\text{FloydWarshall(matrix } W, \ \text{integer } n) \\
\text{ for } k \leftarrow 1 \text{ to } n \text{ do} \\
\quad \text{ for } i \leftarrow 1 \text{ to } n \text{ do} \\
\quad \quad \text{ for } j \leftarrow 1 \text{ to } n \text{ do} \\
\quad \quad \quad d_{ij}^{(k)} \leftarrow \min(d_{ij}^{(k-1)}, d_{ik}^{(k-1)} + d_{kj}^{(k-1)}) \\
\text{ Return } D^{(n)}
\]

2.3 Virtual Community

2.3.1 General Community

The Internet offers a unique medium for communications which knows no geographical boundary. More and more virtual communities appear with the rapidly growth of information technology and Internet. Rheingold (1993) deals with the emergence of socially-motivated communities of interest on the Internet. He describes virtual communities as “social aggregations that emerge from the network when enough people carry on those public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace.” Hagel and Armstrong (1996) take the business perspective and define virtual communities as "virtual enterprises". In 1999, Schubert adopted the third perspective that socially motivated aggregations of people in electronic networks dispose of certain success factors that can be transferred to the concept of a virtual community of transaction. She defines virtual communities as the union between individuals or organizations who share common values and interests using electronic media to communicate within a shared semantic space on a regular basis. Their communication is thus independent from restrictions of time and place.” Hagel and Armstrong (1996) highlight four categories of community based on needs of transaction, interest, fantasy and
relationship. *Communities of transaction* are communities in which the core activity that brings together the active participants is engagement in specific transactions. *Communities of interest* are those in which participants share common interests in some specific topic or concern. Typically these would involve a higher degree of inter-personal communication. *Communities of fantasy* provides participants with a place where they create new environments, personalities, or stories on their own will. *Communities of relationship* are created by the coming together of individuals who share similar experience, world views and who wish to share these with like-mined people. Typically these communities enable strong personal connections because individuals’ identities are often made transparent.

2.3.2 Professional Community

Members in professional communities are different from general communities. They must have professionalism and professional intellect. Westheimer (1998) and Scribner, et al. (1999) consider that members generally have shared norms and values, and they carry out critical reflection and continue the professional dialogues with one another. Some teachers’ professional communities, composed of teachers, appear in 1991 (Bullough and Githin, 1991). In Louis, Marks, and Kruse’s analysis (1996), they indicate that the distinctive and critical characteristics of teachers’ professional community are (1) shared norms and values, (2) a focus on student learning, (3) collaboration, (4) deprivatization of practice, and (5) reflective dialogue. Most members of the virtual community, SCTNet, have teaching background, and they learn and develop teachers’ professionalism across schools instead of learning alone within their schools in the community. Therefore, SCTNet is one of teachers’ professional communities.

2.4 Virtual Team

2.4.1 Definitions of Virtual Teams
Virtual teams are a relatively new phenomenon and, by definition, work across time and distance through the use of information and communication technologies (ICTs) (Townsend et al., 1998). Information technology now enables globally distributed people to collaborate on issues and challenges across space, time, and organizational boundaries (Lipnack and Stamps, 1997). Virtual teams may communicate and work synchronously or asynchronously through technologies such as electronic mail, bulletin board, audio/video/data conferencing, automated workflow, electronic voting and collaborative writing (Coleman, 1997). Virtual teams are playing an increasingly important role in organizational life and are often assigned the most important tasks in an organization, such as multinational product launches, negotiating mergers and acquisitions among global companies and managing strategic alliances (Maznevski and Chudoba, 2000).

2.4.2 Relations in the Virtual Teams

An important social process that can be successfully addressed by a virtual team facilitator is the build of relationships with team members. The link between team effectiveness and team member relationships is an important area of study in virtual teams (Paulleen and Yoong, 2001a). Effective communication is the key to successful virtual teams and one of the keys to effective communication is how well team members are able to build and maintain their personal relationships (Lau et al., 2000). Stronger relational links have been associated with higher task performance (Warkentin and Beranek, 1999) and the effectiveness of information exchange (Warkentin et al., 1997). Kimball (2000) stated that the purpose of building and maintaining relationships in teams is to ensure that individuals develop at least enough harmony to be able to get their group work done. Walther and Burgoon (1992) also considered that strong relational links are associated with enhanced creativity and motivation, increased morale, better decisions and fewer process losses. The build of relationships with virtual team members is a fundamental concern for virtual team
facilitators (Pauleen and Yoong, 2001b). Social network analysis is used to analyze the
relations and interaction patterns among different users from which we analyze virtual teams
in this research.

2.5 Knowledge Management

Knowledge is intangible, boundaryless, and dynamic, and if it is not used at a specific
time in a specific place, it is of no value. There are two kinds of knowledge: \textit{explicit}
knowledge and \textit{tacit} knowledge (Nonaka and Konno, 1998). Explicit knowledge can be
expressed in words and numbers and shared in the form of data, scientific formulae,
specifications, manuals, etc. Tacit knowledge is highly personal and hard to formalize, and
difficult to communication or share with others. Examples include subjective insights,
intuitions, and hunches. Knowledge map may serve as directories pointing to explicit
knowledge, and social network serves as directories pointing to individuals withholding the
needed but inarticulate knowledge (Lin and Lin, 2001).

Lin and Lin (2001) proposed a virtual organizational learning (VOL) model to formally
elaborate learning in a virtual organization. They view a VOL model as a transactive
memory system (Figure 2.2). In the VOL model, a knowledge map captures the knowledge
covered in a virtual organization. There are four main mnemonic functions in order to
manage knowledge map in a virtual organization.

(1) \textit{Knowledge allocation function} can links knowledge objects of knowledge maps to
individuals of social networks, so that specific knowledge can be allocated through social
networks.

(2) \textit{Social network updating function} can refresh the relationship between individuals during
the interaction along the evolution of a virtual organization.

(3) \textit{Knowledge maintenance function} maintains the knowledge map from performing
business processes in a virtual organization by inserting, modifying, or deleting
knowledge objects and their dependencies.

(4) **Collaborative knowledge retrieval function** is used to retrieves problem solving or decision making related knowledge by collaboratively accessing the transactive memory of individuals.

![Figure 2.2 A Virtual Organizational Learning Model (Lin and Lin, 2001)](image)

The transactive memory system is important to accelerate learning in virtual organizations. Lin and Lin (2001) summarized the values of transactive memory system as follows.

1. Knowledge map can be created quickly and maintained easier in a virtual organization.
2. Relative expertise of members in a virtual organization can be easily identified.
Members with different expertise may be spatially distributed in organizations.

(3) Facilitate faster job training and role orientation for individuals in a virtual organization.

(4) Provide integrated and relevant knowledge to members in a virtual organization.

(5) Make mental model explicit and allow people with expertise in many different domains in a virtual organization to communicate with each other.

(6) Encourage people to share knowledge with others.
Chapter 3 Design of the Study

3.1 Research Process

Figure 3-1 illustrates the research processes of this research. Literatures related to social networks analysis, virtual teams and virtual communities are surveyed first as the basis on developing a complete online social network analysis system of virtual teams. After integrating concepts extracted from literatures, we propose the architecture of social network analysis system, and then develop it for virtual teams on the SCTNet. We then analyze social network attributes of task-based virtual teams by using data mining techniques, such as clustering and classification. The performance of social network analysis system is evaluated through experiments with various criteria.
3.2 Research Framework

Figure 3.2 illustrates a conceptual framework of social network formation in virtual teams of a virtual community. Members of a virtual community come from different organizations. They have their own social networks in real world. When they form virtual teams, their social networks are developing through their interactions using various communication means, such as discussion board, guest book, file sharing, e-mail, online message, etc. When their tie is tight, their social connection is strong. For team members with high frequent interactions implying close relationship, sub-group is formed. For an individual joining different teams, different social networks are formed at the same time.

There are two main objectives in this research. First, we will develop an online social network analysis system to facilitate virtual team collaboration among team members. Second, we will identify social networks of task-based teams during project periods. From the results of social network analysis, we hope to answer the following questions. How does a team's interaction correlate to its performance? Do teams in the comparative performance have the similar interaction patterns in the progress of project? We will adopt data mining techniques, such as clustering and classification, to analyze social networks of virtual teams.
3.2.1 The development of Social Network Analysis System

![Diagram of Social Network Analysis System](image)

**Figure 3.3 The Development Process of Social Network Analysis System**

Figure 3.3 illustrates the development process of the social network analysis system.
The inputs of this system are online message logs and various interaction activity logs, such as discussion board, guest book, file sharing, website recommendation, e-mail, project log. We describe system development tasks as follows.

