Chapter 15
The Application of Social Network Analysis: Case of Smart Roofing

Tugrul U. Daim, Monticha Khammuang and Edwin Garces

Abstract The use of social network analysis (SNA) becomes popular in social science research in the recent years. It is a practical application because it helps organizations to have better conceptualized and new understandings of the interactions. It could help organizations interpret and understand complexity, systems, pattern of changes, and structure of interactions. Moreover, SNA applications have been applied in many complicated fields to identify knowledge leaders in organizations, measure collaboration of teams, illustrate the hidden patterns of structure, and exploring the paths of interactions. In addition, many software programs were developed for personal or limited distribution by mathematicians, sociologists, graph theorists, and information technology specialists enabling SNA to facilitate the analysis of data and the creation of sociograms easier than before. Applying SNA in organizations could benefit many internal activities. It could help organizations to identify the group of experts for technology roadmapping (TRM) or R&D related activities, to know who the most appropriate expert for future collaboration may be, and to see the pattern of the interactions for future R&D planning. This chapter proposes an analysis of smart roofing using SNA to identify the group of experts, the interactions among experts, and the patterns of these interactions to help researchers to gain a better understanding of the current situation of smart roofing research and development programs and also to help them to prepare related future plans in order to promote the progress of smart roofing research and development programs.

Keywords Social network analysis · SNA · Practical approach · The application of SNA · The use of SNA

T.U. Daim (✉) · M. Khammuang · E. Garces
Department of Engineering and Technology Management, Portland State University, 900 SW 4th Avenue, Lower Level, Suite 50-02, Portland, OR 97201, USA
e-mail: tugrul.u.daim@pdx.edu

T.U. Daim et al. (eds.), *Anticipating Future Innovation Pathways Through Large Data Analysis*, Innovation, Technology, and Knowledge Management,
DOI 10.1007/978-3-319-39056-7_15

15.1 Introduction

The use of social network analysis (SNA) is new for the evaluation in social science. It has yet to be completely studied in this field. The use of SNA has gradually risen over the past ten to fifteen years (Durland and Fredericks 2005).

There are three factors related to the increasing usage of SNA. Firstly, SNA is a powerful tool that helps organization to have a clear understanding of interactions. After the Dot Com boom in 1990s, online networking tools for individuals to create and explore their personal and business networks grew up rapidly. Some companies mine data and sell data back to other business. Many companies such as IBM, Accenture, and Mars are also using SNA to determine the influencers, the relationships among teams/projects, and patterns of interaction among teams (Durland and Fredericks 2005).

SNA was developed for understanding the complexity and system of the networks so it could help organizations to do the evaluation of designs and program development effectively because it could explain the complexity and interactional nature of structures (Durland and Fredericks 2005).

15.2 Basic Concepts in SNA and Their Relation to Expert Identification

15.2.1 What Is Social Network Analysis?

Social network analysis (SNA) is a general approach for investigating social structures or networks and the relationship among them (Wellman and Berkowitz 1988). It represents a concept of social structure in terms of a network connecting members and channeling resources together. Moreover, it focuses on the characteristics of the network rather than on the characteristics of the individuals and point out a group of networks as personal communities (Wetherell et al. 1994). It also focuses on individual actors making alternatives without considering the behavior of others. This approach neglects the social context of the actor. SNA considers the relationships between actors as the first priority, and individual properties are second priority (Knoke and Kuklinski 1982). Moreover, another important function of SNA is to study how structural regularities influence actors' behavior (Knoke and Kuklinski 1982). It is clear that original application of SNA is to investigate the interactions and we could classify the investigation into two major patterns: the ego network and the global network. The first one is to analyze the network of one person. The second one is to analyze all relations among the participants in that network (White 2000).

SNA has two very unique characteristics which differentiates it from other analysis tools. Firstly, it helps to describe and understand relational data better than others because of its own set of measures and analysis tools. Relational data represent a relationship between two components and also the value of that relationship. SNA focuses on the social context and behavior of relationships among actors rather than on the rational choices individual actors make so this characteristic differentiates SNA from other methodologies (Durland and Fredericks 2005). Another unique character is that other methodologies result in an understanding of importance or the significance of the correlations whereas SNA gives a path into the complexity that often starts with a small thing that opens up into something much bigger. This does not mean that SNA has a never-ending path into an analysis, but it provides many more hidden points into the function of programs (Durland and Fredericks 2005). By the way, the results from SNA are more complex, and the significance from what we found cannot be measured with one statistic.

15.2.2 What Are the Components of the Network?

The main components of the networks are actors and their relations. Actors can be called nodes, vertices, or points. Relations can be called as arcs, edges, or ties (Scott 1988). The picture of components of the network can be seen in Fig. 15.1.

Points in Fig. 15.1 are used to represent the actors, and lines are used to represent relations.

15.2.3 How Many Types of Relations Are There Within a Network?

Generally, relations can be divided into two groups: directed or undirected. In a directed relation (see in Fig. 15.2), the edges have their own direction or arrow. This kind of edges is similar to arcs and can be used as ordered pairs of vertices. We

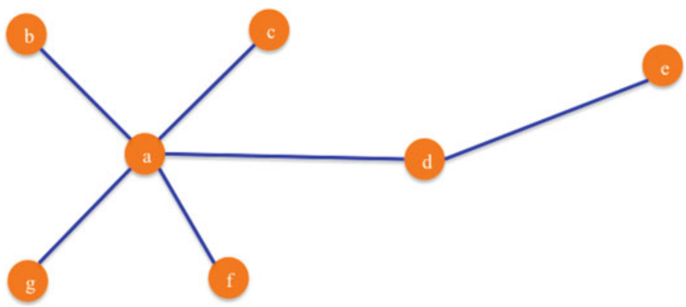

Fig. 15.1 Components in a network

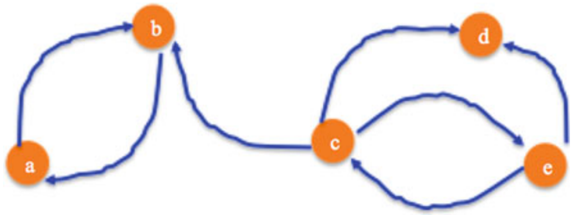

Fig. 15.2 Directed relation

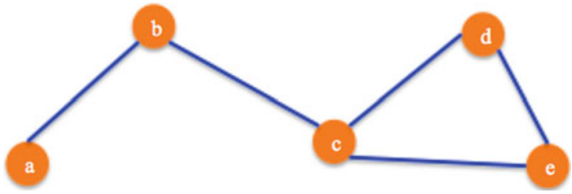

Fig. 15.3 The undirected relation

use directed relations to show relational phenomena that has a sense of direction (Borgatti and Everett 1997).

In an undirected relation (see in Fig. 15.3), the edges have no order pairs. This kind of relation is used where there is no direction or the direction does not make sense or it is not clear about exactly the direction (Borgatti and Everett 1997).

For general research, we expect that the networks have different structures and have their unique form of relations.

15.2.4 What Are the Types of Network Data?

Data used in SNA consist of an array/table of measurements. The rows of the array could represent the cases, subjects, or observations. Each cell of the array shows a relationship between the actors. We could see the example of network data in Table 15.1, which describes the network of friendship relations among four people (Hanneman and Riddle 2005).

The difference between conventional data and network data is that conventional data focus on actors and attributes, whereas network data focus on actors and

Table 15.1 Example of social network data

Who reports liking whom?				
Chooser	Choice			
	Bob	Carol	Ted	Alice
Bob	–	0	1	1
Carol	1	–	0	1
Ted	0	1	–	1
Alice	1	0	0	–

relations. This difference causes the researchers to design the research before collecting data in order to conduct sampling, develop measurement, and handle the resulting data.

We can separate social network data into two groups: ***1-mode and 2-mode***. The first group shows edges based on directed contact between actors in the network, and all of the nodes are of the same type such as people, organizations, and ideas, whereas the 2-mode data show nodes from two different classes and ties are across classes (Borgatti and Everett 1997; Hanneman and Riddle 2005).

15.2.5 How to Classify Level of Network for Investigation?

We can classify the level of networks into three levels: ego network, partial network, or global network.

Ego network: This level of network focuses on the individual, rather than on the whole network (see in Fig. 15.5). At this level, we collect information from the connections where the actors are connected to each focal ego. This information is useful for researchers because it could enable them to see the incomplete picture of the whole network and to understand how networks affect individuals. However, researchers can obtain only some information from ego network level. In ego networks, we cannot measure the overall density of the population. If we have some reasonable explanation to explain about alters in terms of their social roles, rather than as individuals, ego networks can tell us more about their local social structures such as alters connected to an ego by a friendship relation as kin, co-worker, member of the same church, co-author, etc., see in Fig. 15.5 (Hanneman and Riddle 2005).