(1) **Relation analysis.** The relation between two team members may be direct or indirect. If two persons have direct interactions, their relations are stronger. Their direct interactions may be via mailing to others, sending messages to others, downloading others’ shared files, replying others’ posts, and linking to others’ recommended websites. Indirect relations between two persons include mailing to the same third person, sending messages to the same persons, downloading the same files, replying to the same discussion subjects, linking to the same websites. In this research, we only consider direct relations both in direction and multiplexity because they are more meaningful in social networks. The higher interaction in a pair, the stronger strength their tie is. According to types of relations, relations are assigned by different weights.

(2) **Roles definition and group information computing.** In this step, we use the properties of social networks to specify the roles that individuals play. These properties denoted by graphs include centrality and prestige. With these properties, we can know the prestige individuals hold, and define individual characteristics within the group. The role an individual plays may be isolated, a transmitter, a receiver, or a carrier. The role a team member acts may demonstrate various degrees of significance in a team’s performance. Besides individual characteristics, the group itself can be represented by such properties as (a) the size and density of the group, (b) the types of actors in the group, (c) the number and size of sub-groups in the group, (d) the average reachable rate and transitive rate in the group, and (e) the centrality and prestige of the group.

(3) **Sub-group identification** Sub-group identification is used to find users with highly connected relations in a group. The method we use in this research is strongly connected component (SCC). Members of a SCC at least have a path to other members in the same
SCC. From the SCC identification, team leaders and members can understand changes of structure in the team more easily.

(4) Visualization Social networks can be visualized either in ego-centered network or whole network. Ego-centered network can show an individual’s range and breadth of connectivity. Visualizing whole network helps identify the global structure of the network, and dig out the status of cleavage. Different team roles are granted with different views. Individuals as a general member can only view ego-center graphs, while team leaders can read the team’s whole network because they hold the responsibility in managing the team. In this research, we adopt a visualization tool called webdot developed by AT&T labs for graph drawing.

3.2.2 Discovering Interaction Patterns of Task-based Virtual Teams

In this section, we use social network analysis to analyze task-based virtual teams at their different periods. Each team is presented by several characteristics that are computed from social network analysis and graph theory. These attributes are density, reachable rate, transitive rate, indegree centrality, outdegree centrality, closeness centrality, betweenness centrality, proximity prestige and the number of strongly connected components (SCC). These attributes can represent various interactions of the group members. The evolution of these attributes enables the discovery of the interaction patterns of different groups in groups’ life cycles. With these discoveries, team leaders can obtain concise information about their teams’ performance, and community managers can cluster groups to capture stereotypes of virtual teams in the community. The obtained knowledge will benefit group development in a professional community. Furthermore, community managers can make use of these rules that computed from classification technique to predict similar groups’ performance in the future. In order to discover patterns and identify social network attributes that may be relative with teams’ performance, we focus on three analysis methods of teams’ development:
(1) a clustering technique, \( k \)-means, to cluster social networks and observe teams’ development patterns, (2) role identification to analyze team members’ social network and explain how the role influences a team’s performance, and (3) a classification technique, decision tree, is used to generate classification rules to predict a team’s performance.

### 3.3 Applying Social Network Analysis System to a Virtual Community

There are three aspects of applying social network analysis to a virtual community. Individuals, groups and the community can benefit from the system.

For a community manager, the activities and changes of groups are important to understand the community evolution. The social network analysis system not only provides managers information of all groups but also clusters them with similar properties. From these data, managers can capture the trends of task-based groups and forecast the teams’ processes according to the patterns in the past and then provide proper suggestions to these groups.

For a group leader, the group social network analysis help understand the interaction among group members for the intention to increase members’ enthusiasm in getting involved in group activities. Effective communication is the key to successful virtual teams, which depend on how well team members are able to build and maintain their personal relationships (Lau et al., 2000). Therefore, a group leader can gain insights into the structure of the group relationships based on the social network analysis, and then improve the group’s effective communications.

For a group member, the presentation of ego-center social network helps individuals understand how themselves have been interacting with team colleagues.

### 3.4 The Design of Evaluation

We will conduct experiments with various criteria to evaluate the performance of the
proposed social network analysis system.

**Individual social network.** In the experimentation, subjects are asked to fill the strength of peer relation in the degree between 0 and 10 before presenting individual social networks. After viewing the social network presented by the social network analysis system and social network attribute values, they are asked to adjust the degree of strengths. Besides, each subject fills in a questionnaire to express his or her opinions about the system. We can obtain certainty and utility value according to the difference between system results and individual’s adjusted values, and novelty values according to the difference between original and adjusted values. We evaluate simplicity values from the questionnaire.

**Group-level social network.** Each virtual team leaders is asked to fill in a questionnaire after using the social network analysis system. We can understand how useful the system for team management.
Chapter 4 The Development and Evaluation of Social Network Analysis System

In this chapter, we describe the development and evaluation of the proposed social network analysis system. In the system, we provide a visualization interface for team members to easily understand the structure of the team and the relations of team members. Figure 4.1 depicts the architecture of the system that we develop in this research. We input the data of a team (or a user) from the database server, and then we analyze these data with social network analysis and graph theory. The main output from the analysis is a visualization layout file with web dot’ format. Web dot is a CGI program, developed by AT&T, which can draw graphs according to the layout file we generate. What an end user sees in the system includes not only the social network graph but also the detailed information of social network analysis.

Figure 4.1 The Architecture of Social Network Analysis System
Based on the architecture of social network analysis system, we describe functions designed for ego-centered and whole network in subsection 4.1, the evaluation process in 4.2, the evaluation results in 4.3, and managerial discussions in 4.4.

4.1 Functions of Social Network Analysis System

In order to make the social network analysis system useful and friendly for virtual teams in the community, we build the functions with many concepts of social network described in Chapter 2. The system consists of two main components. One is the visualization of social network (Subsection 4.1.1), and the other is the evolution trend of social network (Subsection 4.1.2). We also add notes to social network for team members to understand the denotation of social networks.

4.1.1 Visualization of Social Network

One contribution of the system is the visualization of social network of virtual teams on the fly. A team member can know how strong he or she relates to other team colleagues, and what communication channels of their interaction by traversing the social network. A team leader can know the roles of its members play, which information can further facilitate morale inspiration and task assignment. The information provided by the system is listed in Figure 4.2.
4.1.1.1 Ego-centered network

Ego-centered network positions an individual at the center of a team social network for the person to traverse his or her relationships with other team members. An individual can choose one of six sources of data or the integration of six data to form an ego-centered network. Figure 4.3 shows the user interface of the ego-centered network function.

Figure 4.2 Information Presented by Social Network Analysis System

Figure 4.3 The User Interface of the Ego-centered Network Function
A user can choose data sources, time period, link weight, graph size, and graph layout. The system calculates all teams’ data every 30 days. When a user makes a new query, he or she can choose to view only a period of data or an accumulative data of the team. The layout provided in this system includes two types: *dot* and *neato*. Dot is used to make hierarchical layouts of graphs (Figure 4.4), and neato is used to make "spring" model layouts of graphs (Figure 4.5).

![Figure 4.4 The “Dot” Layout of an Ego-centered Network](image)

![Figure 4.5 The “Neato” Layout of an Ego-centered Network](image)
Figure 4.6 Result of an Ego-centered Network

Figure 4.6 is an example of an ego-centered network. The right frame of the window shows the ego social network information with various attributes. The main frame shows the ego-centered network. The left bottom frame denotes the corresponding members’ names to the network node numbers. The shape and color of the network are explained below this left bottom frame, where circle node represents woman, rectangle is man, and rhombus represents the user without gender information.

4.1.1.2 Whole network

A whole network describes the ties that all members of a team interact with each other. Figure 4.7 shows the user interface of whole network visualization. A user can choose how to generate the whole network with the following parameters: (1) data source, a user can choose various data sources, (2) time period, a user can choose data from a specific period or
accumulated data, (3) display, a user can choose to view whole network with sub-groups (Figure 4.8), with detailed sub-group information (Figure 4.9) or without sub-groups-information (Figure 4.10), (4) link weight, whether to assign weight to relation or not, (5) node information, whether to show node information or not, (6) evolution difference view, whether to calculate the difference between different periods, where we use different color and type of line and node to represent the difference, (7) display size, users can decide the size of display according to their team size, (8) layout using dot or neato layout of webdot, (9) role configuration, which a user can set up the parameters of different roles (isolate, transmitter, receiver, and carrier).