Partial network: There are some cases that we cannot track down the full network. This partial network is an alternative approach to begin with a selection of focal egos and identify connected egos. Then, we determine the first stage of egos connected to one another. This partial network approach is suitable for collecting data from very large populations. For instance, we collect data of female university students about their close friends and ask them to identify which of their friends know one another. This partial network approach could give us a clear and reliable overall picture of networks in which individuals are embedded. This kind of data could be very useful to help us understand the opportunities and constraints ego has as a result of the path they are related in their networks. Moreover, this kind of network also gives us some information about the network as a whole. It represents micro-network data sets or a sampling of local areas of larger networks (Hanneman and Riddle 2005).

Global network or complete network: This kind of network (see in Fig. 15.4) focuses on multiple attributes of actors and also multiple kinds of ties that connect

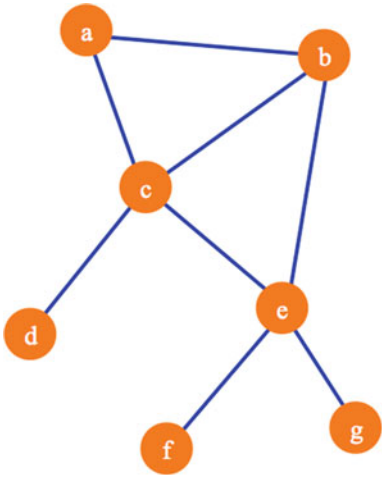

Fig. 15.4 The global network

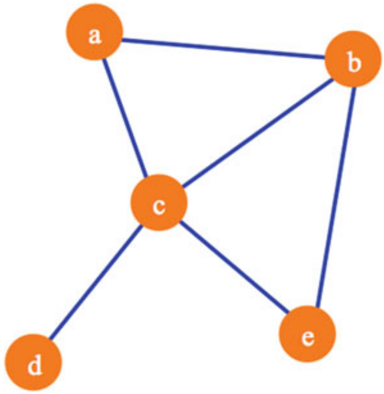

Fig. 15.5 The ego network of "ego c"

actors in a network. For instance, we might want to know which faculty have the same group of students, serve on the same community, and have one or more fields of expertise and co-author in common. These actors might be tied together closely in one relational network; however, they might be quite far from one another in a different relational network. The establishment of actors in multi-relational networks and the topology of networks combined of multiple relations are the most interesting part of SNA. When researchers collect data about relations among actors, we are trying to sample from a population of possible relations. Network correlation, multi-dimensional scaling and clustering, and role algebras are related to the study of global network or complete network data (Fig. 15.6) (Hanneman and Riddle 2005).

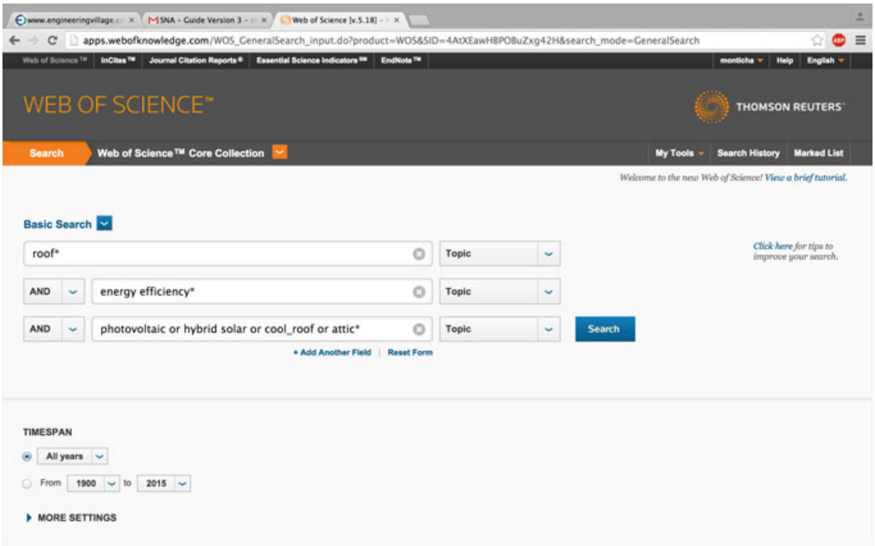

Fig. 15.6 Web of Science search screen with specific keywords

15.2.6 *How to Identify Network Structure?*

Researchers need to identify network structure because it shows the characteristic and the cohesion of that network. We could use "***Connectivity***" to identify the structure of the network (Hanneman and Riddle 2005).

Connectivity: This term is used to explain how actors in one part of the network are connected to another actor of other part of the network or how two actors are connected to each other in undirected network data. Actors and their connections play important role in SNA so it is necessary to begin to investigate the networks by examining these very connectivity properties. Firstly, we should look at the whole network and then focus on the number of actors, the number of possible connections, and the number of actual connections. The differences in the size of networks and how the actors are connected could tell us about human populations. Population size is one of the most critical factors in sociological analysis. The connection of a small group is different from a large group in many ways. The ways they are connected to each other could be a key indicator of the cohesion, solidarity, moral density, and complexity of the network. Individuals and networks have different basic demographic features. Individual actors might have many or few ties. Individuals might be the source of ties or might be the actors that receive ties, but do not send them, or might be both. The number and kinds of ties that actors have are critical factors to determine how much they embed in the network, what constraints related to their behavior exist, what the range of opportunities is, or how much influence and power they have (Hanneman and Riddle 2005).

In order to analyze connectivity, researchers could use “reachability, density, distance, a path/a walk/a cycle, number of walks, and flow”:

Reachability. It is used when a target actor is reachable by another. In general, if the data are directed, it is possible that actor A might be able to reach actor B, but actor B might be unable to reach actor A. In undirected data, each pair of actors might be able or unable to reach to one another. If there is the case of unreachability in a network, there might be the potential of a division of the network or it could be interpreted that the target population is composed of more than one subpopulation (Hanneman and Riddle 2005).

Density. The density is used to indicate the level of connectedness of a network. It is calculated by using the number of lines in a graph divided by the maximum number of lines (in case that every author is connected to every other one). Consequently, its value is between 0 and 1. For example, if the value of the density of the central network is 0.05, this network is very loose and is not a dense network at all (Otte and Rousseau 2002).

Distance. It is the distance between an actor and others. It is used to capture how individuals are embedded in networks. Knowing number of actors stay at various distances from each actor is very important in order to understand the differences among actors in terms of their limitations and advantages. Sometimes there are multiple paths between two actors. Multiple connections may indicate a stronger connection between two actors than a single one. The distances among actors in a network could be a critical macro-characteristic of the whole network. If the distances are great, it may take a long time for information to diffuse across population or it could be that some actors are quite unaware of, and influenced by others even if they are technically reachable. The variance across the actors in the length that they have from other actors could be a basis for differentiation and stratification. Actors who are closer to more others may be able to put more power than those who are more distant (Hanneman and Riddle 2005).

A path/a walk/a cycle. A path is a sequence of nodes and edges: starting with one node and ending with another node. It also represents the tracing of the indirect connection between the two nodes. On a path, it is impossible to go backward or revisit the same node twice, whereas *a walk can be* any sequence of nodes and edges and it is possible to go backwards on a walk. A path which starts and ends at the same node is named *a cycle* (Otte and Rousseau 2002).

Number of walks/paths. This metric is used to count how many linkage actors have been compared to one another. These data provide a way of thinking about the strength of ties or relations. Actors connected at short distances might have stronger connections if they are connected many times or even if they have many more number of paths. The numbers of walks/paths could be found by raising the matrix to that power. These differences help researchers to understand how information moves in the network, which actors have stronger power, and also other important properties (Hanneman and Riddle 2005).

Flow. This metric is used to identify the movement of information from actors to actors. It is used to answer how many different actors in the neighborhood lead to a target. Flow also helps to assess the strength of the ties (Otte and Rousseau 2002).

15.2.7 How to Identify Key/Central Nodes in Network?

Researchers need to identify key/central nodes in network of the networks because these nodes show the key players of that network. The term used to identify key nodes in a network is called "***Centrality***" (Freeman 1979).

Centrality refers to location, indicating where an actor resides in a network. This term could help to formalize intuitive notions about the distinction between insiders and outsiders. In order to analyze connectivity, researchers could use "degree centrality, closeness, betweenness centrality" as basic centrality elements (Freeman 1979; Borgatti 2005).

- *Degree centrality*. It is the number of connections that a node has with other nodes. For example, having higher degree of centrality means that this scientist has collaborated with many colleagues. Moreover, we could measure the degree centrality of the whole network. Lower degree of centrality of the whole network indicates that many authors in this network are not connected to each other (Freeman 1979).
- *Closeness*. Another way of analyzing centrality is using the closeness factor. This closeness indicator is more general than the degree centrality, because it includes the structural position of actors in the whole network. A high value of closeness for an actor means that actor is related to all others through a small number of paths (Freeman 1979).
- *Betweenness*. This indicator relied on the number of shortest paths passing through an actor. Actors who have a high value of betweenness seem to play important role of connecting different groups or they might have higher power in communication, communication control, and communication flow than others (Borgatti 2005).