Figure 4.11 illustrates an example whole network. The right frame of the window shows social network information of the team. The information includes basic information of the team, graph statistic, and roles information that members play. The main frame illustrates the whole network. The left bottom frame shows the corresponding members’ names to node numbers and the sub-group information.

<table>
<thead>
<tr>
<th>工作坊社會網絡查詢 (whole network)</th>
</tr>
</thead>
<tbody>
<tr>
<td>資料來源</td>
</tr>
<tr>
<td>時間範圍</td>
</tr>
<tr>
<td>開示層級</td>
</tr>
<tr>
<td>開示權重</td>
</tr>
<tr>
<td>開示node資訊</td>
</tr>
<tr>
<td>開示時間差異</td>
</tr>
<tr>
<td>圖形大小</td>
</tr>
<tr>
<td>layout方式</td>
</tr>
<tr>
<td>其它設定 (僅適</td>
</tr>
<tr>
<td>宜)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Figure 4.7 The User Interface of Whole Network Function
In Figure 4.8, the notation \( g(n,d) \) in a node notifies a subgroup \( g \) and its properties in subgroup size and density denoted by \( n \) and \( d \) respectively. The gray level of the node also denotes the density. Larger the density is, darker the gray level of the node is. A node is black if its density is 1, and white if its density is 0.

In Figure 4.9, the notation \((x_1+x_2, y_1+y_2)\) in a node notifies that the user has \((x_1+x_2)\) indegree, where \(x_1\) is the degree from the same sub-group and \(x_2\) outside from the sub-group, and \((y_1+y_2)\) outdegree, where \(y_1\) is the degree to the same sub-group and \(y_2\) to those nodes.
outside of its sub-group.

Figure 4.10 Whole Network Display without Sub-group Information

In Figure 4.10, the notation $i(x,y)$ notifies the identifier of a user $i$, and his or her indegree and outdegree relation with $x$ and $y$ respectively.

Figure 4.11 An Example of a Whole Network
4.1.2 The Computation of Social Network Evolution Trend

The social network analysis system also calculates the evolution trend of individuals and teams by various social network attributes through different periods. The operation interface is shown in Figure 4.12, and some examples are showed in Figure 4.13 and 4.14. Notice that the red line is a user’s data and the blue dot line is the average value of the team. The two extremities of a green line are maximum and minimum values of the team during that period. A team leader can also obtain the evolution trend of his or her team with various social network attributes across different periods. Figure 4.15 shows the operation interface, and Figure 4.16 is an example showing a team’s indegree evolution trend. These attributes are indegree, outdegree, transitivity, reachability, indegree centrality, outdegree centrality, closeness centrality, betweenness centrality, prestige, and farness.

Figure 4.12 The Operation Interface of Calculating Statistics of Individual Social Network Attributes
Figure 4.13 An Individual’s Outdegree Evolution Trend

Figure 4.14 An Individual’s Relative Outdegree and Transitivity Evolution Trend

Figure 4.15 The Operation Interface of Calculating Statistics of Team’s Social Network Attributes
4.2 The Design of Evaluating Social Network System

We conduct survey to evaluate the performance of the social network system in order to understand its strength and weakness. We select teams satisfying the following criteria from 230 SIGs of the SCTNet: (1) the team size less than 40 members, and (2) the team is formed before March 31, 2002. The reasons of such filtering criteria are explained as follows. For a team with too many members, the social network is uneasy to read, and this will prohibit subjects from answering questions while evaluating the system. Because we calculate social network attributes per month, team members may not know each other well for those teams formed lately.

There are 182 SIGs satisfying these two criteria. We gather activity logs (include discussion board, guest book, file sharing, website recommendation, e-mail and online message) of all SIGs from the date that teams formed to May 01, 2002. We then compute social network attributes of these teams. The survey is scheduled from May 21, 2002 to June 3, 2002 on the SCTNet. Subjects belonging to the selected teams are asked to answer the presented social network and questionnaires while they log on SCTNet during the survey.
4.2.1 The Evaluation Design

We conduct the survey according to the following four steps: (the detailed screen captures are attached in Appendix A)

**Step 1 Subject’s self-evaluation.** When a subject chooses a SIG, the system will list the team members, and ask the subject to fill out the degree of relationship (between 0 and 10) between the subject and every teammate. Score “0” represents that the subject does not know the member at all. The higher score the stronger relation they may have.

**Step 2 System’s suggestion.** The social network analysis system has calculated the relation strength between the subject and other team colleagues according to the group interaction logs. The social network and attributes generated by the system are shown to the subject. In order to align with the subject’s score range, we standardize system generated attribute values ranging between 0 and 10. Since users may choose a maximum value as the reference score to compare their relative strength of relations with others, we standardize system values by dividing the raw values by the maximum value $X^*$ in the team. The standardized value $S_s$ of original value $O_s$ is defined as $S_s = \frac{O_s}{X^*} \times 10$.

The system shows the ego-centered network and social network analysis information to the subject. The information includes all social network attributes we compute, the evolution trend of these attributes, and the role the subject plays. We also provide a detailed explanation and examples to let the subject understand what these social network attributes mean. The subject can also view the whole network and its attributes listed in Subsection 4.1.1.2.

**Step 3 Subject’s adjustment.** In this step, the system will list both the system value computed in step 2 and the value set by the subject in step 1. A user’s perception of his or
her social network may be different from the values generated by the system from group interaction logs. In this step, the subject is asked to adjust the relationship score he or she set in step 1 while viewing the corresponding value set by system. A subject has two alternatives for the adjustment. One is to keep the initial value set in step 1, and the other is to assign a new score. If the subject thinks the initial value inaccurate after viewing the results of social network analysis system, he or she can replace suitable values for specifying the relation strength.

**Step 4 Filling out questionnaires.** Each subject is asked to answer a questionnaire, which includes three sections. In the first section, we ask users’ cognition about their team. These question are “do you know who have close relationships in the team?” and “do you think it is important to understand relations among all members during the team development?” In the second section, we ask their attitudes toward the benefits of the social network system. These questions are “can the system increase your relations with team colleagues?”, and “can the system improve your team development?”. In the third section, we ask their judgment of the system performance. These questions are “are the system results accurate?”, and ”does the system give you some different information, which surprises you?”.

**4.2.2 Evaluation Criteria of the System Performance**

The criteria we use to evaluate the performance of the social analysis system include *simplicity, utility, certainty* and *novelty*. We define $V_1$ as subjects’ initial values filled in Step 1, $V_2$ as the system’s results presented in Step 2, and $V_3$ as users’ final adjustments in Step 3. Values from questionnaire can be used to compare with $V_1$, $V_2$ and $V_3$. Figure 4.17 shows how we compute these evaluation criteria. Utility is determined by calculating the correlation coefficient between $V_2$ and $V_3$. The difference of $V_1$ and $V_3$ is tested by pair-t test. Certainty is obtained by comparing $V_2$ and $V_3$. We check the proportion of $V_2$ and $V_3$ that have the same values or same orders. The value of novelty is obtained from the change
proportion of $V_1$ and $V_3$. 

**Figure 4.17 Evaluation Criteria of the System Performance**

**Simplicity.** It is a factor to indicate the degree of human comprehension of the patterns generated by the social network analysis system. The more complex of the network structure the more difficult for a person to understand its embedded relations. In this research, we hope that team members can easily understand the social network structure and descriptive attributes generated by the social network analysis system. We calculate simplicity from subjects' answers to the first section of questionnaires.

**Utility.** The system has high utility if it is very helpful for subjects to tuning their subjective perception to their social networks with team colleagues. We measure utility value from two aspects. The simplest method is to summarize the answers of the questionnaire. This value can be thought as the proportion of users that agree its usefulness. We can also compute the *correlation coefficient* between $V_2$ and $V_3$. If their correlation coefficient is high, the system has great influence on users' adjustment toward final values.
We also adopt statistical method, *pair-t test*, to test $V_1$ and $V_3$. If these two values have significant difference, it means that their difference is caused by the system. Therefore, we can claim that the system is useful by the evidence that users are influenced to change their initial perceptions.

**Certainty.** The system results should be valid or trustworthy to users. We calculate certainty firstly by measuring the proportion that the $V_2$ values are equal to $V_3$ values. Because the system information may not precisely match users’ cognition, we compare the sequence of $V_2$ and $V_3$ to obtain higher certainty. If a value generated by the system is not equal to the subject’s value but they have the same relative ordering, we also view that this system result is accurate. We also obtain certainty value from the questionnaire.

**Novelty.** If the system results can bring new clues which surprise team members, we consider the system has its novelty. We can observe it from the degree of change from the final adjusted value to the value set in Step 1. Novelty means that results from the system can bring team members additional information that they may not know before. We also obtain certainty value from the questionnaire.

### 4.3 Results of Performance Evaluation

#### 4.3.1 Information of Subjects

During the two-week survey period (05/21/2002–06/03/2002), 99 subjects from 53 SIGs responded to the survey. Among 99 subjects, 20 users are team leaders, 24 users are assistants, and 55 members are general members. Each subject practiced the social network analysis system and filled in the degree of relation strength between himself or herself with other team members. There are 1912 community members in 53 SIGs; that is, 99 subjects may answer their relations with 1912 members in total.

The profiles of these subjects responding to the survey are illustrated in Figure 4.18 and Figure 4.19. There are two measures on SCTNet to indicate the degree of involvement of
members to community activities, such as discussion, file sharing, and communications. One is energy index which refers to the degree of contribution for a member to share knowledge; the other is a caring score which refers to the degree of responsiveness for a member to return others’ sharing. By the end of survey period (06/03/2002), 8280 community members had been joining 230 SIGs. Their average energy score is 79.09 and caring score is 28.59. From Figure 4.18 and 4.19, the majority of subjects in the survey has highly involved in team activities.

![Figure 4.18 Statistics of Subjects’ Energy Score](image)

![Figure 4.19 Statistics of Subjects’ Caring Score](image)
4.3.2 Evaluation Results with Four Criteria

From the questionnaire, we found that 77.78% members could identify their relations with their team colleagues. Among three types of SIG members, the proportion of general members is the lowest, which only 69.09% of them can identify their relations with other members. However, 87.5% of SIG assistants and 90% of SIG leaders can recognize their relations with team colleagues, which is higher than general members. To know the peer relations with others may be easy, but it is more difficult to understand the whole relational structure among team members. Only 50.51% of subjects in average can understand the relations among teammates, and specifically 49.09% of general members, 50% of team assistants, and 55% of leaders can do so. We are surprised by the low proportion of leaders and assistants since their roles are important to teams’ development. If they cannot understand the relations among their team members, it is difficult to have high performance.

Although most subjects cannot realize their relations with colleagues clearly, 82.83% of subjects consider they will have more activities and better performance if they can know their ego-centered social network. 89.90% of them think the whole social network can facilitate the development of their teams. In general, from the answers of the first section of the questionnaire, we know it is important for team members to understand their teams’ social networks. These findings are summarized in Table 4.1.
Table 4.1 Results of Users’ Cognition of Social Networks

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can understand who are close to me in my team at any time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>18.18%</td>
<td>12.73%</td>
<td>69.09%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>4.17%</td>
<td>8.33%</td>
<td>87.50%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>0.00%</td>
<td>10.00%</td>
<td>90.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>11.11%</td>
<td>11.11%</td>
<td>77.78%</td>
</tr>
<tr>
<td>I can understand all relations of all members in my team</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>38.18%</td>
<td>12.73%</td>
<td>49.09%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>29.17%</td>
<td>20.83%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>25.00%</td>
<td>20.00%</td>
<td>55.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>33.33%</td>
<td>16.16%</td>
<td>50.51%</td>
</tr>
<tr>
<td>I think it will improve the performance of my team if all users can understand relations in the team</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>3.64%</td>
<td>10.91%</td>
<td>85.45%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>0.00%</td>
<td>4.17%</td>
<td>95.83%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>0.00%</td>
<td>5.00%</td>
<td>95.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>2.02%</td>
<td>8.08%</td>
<td>89.90%</td>
</tr>
<tr>
<td>I think I can improve my performance if I can understand relations in the team</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>9.09%</td>
<td>12.73%</td>
<td>78.18%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>4.17%</td>
<td>8.33%</td>
<td>87.50%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>0.00%</td>
<td>10.00%</td>
<td>90.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>6.06%</td>
<td>11.11%</td>
<td>82.83%</td>
</tr>
</tbody>
</table>

After knowing subjects’ cognition, we evaluate the system’s performance based on three criteria, utility, certainty, and novelty as shown in Table 4.2.
Table 4.2 Results of Evaluation of Social Network Analysis System

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Role</th>
<th>System Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Role</td>
<td>Correlation coefficient of $V_2$ and $V_3$</td>
</tr>
<tr>
<td>Utility</td>
<td>Member</td>
<td>0.399</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>0.558</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>0.566</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>0.444</td>
</tr>
<tr>
<td>Certainty</td>
<td>Role</td>
<td>$V_2$ and $V_3$ have the same value</td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>29.52%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>39.02%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>31.56%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>32.01%</td>
</tr>
<tr>
<td>Novelty</td>
<td>Role</td>
<td>Change rate of $V_1$ and $V_3$</td>
</tr>
<tr>
<td></td>
<td>Member</td>
<td>16.94%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>27.57%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>43.07%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>23.95%</td>
</tr>
</tbody>
</table>

**Simplicity.** From the questionnaire, only 58.59% subjects think the system is easy to understand (Table 4.3). We are not surprised by this result because 70.45% members of SCTNet are teachers of primary or junior schools. Most of them do not have technology background and are not familiar with the concept of social network analysis. Even we have made detailed explanations and examples during the survey, some subjects respond that some attributes in social network analysis are not easy to understand. Experiment time is another factor that influences the result. A subject needs to spend about 30 minutes to complete an survey. The speed of Internet will impact their intension of reading these explanations and understand the presentation of social networks.
**Table 4.3 Simplicity Measured in the Questionnaire**

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data in social network analysis system is easy to understand</td>
<td>Member</td>
<td>18.18%</td>
<td>21.82%</td>
<td>60.00%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>29.17%</td>
<td>20.83%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>20.00%</td>
<td>15.00%</td>
<td>65.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>21.21%</td>
<td>20.20%</td>
<td>58.59%</td>
</tr>
</tbody>
</table>

**Utility.** The value of utility in the questionnaire is 62.63% (Table 4.4). General members think it is more useful than team leaders and assistants. That is because leaders and assistants have more interactions with other members in order to maintain the development of the team. Therefore, more than 50% assistants and leaders agree that the system is useful to them. The correlation coefficient between system and users’ decision is 0.444. It means that the system is positive and middle associate with users’ decision. The result in Table 4.2 also tells that the system is more useful to assistants and leaders because correlation coefficient of assistants and leaders is bigger than common members. We also test users’ first and final decisions by pair-\( t \) test. The result reveals that the two values have significant difference at 5% level of significance (95% significance). The results of assistants and leaders have significant difference, but general members do not. So that we can know the system has higher influence to assistants and team leaders. From the results of questionnaire, correlation coefficient or pair-\( t \) test, we can say the system is useful. This explains why users change their initial decisions after viewing the analysis of social networks.
In the questionnaire, we also ask usefulness of the system in different aspects. These aspects that we focus on include roles type users play, activity status control, relationships strengthen, relationships understanding, individual performance improvement, and team’s performance improvement. The results are listed in Table 4.5. No matter in what aspects, more than half of users approve the system’s value. From the results of questionnaire and statistical methods, we believe that the system is useful to all users, no matter leaders or general members.