15.2.8 What Is SNA Process?

According to Otte and Rousseau (2002), Hansen (2009), the process of SNA can be separated into 3 steps.

Step 1: Designing of the analysis

Researchers need to define the objective and clarify the scope of the analysis. They also need to determine what kinds of networks and what kinds of relations they want to study. Moreover, they need to formulate the hypotheses and research questions to set the right path to the analysis (Otte and Rousseau 2002).

Step 2: Collecting network data

There are two main methods to collect data for SNA. Firstly, we could use questionnaires and interviews to collect data about the relationships within a specific group. In this case, researchers need to gather background information such as using interviewing senior managers and key staff to understand specific needs and hidden issues. This way is suitable for organizations to identify the relationship among teams/projects, the flow of information among teams/project, or the most influencer in the network. After that, researchers need to develop the survey/interview methodology, design the questionnaire, and survey/interview the individuals/teams/and units in the target network. Secondly, researchers could gather data from academic websites such as for Web of Science or Compendex. They could get the data about the authors and the co-authors in order to identify relationship between authors/coauthors. This method is suitable for knowledge management, collaboration, or other academic purposes. After that, researchers could go directly to that website and search information by using the specific keywords in order to collect target data. These keywords are very important because website/program will gather the article/journal from giving keywords. If keywords are not that specific, the result might be too large and difficult to use for analysis, but if keywords are too specific, the result might be too small and we cannot find the relationships in the target network (Otte and Rousseau 2002; Hansen 2009).

Step 3: Measuring network data and analyzing network data

Researchers could use SNA software to calculate the basic terms used for SNA and create the visualization of the network. The examples of software are (Huisman and Van Duijn 2005);

- Pajek (Windows, free)
- UCInet (Windows, shareware)
- Netdraw (Windows, free)
- Mage (Windows, free)
- GUESS (with all platforms, freeware)
- R packages for SNA (with all platforms, freeware)
- Gephi (with all platforms, freeware).

15.2.9 Application of SNA in Organizations

There are many advantages of applying SNA. One of the biggest advantages of SNA is that it can be visualized using the appropriate tools very clearly. This leads to a deeper understanding of the structures and relationships of a network. The analysis of social networks focused on the fact that a relation between persons and a relation between organizations are important, because they make and display attitudes, communication, and information flow of products. SNA provides the methods to investigate these relationships, to represent graphically, evaluating and building on that to develop it further (Krause and Croft 2007). Organizations could apply SNA in many business activities such as they could apply SNA in merger acquisition assessment in order to control the network management with strategic alliances and collaborations. In this field, they could measure how successful the integration after mergers is. They also can apply SNA to identify how good the network is connected, how stable the network is, and whether or not there are any holes in the network (Krause and Croft 2007).

We can apply SNA to find the experts in order to build communities of practice, tighten the internal knowledge, and build up the knowledge management system by survey and analyze leaders' opinions to identify who has the most influence, what the organizational causes of conflicts are, or how efficient the information is (Fritsch and Kauffeld-Mon 2010). Last but not least, viral marketing like word of mouth (Staab 2005), procurement, supply chain management, and human resource development also can use the application of SNA as well (Staab 2005).

15.2.10 The Importance of Expert Identification

It is critically important for organizations to identify experts within and outside the organization as they need their help to manage tacit knowledge within organization in an effective way. This kind of knowledge is usually transferred better through mentoring or face-to-face interactions among those internal experts. In a very highly competitive business environment, managing explicit knowledge is also important. Organizations need this kind of knowledge to create innovation within organization by transferring knowledge from external experts. With this combination between tacit and explicit knowledge, organizations could gain the true value added intellectual assets of an organization and they could maintain the organization's core competency and innovate (Yang and Huh 2008). It is also important to identify the right experts for collaboration purposes or to develop strategic technology road maps (Müller et al. 2012; Daim and Oliver 2008).

15.3 Methodology

In this research, the methodology was divided into three steps: data collection using web of science, expert identification using RStudio and Gephi, and expert data analysis.

Step 1: Data collection using Web of Science. Because we need expert data in smart roofing field, we need to collect expert data from reliable multi-databases and easy-to-access sources so we chose to use Web of Science as our database. Web of science is an online scientific citation database which provides a comprehensive citation search. It gives access to multiple databases and allows us to do in-depth exploration of academic fields.

After choosing the source, then we have to create “Keywords” used to search for smart roofing data. The size of the data is very important for the expert analysis. If data are too large, the analysis might be too rough. We might be unable to analyze the whole network, but if data are too small, we might be unable to get any significant results and may not find the right network of experts. That is why choosing the right size of data is very important for this kind of research.

The strategy we used to identify keywords was to “***search with generic keywords and then narrow down using specific keywords.***” When we searched by using generic keywords, we got more than 1000 data points, and then, we narrowed down by using specific names. Finally, we got down to 80 data points, which was suitable for our analysis. For example in the case of Web of Science, system will run the query by starting to look for articles that contain “roof with any following word” in article’s topic, and then, it will start to search in article’s title and collect articles that contain photovoltaic or hybrid solar or cool roof or attic in their title from a time period which in this research we selected as being from 1900 to 2015. The search screen we used can be seen in Figs. 15.6, and 15.7.

After that, we have to go directly to each article and check whether the article is exactly related to our topic or not by clicking on each topic, to see the detail and delete the unrelated articles. After we check and clean all data, we could save the data with full record and cited references and be ready to move to the next step.

Step 2: Expert identification using RStudio and Gephi. In this step, we will interpret the data we got from previous step by using RStudio with Shiny package. RStudio is a free and open source programming language used for statistical computing and graphics, whereas Shiny package is a package used to calculate

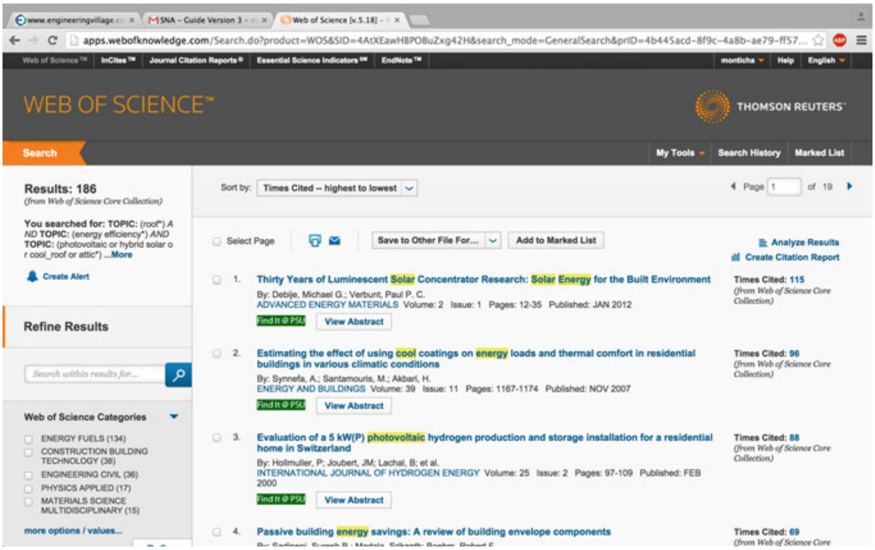

Fig. 15.7 Web of Science using specific keywords

SNA basic elements such as betweenness, degree, closeness, and centrality. This package is developed in house at the authors' institution, and it supports for databases from Web of Science/Compendex. The way the RStudio with Shiny package was used can be seen in Figs. 15.8, 15.9, and 15.10.