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social network analysis system is useful and valuable for me</td>
<td>Member</td>
<td>7.27%</td>
<td>21.82%</td>
<td>70.91%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>12.50%</td>
<td>37.50%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>15.00%</td>
<td>30.00%</td>
<td>55.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>10.10%</td>
<td>27.27%</td>
<td>62.63%</td>
</tr>
</tbody>
</table>
Table 4.5 Utility Measured in Different Aspects

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understand roles that all members play and their position in the team</td>
<td>Member</td>
<td>3.64%</td>
<td>21.82%</td>
<td>74.55%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>0.00%</td>
<td>25.00%</td>
<td>75%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>5.00%</td>
<td>20.00%</td>
<td>75%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>3.03%</td>
<td>22.22%</td>
<td>74.75%*</td>
</tr>
<tr>
<td>Understand and control the activity status in the team</td>
<td>Member</td>
<td>3.64%</td>
<td>14.55%</td>
<td>81.82%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>8.33%</td>
<td>20.83%</td>
<td>70.83%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>5.00%</td>
<td>20.00%</td>
<td>75%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>5.05%</td>
<td>17.17%</td>
<td>77.78%</td>
</tr>
<tr>
<td>Strengthen the relations between me and other members</td>
<td>Member</td>
<td>12.73%</td>
<td>27.27%</td>
<td>60%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>8.33%</td>
<td>29.17%</td>
<td>62.5%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>20.00%</td>
<td>25.00%</td>
<td>55%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>13.13%</td>
<td>27.27%</td>
<td>59.60%</td>
</tr>
<tr>
<td>Improve individual performance in the team</td>
<td>Member</td>
<td>10.91%</td>
<td>27.27%</td>
<td>61.82%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>8.33%</td>
<td>45.83%</td>
<td>45.83%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>10.00%</td>
<td>45%*</td>
<td>45%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>10.10%</td>
<td>35.35%</td>
<td>54.55%</td>
</tr>
<tr>
<td>Understand the relations of all members</td>
<td>Member</td>
<td>3.64%</td>
<td>12.73%</td>
<td>83.64%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>8.33%</td>
<td>29.17%</td>
<td>62.5%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>10.00%</td>
<td>20.00%</td>
<td>70%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>6.06%</td>
<td>18.18%</td>
<td>75.76%</td>
</tr>
<tr>
<td>Improve the team's performance</td>
<td>Member</td>
<td>1.82%</td>
<td>23.64%</td>
<td>74.55%*</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>4.17%</td>
<td>25.00%</td>
<td>70.83%*</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>0.00%</td>
<td>35.00%</td>
<td>65%*</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>2.02%</td>
<td>26.26%</td>
<td>71.72%</td>
</tr>
</tbody>
</table>

Certainty. The certainty value in the questionnaire is 55.56% (Table 4.6). The value is not high and it may be caused by the standardization method we use. The weigh range of a relation users fill in is from 0 to 10, but their real interaction frequency in the virtual team does not satisfy the range. In order to let users can compare easily, we standardize all frequency by dividing the max frequency in the team and multiple them by 10. The method
will have problem if there is an outlier maximum value because other values will very small after standardization and users may think these values are wrong.

Table 4.6 Certainty Measured in the Questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data in social network analysis system is accurate</td>
<td>Member</td>
<td>14.55%</td>
<td>23.64%</td>
<td>61.82%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>25.00%</td>
<td>25.00%</td>
<td>50.00%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>20.00%</td>
<td>35.00%</td>
<td>45.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>18.18%</td>
<td>26.26%</td>
<td>55.56%</td>
</tr>
</tbody>
</table>

In addition to questionnaire, we also compute certainty from the system values and users’ final decisions. When we just consider the exact match of values between system generated and users’ decisions, only achieve 32.01% accuracy. If we define the system value is accurate when it has the same order as users’ decision even they do not have the same value, the certainty becomes 61.19%. From Table 4.2, we can identify that the system is more accurate for assistants (75.70%) and team leaders (98.23%) than common members (48.30%) when considering value and order. Because assistants and leaders have to maintain the development and are familiar with other members, they have higher probability to accurately recall members’ relations.

**Novelty.** The novelty value in the questionnaire is 63.64% (Table 4.7). Most users consider the social network analysis system can bring some special information to them. We measured the value from the change rate of users’ final decisions. If they change their initial choices, it means that they get additional information from the system. The average value of change rate is 23.95%. About a quarter of users change their minds because of the results of the system. Team leaders have higher value than assistants and common members (Table 4.2).
Table 4.7 Novelty Measured in the Questionnaire

<table>
<thead>
<tr>
<th>Question</th>
<th>Role</th>
<th>Disagree</th>
<th>No Sense</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social network analysis system gives me surprise and information that I don't recognize before</td>
<td>Member</td>
<td>14.55%</td>
<td>18.18%</td>
<td>67.27%</td>
</tr>
<tr>
<td></td>
<td>Assistant</td>
<td>12.50%</td>
<td>29.17%</td>
<td>58.33%</td>
</tr>
<tr>
<td></td>
<td>Leader</td>
<td>15.00%</td>
<td>25.00%</td>
<td>60.00%</td>
</tr>
<tr>
<td></td>
<td>All Users</td>
<td>14.14%</td>
<td>22.22%</td>
<td>63.64%</td>
</tr>
</tbody>
</table>

4.4 Discussions

Social network analysis has developed a range of concepts and methods for detecting structural patterns, identifying patterns of different types of relationship interrelate, analyzing the implications that structural patterns for the behavior of network members, and studying the impact on social structures of the characteristics of network members and their social relationships (Berkowitz, 1982; Wellman, 1988; Scott, 1991; Wasserman and Faust, 1994). From evaluation results, most subjects agree that social network analysis can facilitate the team development and improve users’ performance in virtual teams. The value of simplicity criterion is not very high because not all of users are used to using information technique, and familiar with social network analysis. If we want the system to become easier for people, it needs not only a detailed description of all attributes, but also a good transform mechanism to transform social network concept from academic jargons to casual descriptions.

The average values of certainty in system results and questionnaire are not as high as we expected. From the system results, we discovered that certainty values of assistants and leaders are higher than those of common members. The roles of team leader and assistant have high predicted accuracy due to highly involved team activities. However, most of mis-match between user and system values come from different ranking scales, which mainly from extremely high degree of relation strength distorted the distribution of scores. The large value makes strength of other relations become too small, and subjects may think the value system calculated is inaccurate. Because of the different relative score computation
among subjects, the system may provide different standardization methods for different patterns in mitigating this problem.

In measuring utility, subjects consider that the social network analysis system is useful in different aspects. Statistically, the social network analysis influences users’ decisions in updating their cognition of team members’ relations. The system also surprises subjects by providing additional information of team members’ relations. The novelty value from questionnaire results is larger than system results, which may be caused by the standardization method used in social network analysis system.
Chapter 5 The Discovery of Group Interaction Patterns

With the social network analysis system developed in this study, we can visualize the team social network along with the team development. The social networks cannot only provide information of single team’s social networks, but also contribute data for data mining techniques to discover interaction patterns of virtual teams. This chapter is aimed at demonstrating how data mining techniques, such as clustering and classification, to discover group interaction patterns with task-based SIGs at different periods. Every SIG is presented by several characteristics computed from social network analysis and graph theory. These attributes are density, reachable rate, transitive rate, indegree centrality, outdegree centrality, closeness centrality, betweenness centrality, proximity prestige and the number of strongly connected components (SCC). These attributes can represent various interactions of the group members. The evolution of these attributes enables the discovery of the interaction patterns of different groups in groups’ life cycles. With these discoveries, team leaders can obtain concise information about their teams’ performance, and community managers can cluster groups to capture stereotypes of virtual teams in the community. The obtained knowledge will benefit group development in such professional community. Furthermore, community managers can make use of these rules generated by classification techniques to predict similar groups’ performance in the future. In order to discover patterns and identify social network attributes that may be relative with teams’ performance, we focus on three analysis methods of teams’ development. First, we use clustering technique to cluster social network data and observe teams’ development patterns. Second, we analyze team members’ roles from their social networks, and explain how the role influences teams’ performance. Finally, we demonstrate the prediction of teams’ performance using classification rules generated by the classification technique, decision tree once we can collect sufficient group interaction data.
We are interested in identifying social networks of teams during the project period. Do the interaction and communication of members in the same team will influence their performance finally? Do teams in the comparative performance have the same interaction patterns in the progress of project? We hope to identify social network attributes which classify them to various levels of performance. In order to answer these questions, we adopt data mining techniques, such as clustering and classification, to analyze the data gathered from social network analysis.

We will introduce the training project and task-based teams in Subsection 5.1, clustering results in 5.2, communication channels of teams in 5.3, difference of roles in social network analysis in 5.4, classification results in 5.5 and consequence in 5.6.

5.1 Training Project and Task-based Teams

In order to apply social network analysis to the virtual teams’ activities, we choose some task-based teams from the SCTNet. In 2001, an on-job training project was launched by recruiting teachers from primary schools to exercise the collaborative lesson plan development to integrate information technology for subject teaching. In the first two weeks of the project, expert teachers introduced the strategies of adopting information technology to teaching. Members of a team collaborate to develop a lesson plan through SCTNet’s group supporting functions from the third to seventh week. In the last two weeks, teachers excised lesson plans they had designed, and domain experts commented their practices, graded their team performance. Table 5.1 summarizes the agenda of the nine-week on-job training project (Huang, 2001).
Table 5.1 The Agenda of Nine-week On-job Training Project (Huang, 2001)

<table>
<thead>
<tr>
<th>Week</th>
<th>Phases</th>
<th>Activities</th>
<th>Milestones</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>Introduction of the strategies and practices in how to adopt IT to teaching by two experienced primary school teachers</td>
<td>1. Participant teachers can use the SIGs on the SCTNet 2. Team-up</td>
</tr>
<tr>
<td></td>
<td>Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Preparation</td>
<td>Introduction of the strategies and practices in how to adopt IT to teaching by the other two experienced primary school teachers</td>
<td>1. Group leader are elected 2. “Team portrait” homework is assigned (including introduction of members, the name, goal, and task of the team.)</td>
</tr>
<tr>
<td></td>
<td>Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Design</td>
<td>Design lesson plans of using IT to teach through group interaction and collaboration</td>
<td>Lesson plans of using IT to teaching are developed</td>
</tr>
<tr>
<td>4</td>
<td>Phase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Revision</td>
<td>Prototyping lesson plans</td>
<td>The revision of lesson plans</td>
</tr>
<tr>
<td>7</td>
<td>Phase</td>
<td>Finishing lesson plans</td>
<td>1. Finished lesson plans 2. Decide the date of bring lesson plans into practice</td>
</tr>
<tr>
<td>8</td>
<td>Reflection</td>
<td>Practice teaching</td>
<td>Practice developed lesson plans</td>
</tr>
<tr>
<td>9</td>
<td>Phase</td>
<td>1. Reflective dialogues on the outcomes of practicing lesson plans 2. Comments from experts</td>
<td>Review and reflection</td>
</tr>
</tbody>
</table>

There are six teams when the training project started on April 11, 2001. Each team focuses on its interested subject. Team A1, A2, …, A6 focused on art and humanity, science, social study, math, integrated activity, and language respectively. There were 6 to 12 teachers participating in each team. Members of Team A2 to A6 came from different schools, and Team A1 was formed by teachers from the same school. We classified six teams into two sets according to their outputs graded by expert teachers. Team A2, A4, and A6 are assigned to $G$ set due to their high performance, and Team A1, A3, and A5 are in $B$ set.
which show low performance. Although members of Team A1 came from the same school, Huang (2001) found that “the products of teams that the members came from different schools were not certainly superior to those teams that members come form the same schools, and vice versa.”

During the training project, these teams are asked to use SIGs in the virtual community SCTNet, to communicate and collaborate with each others. The facilitation functions provided by SIG include discussion board, guest book, calendar, chatting room, file sharing, website recommendation, news announcement, e-mail, online message, and activity statistics of team members. Team members can communicate and interact synchronously and asynchronously by using these functions. Social network are built through interacting with SIG functions. In this research, we gather data from six functions (discussion board, guest book, file sharing, websites recommendation, e-mail, and online message) and aggregate these data shown in Table 5.2.

<table>
<thead>
<tr>
<th>Team\ Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>6</td>
<td>28</td>
<td>37</td>
<td>35</td>
<td>11</td>
<td>15</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>141</td>
</tr>
<tr>
<td>A2</td>
<td>11</td>
<td>6</td>
<td>18</td>
<td>12</td>
<td>11</td>
<td>16</td>
<td>7</td>
<td>15</td>
<td>9</td>
<td>105</td>
</tr>
<tr>
<td>A3</td>
<td>2</td>
<td>3</td>
<td>23</td>
<td>11</td>
<td>7</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>60</td>
</tr>
<tr>
<td>A4</td>
<td>2</td>
<td>10</td>
<td>14</td>
<td>11</td>
<td>14</td>
<td>13</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>77</td>
</tr>
<tr>
<td>A5</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>A6</td>
<td>11</td>
<td>18</td>
<td>35</td>
<td>26</td>
<td>13</td>
<td>19</td>
<td>41</td>
<td>16</td>
<td>9</td>
<td>188</td>
</tr>
</tbody>
</table>

5.2 Clustering Teams Using Social Network Attributes

Social network emphasizes the relationships built through the interactions among people. For example, when a user posts an article in the discussion board, no relation forms if no one replies to it. The relation forms when one person replies it. The similar formation occurs
with other group interaction functions. There are many characteristics in social network analysis. We focus on group level characteristics to identify teams’ interaction patterns and activities first. The attributes we use include density, reachable rate, transitive rate, indegree centrality, outdegree centrality, closeness centrality, betweenness centrality, proximity prestige and the number of strongly connected components. We compute the means and standard deviation of these attributes in a weekly basis.

After obtaining teams’ social network values, we want to know whether these teams belonging to the same category have the similar activity pattern during these nine weeks. In other words, we want to test whether teams having high score of performance are clustered together in all periods. Because a team is described by many social network characteristics, clustering is used for grouping similar teams together. In this study, we adopt $k$-means clustering, and $k$ is set to 2 or 3 depending on the purity of the clustering results, and we choose the highest degree of purity one as the last result. Figure 5.1 is the result of six teams across different time periods with aggregated social network values.

![Figure 5.1 The Results of Clustering Six Teams across Different Time Periods](image-url)
In the preparation phase, members of most teams had similar social network patterns because most activity in this phase were proposed and controlled by the project manager. Therefore, the difference is small even there are two clusters in the phase. During the design and revision phases, every team was asked to collaborate autonomously, the project manager intervened only when teams requested for help. We found that all teams acted differently and had diverse relationships. In the final week of the reflection phase, we can easily distinguish two categories from clustering. Team A1, A3 and A5 are in the same cluster which have bad performance. Team A2 and A4 are in the same one, and Team A6 itself is a cluster. Although Team A2, A4 and A6 are not in the same cluster, their social network characteristics are different from those of Team A1, A3 and A5. This result implies that social network of team members is correlative with teams’ performance, especially after the middle phase.

**Figure 5.2 The Result of Clustering Ordered by Density**

We can specify which teams have similar social network patterns during their life cycle from Figure 5.1. In order to know which clusters are better, we re-cluster all data to 3 clusters and use “density” as a criterion to arrange the nodes’ position in Figure 5.1. Density
is a characteristic that denotes the proportion of all ties that could be present. It is related with many other characteristics in social network analysis, such as degree, centrality, prestige, SCC and so on. If a team is always in high density during the progress of development, it may have high performance and cohesion. Figure 5.2 is the graph after rearranging according to the density value of every cluster, and Table 5.3 is the proportion of time of six teams according to their density value.

<table>
<thead>
<tr>
<th>Density</th>
<th>Team A1</th>
<th>Team A2</th>
<th>Team A3</th>
<th>Team A4</th>
<th>Team A5</th>
<th>Team A6</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>3 weeks</td>
<td>5 weeks</td>
<td>1 week</td>
<td>6 weeks</td>
<td>0 week</td>
<td>4 weeks</td>
</tr>
<tr>
<td>Medium</td>
<td>1 week</td>
<td>3 weeks</td>
<td>2 weeks</td>
<td>3 weeks</td>
<td>5 weeks</td>
<td>2 weeks</td>
</tr>
<tr>
<td>Low</td>
<td>5 weeks</td>
<td>1 week</td>
<td>6 weeks</td>
<td>0 week</td>
<td>4 weeks</td>
<td>3 weeks</td>
</tr>
</tbody>
</table>

*The bold numbers represent the highest frequency of every team

The density of team A2, A4 and A6 was always above the medium score in the nine weeks. Team A4’s performance was the best because it was either in high or medium degree of density. Although Team A6 did not perform well in the early stage, it performed actively in the later stage. Team A3 performed the worst as its members communicate rarely at most of the time. Team A5’s performance was good at the first sight because it achieved the medium degree of density half of project time, but its trend of density curve is downward. The interaction of Team A5 is only high in the early stage, but it does not last to the final stage. We can conclude that the performance of teams is influenced by the relationship and interaction of members, especially in later phase of team development. However, we have no sufficient evidence to conclude that the relations between a team’s early stage performance to its final outcomes.
5.3 Main Communication Channels of Tasked-based Teams

From the aggregated data analysis, the status of social network of a team is a critical factor in affecting the performance of a team. It may be insightful to analyze major channels users communicate and build their social networks. It is also helpful to improve the team interaction which appeals to users.