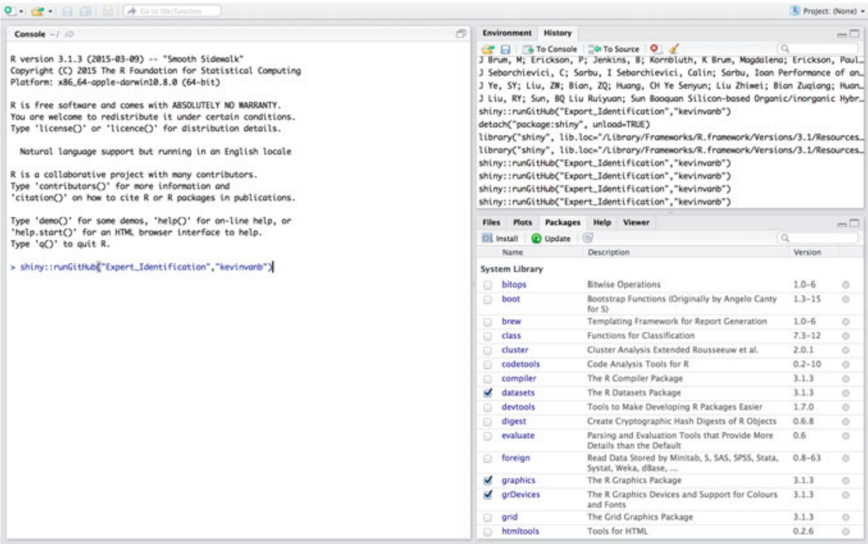

Fig. 15.8 RStudio with shiny package

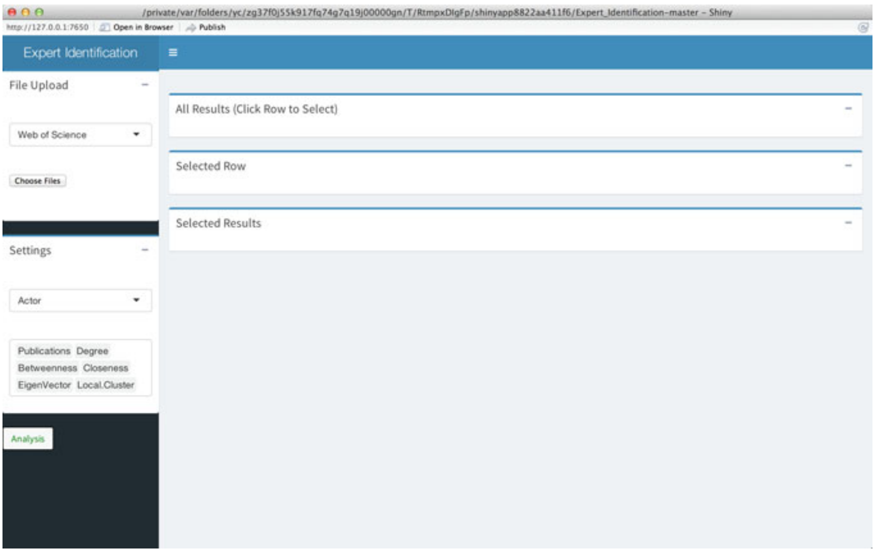

Fig. 15.9 RStudio with shiny package

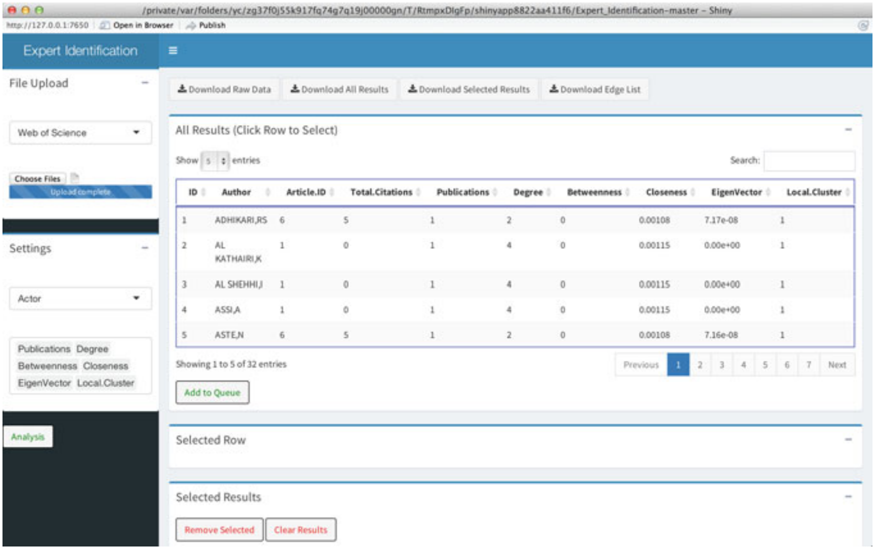

Fig. 15.10 RStudio with shiny package

At the end of this step, we could calculate important terms used for SNA analysis such as betweenness, degree, closeness, and centrality. The result of this step can be seen in Appendix A.

After we got this result, we did the visual analysis using Gephi, which is a free program used to analyze and visualize network data. There are two data sets we used for Gephi. One data set is the one with all the data (186 entries) and the other one is the data set with only the most top ten experts in this field. The way the Gephi was used can be seen in Figs. 15.11 and 15.12.

Step 3: Expert data analysis. After completing steps 1 and 2, we analyze the results to identify the groups of experts for smart roofing, the relations between experts, the structure of this network, and also the centrality of the experts.

15.4 Results and Discussion

In this section, we will analyze two groups of results to identify the experts in smart roofing field. The first group of results is the centrality-related metrics including degree centrality, closeness, number of citations, and betweenness. These results will help us to understand more about the key experts in the network. The second group of results is the expert network pictures or visualizations. These results will help us to better understand the structure of the network, structure of the sub-networks, and the network flow.