In six interaction functions of SIGs, online message belongs to the synchronous method and others are asynchronous. We cluster teams by individual functions, and obtain surprised results. We found that teams’ social network in G set rarely built by e-mail and online message. These asynchronous methods, such as discussion board (or guest book) and file sharing are more popular. Social networks built in B set use e-mail and online message more frequently. Intuitively, e-mail and online message should build stronger relation than discussion board and file sharing. However, in the reality, the information cannot be kept for reusing in group’s memory. Besides, e-mail and online message are more suitable for one-to-one communication. The goal of group formation is to collaborate with lesson plan development, which is important for members to share knowledge broadly. Therefore, discussion board (or guest book) and file sharing serve their best roles to achieve the task.

5.4 Role Difference in Social Network Attributes

Members of a team may play various roles, such as team leader, team assistant or participant. Team leaders and assistants are leaders and the rest of members are participants. Table 5.4 lists average values of these two types of roles by aggregating six team interaction functions.

From Table 5.4, we found that members in the leader role played a more central role in the team, except for Team A5. Members in the participant role in Team A5 demonstrated stronger contribution to social network because Team A5 leaders did not direct and guide others well. For Team A1, either leaders or participants performed well in the early stage,
but this situation did not last to the end. The situation of Team A3 was not similar to that of Team A1. The average value of Team A3 is only high in 3, 4, and 7 weeks, but low in other weeks. Besides, the main source of social network in Team A3 was from the function of e-mail, not from discussion board and file sharing. That is the reason that its value was high but the performance was low. The value of Team A2 was relatively low in G set. Their good performance was mainly from their positive attitudes toward team collaboration. They asked for help when they encountered problems. They also had the habit to meet together to discuss and exchange information. That is the reason that their social networks attributes’ values were low but achieved high performance.

### Table 5.4 Statistics of Interactions among Team Members

<table>
<thead>
<tr>
<th>Performance class</th>
<th>Team</th>
<th>Type</th>
<th>Reachable rate</th>
<th>Transitive rate</th>
<th>Indegree Centrality</th>
<th>Outdegree Centrality</th>
<th>Closeness Centrality</th>
<th>Betweenness Centrality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A1</td>
<td>Leader</td>
<td>0.3210</td>
<td>0.6173</td>
<td>0.1975</td>
<td>0.3827</td>
<td>0.3147</td>
<td>0.0267</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.3155</td>
<td>0.2826</td>
<td>0.1715</td>
<td>0.1509</td>
<td>0.2181</td>
<td>0.0138</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>Leader</td>
<td>0.2778</td>
<td>0.6296</td>
<td>0.1111</td>
<td>0.3519</td>
<td>0.3993</td>
<td>0.0074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.2901</td>
<td>0.2315</td>
<td>0.1667</td>
<td>0.1265</td>
<td>0.2257</td>
<td>0.0198</td>
</tr>
<tr>
<td></td>
<td>A5</td>
<td>Leader</td>
<td>0.1111</td>
<td>0.0926</td>
<td>0.0741</td>
<td>0.0926</td>
<td>0.1616</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.2315</td>
<td>0.2346</td>
<td>0.1204</td>
<td>0.1173</td>
<td>0.2160</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>Leader</td>
<td>0.4091</td>
<td>0.4798</td>
<td>0.1616</td>
<td>0.1818</td>
<td>0.1834</td>
<td>0.0296</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.4411</td>
<td>0.4254</td>
<td>0.1751</td>
<td>0.1706</td>
<td>0.2622</td>
<td>0.0326</td>
</tr>
<tr>
<td></td>
<td>A4</td>
<td>Leader</td>
<td>0.4762</td>
<td>0.7302</td>
<td>0.2540</td>
<td>0.5079</td>
<td>0.3618</td>
<td>0.0813</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.3855</td>
<td>0.3492</td>
<td>0.2018</td>
<td>0.1655</td>
<td>0.1989</td>
<td>0.0209</td>
</tr>
<tr>
<td></td>
<td>A6</td>
<td>Leader</td>
<td>0.4278</td>
<td>0.5667</td>
<td>0.2667</td>
<td>0.4000</td>
<td>0.3595</td>
<td>0.0229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Participant</td>
<td>0.4000</td>
<td>0.6815</td>
<td>0.3037</td>
<td>0.2148</td>
<td>0.2752</td>
<td>0.0296</td>
</tr>
</tbody>
</table>

The values of outdegree and indegree centrality of two types are larger than 0 in all teams. It means that most users not only transmit information but also receive information. In Table 5.4, you can also discover that outdegree centrality of leaders in all six teams is larger than indegree centrality. It shows that a leader play an important role of “transmitter (source)” instead of “receiver (destination)”. The statistic of a participant is contrary to a
leader. Most participants play a “receiver” role in all teams. The values of betweenness centrality of all users are relatively low. It indicates that most users have direct connection with others, and don’t rely on the third party.

Transitive rate is the proportion of group members that a user can communicate directly or indirectly. Reachable rate is the opposite concept. The more people can communicate with a team member, whether indirectly or directly, the larger value of reachable rate is. The highest value of transitive rate and reachable rate of six teams appeared in leaders of Team A4. Even it is the highest in all teams, some members played the isolated role in these nine weeks. If the community manager and the sponsor of training project can obtain the up-to-date team interaction information during the progress of project, they can warn the team in right time, and may improve team performance.

Social network analysis can display the status and relations of a team. Users can obtain a global view of the activity and relationship among team members. The team leaders can grasp the activity status and social relations of members on the fly. They can propose different strategies to different participants according to the measure of social network analysis. For a community manager, if they can capture social network attributes which influence team’s performance, they can direct suitable strategies in team development.

5.5 Classification and Rules Discovery

The clustering technique has discovered that the social network development has its impacts on teams’ achievement. However, these clusters only bring us the status information of virtual teams’ social networks at different time periods. By classification, we can predict a team’s performance from teams’ social network attributes. Classification is the technique usually used to analyze data sets and then classify these data to a rule set. After generating classification rules, users can apply these rules to predict class from unseen data. Therefore, the classification algorithm can predict the class of new instances based on
the training data. The classification software we used in this research is IBM’s Intelligent
Miner (IM). Intelligent Miner provides the fundamental technology and tools to support the
mining process. It provides an easy operation interface and good visualization for the
results of data mining. It also provides a convenient mechanism to prune the classification
tree, and display relative information of nodes and paths.

The data source of classification we use is the aggregated data of social network analysis
gathered during nine weeks and their final results. Because the time order of team progress
may affect the final performance, we also consider it as an attribute in addition to social
network attributes. We use the average and deviation of social network attributes and time
order as the input attributes, and the final result category as the class label. Average values
are average status of all team members and deviation values can show us what difference is
between team members in that attribute. Figure 5.3 is the visualization of classification
results, where the decision tree is pruned to depth five. The characteristics that may affect
the final result are prestige, closeness centrality, reachable rate, the deviation of indegree
centrality and the deviation of prestige. In this training project, the order of time has no
influence on the team performance classification. Table 5.5 is the detailed rules obtained
from the classification tree. There are 14 rules generated from Intelligent Miner’s decision
tree classification. Number 1-5 are the descriptions of left sub-tree and number 6-14 are
those of the right sub-tree. For example, we can interpret the classification results as
follows. Teams with indegree centrality deviation <= 0.0893, prestige <= 0.0711 and
reachable rate deviation <= 1.4673 results in low performance eventually. This rule can be
elaborated as follows, if members of a team, in general, receive little knowledge or cares
from each others directly or indirectly, the team performs poorly at end of project. The
other rule, teams with indegree centrality deviation > 0.0893 and closeness centrality > 0.1
and prestige <= 0.0495 and prestige deviation > 0.1316, explains teams with high
performance. This can be elaborated as follows, if members of a team have a certain degree
of closeness though large difference in receiving knowledge and cares from other teammates, the team may have high performance. These two rules are more general than other rules because they explain 70% teams’ final performance.

From the results of classification, we can see that most key characteristics that affect the final results are related with “indegree.” Indegree centrality emphasizes the numbers of direct inbound ties, reachable rate emphasizes the numbers of direct and indirect inbound ties and prestige not only focuses on the total numbers of inbound ties and their distance. As for closeness centrality, it considers the distance of indegree and outdegree and focuses on how close an actor is to all the other actors in the set of actors; therefore it considers the distance of indegree and outdegree at the same time.
Table 5.5 Rules of Classification Tree

<table>
<thead>
<tr>
<th>Number</th>
<th>Rule</th>
<th>Class</th>
<th>Records</th>
<th>Errors</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Indegree Centrality Dev. &lt;= 0.0893</td>
<td>B</td>
<td>23</td>
<td>3</td>
<td>87.0</td>
</tr>
<tr>
<td>2</td>
<td>Indegree Centrality Dev. &lt;= 0.0893 and prestige &lt;= 0.0711</td>
<td>B</td>
<td>21</td>
<td>1</td>
<td>95.2</td>
</tr>
<tr>
<td>3</td>
<td>Indegree Centrality Dev. &lt;= 0.0893 and prestige &lt;= 0.0711 and Reachability Dev. &lt;= 1.4673</td>
<td>B</td>
<td>20</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>Indegree Centrality Dev. &lt;= 0.0893 and prestige &lt;= 0.0711 and Reachability Dev. &gt; 1.4673</td>
<td>G</td>
<td>1</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>Indegree Centrality Dev. &lt;= 0.0893 and Prestige &gt; 0.0711</td>
<td>G</td>
<td>2</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>Indegree Centrality Dev. &gt; 0.0893</td>
<td>G</td>
<td>31</td>
<td>7</td>
<td>77.4</td>
</tr>
<tr>
<td>7</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &lt;= 0.1</td>
<td>B</td>
<td>2</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1</td>
<td>G</td>
<td>29</td>
<td>5</td>
<td>82.8</td>
</tr>
<tr>
<td>9</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &lt;= 0.0495</td>
<td>G</td>
<td>21</td>
<td>1</td>
<td>95.2</td>
</tr>
<tr>
<td>10</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &lt;= 0.0495 and Prestige Dev. &lt;= 0.1316</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>11</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &lt;= 0.0495 and Prestige Dev. &gt; 0.1316</td>
<td>G</td>
<td>19</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &gt; 0.0495</td>
<td>B</td>
<td>8</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>13</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &gt; 0.0495 and Prestige Dev. &lt;= 0.3064</td>
<td>B</td>
<td>4</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>Indegree Centrality Dev. &gt; 0.0893 and Closeness Centrality &gt; 0.1 and Prestige &gt; 0.0495 and Prestige Dev. &gt; 0.3064</td>
<td>G</td>
<td>4</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

5.6 Managerial Implications

In this chapter we have combined social network analysis and data mining to analyze task-based virtual teams in the professional cyber community. Clustering with teams’ social network attributes was used to examine the relation of teams’ performance and team members’ interaction pattern in the progress of team development. Values of different roles
in all teams were used to analyze the impact of roles that users play on the teams’ performance. Finally, we applied classification technique to classify example teams with various social network attributes, and found some preliminary rules that may affect teams’ performance. From the results of the three methods, we confirm that social network analysis is a useful means to analyze the knowledge activities conducted by virtual teams. Interaction patterns of members in the team and roles that users play have high influence on a virtual team’s development.

We expect that more and more virtual teams will be formed with the continuous growth of information technology. The interaction patterns are critical for teams’ collaboration work. Social network analysis is an effective method to assist these teams in developing a high cohesion group. We hope that the ideas and methods presented in this study will be proved useful in the analysis of virtual teams.

In this study, we only collected six teams’ information in the nine-week training project, and the data used to generate general classification rules for special interest groups in SCTNet is insufficient. We cannot say that the rules we find can apply to all SIGs in the community because the training data may not be enough. In the future, we hope to collect more valid example data to obtain more accurate prediction rules.
Chapter 6 Conclusion and Future Research

Relation in a team is an important factor to its performance. Stronger relational links are associated with higher task performance and the effectiveness of information exchange (Warkentin et al., 1997; Warkentin and Beranek, 1999). Walther and Burgoon (1992) also considered that strong relational links are associated with enhanced creativity and motivation, increased morale, better decisions and fewer process losses. Relation realization is important, but not all users can know it well. Therefore, this research is aimed at fulfilling such objectives as developing an online social network analysis and discovering interaction pattern of task-based teams.

In this thesis, a social network analysis system has been prototyped to facilitate virtual teams’ tasks in the professional virtual community, SCTNet. From ego-centered network, a team member can identify whom he or she has relations with, and how strong their relations are. Generalized from various sources of relation information from SIGs, such as discussion board, file sharing, guest book, e-mail, and online message, a team member can realize the role he or she within the team. Social network evolution trend enables the team leaders and participants to keep track team development status. Whole network displays a global view of the activity and relationship among team members. Team members can understand how information flows through network ties, how people acquire information and resources, and how cleavages and coalitions operate. A leader has insight into the structure and the participation of individuals in a team from whole network to identify roles that members play.

After implementing the system, 99 subjects filled in questions for evaluating the system. Evaluated by the four evaluation criteria, simplicity, utility, certainty, and novelty, we found that most subjects were satisfied with the system in of the four criteria. Team leaders and assistants have higher satisfaction than general participants. It implies that the system conforms with the cognition of team leaders and assistants since team leaders and assistants
interact frequently with team members in leading the team. From the results of questionnaire, most subjects assert that the social network analysis system can facilitate them to enhance the teams’ performance.

Besides building the social network analysis system, we analyze six task-based virtual teams to demonstrate the methods in discovering teams’ interaction patterns. These teams are built for an on-job training project. We combine social network analysis and data mining to discover their interaction patterns. From the discovered patterns, we found that team performance is influenced by the relationship and interaction of members, especially in later stage of team development. Moreover, interaction patterns of different roles that team members play highly influence a virtual team’s development, especially team leaders. A team leader always plays an important role of “transmitter (source)” instead of “receiver (destination)” and most participants play “receiver” roles in the team from this study. The characteristics that may affect the final result are prestige, closeness centrality, reachable rate, the deviation of indegree centrality and the deviation of prestige after classifying six task-based teams. Most key characteristics that affect the final results are related with indegree. It implies that successful teams have high degree of connectivity with members. However, in this research, we only collected six teams’ data and more teams’ data can be collected in the future to generate general classification rules. We hope to collect more valid example data to obtain more accurate prediction rules in the future.

Many directions can be further investigated in future research, such as

1. One research limitation is that we did not consider the property of the peer relation. We only know that two members have a relation, but we do not identify whether it is positive or negative. It will bring richer information by analyzing the relation using content analysis in future research.

2. Most subjects were not familiar with the domain knowledge of social network analysis. The concept is related to graph theory and it is uneasy for them as teachers to understand.
The mechanism to transform social network structure from abstract terms to common descriptions may help understand the relationship.

3. In this system, we only compute users’ relations in the same team. If they join many teams at the same time, their intersectional relations are not considered. In future research, the relation crossing different teams is an additional factor to investigate.

4. The method used to find subgroup in this research is SCC. SCC is a loose structure to find paths in the subgroup. In social network analysis, many other methods can be used to find subgroups, such as cliques, n-cluques, n-clans, k-plexs, k-cores. Different methods represent different characteristics of subgroups, and future research can include these methods to explain teams’ structures.

5. The social network analysis system was implemented with script language and CGI program. It doesn’t provide a good interaction interface for users to operate on the fly, especially in the graph visualization. A more flexible interactive interface can be developed by java applet in the future.

6. Only six teams’ data are used to discover interaction patterns of virtual teams. It needs more data to obtain more accurate classification rules in future research.
References


Haythornthwaite, C., Wellman, B., and Mantei, M. “Work relationships and media use: A


Sabidussi, G. “The centrality index of a graph,” *Psychometrika* (31), 1966, pp. 581-603


Seidman, S. B. and Foster, B.L. “A note on the potential for genuine cross-fertilization between anthropology and mathematics,” *Social Networks* (1), 1978, pp. 65-72

Seidman, S. B. “Network structure and minimum degree,” *Social Networks* (5), 1983, pp. 97-107


Wellman, B. and Wortley, S. “Different strokes from different folks: Community ties and social support,” American Journal of Sociology (96), 1990, pp. 558-588

Appendix

Appendix A.1 Announce of Evaluation

Appendix A.2 Step 1 of Evaluation
Appendix A.3 Step 2 of Evaluation
Appendix A.4 Step 3 of Evaluation

Appendix A.5 Step 4 of Evaluation