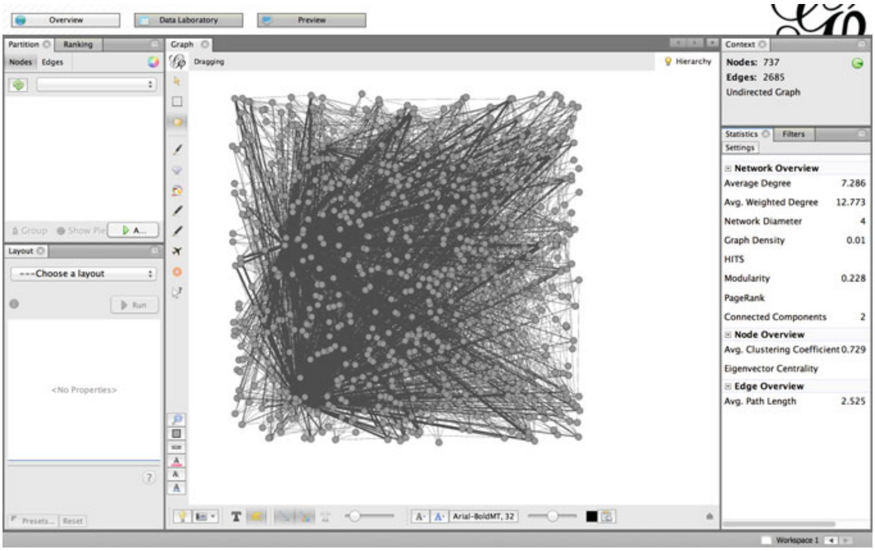

Fig. 15.11 Working screen of gephi program

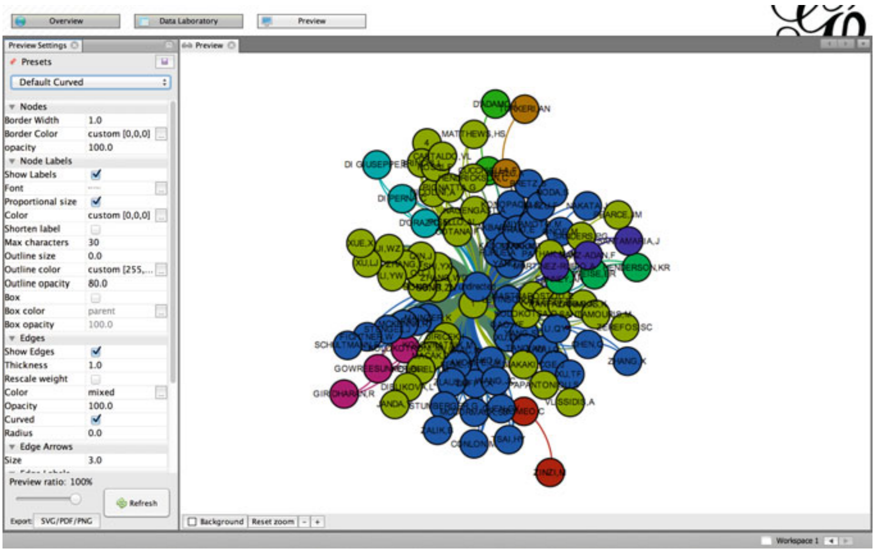

Fig. 15.12 Visualization data created by gephi

15.4.1 *Centrality Analysis*

15.4.1.1 Degree Centrality

Table 15.2 shows the degree of centrality from Web of Science database. This data set is ranked from the largest value of degree centrality to the lowest one. According to these data, we could see that LEVINSON-R, JI-WZ, LI-YW, QIN-J QU-J, SHI-YX, SONG-JR, SONG-ZN, XU-LJ, XUE-X, ZHANG-T, and ZHANG-WD have the highest values of degree centrality. It means that these people have stronger collaborations than the others. They might be collaborating with many colleagues. We might assume that they may know many experts in this field or they might have had more opportunity to work, share, or communicate with other experts compared to those who have lower values of degree centrality.

15.4.1.2 Closeness

Table 15.3 shows the closeness values from Web of Science database. This data set is ranked from the highest value of closeness to the lowest one. According to these

Table 15.2 Degree of centrality results (top 20) from Web of Science database

ID	Author	Closeness
1	LEVINSON, R	0.000103
2	JI, WZ	0.000103
3	LI, YW	0.000103
4	QIN, J	0.000103
5	QU, J	0.000103
6	SHI, YX	0.000103
7	SONG, JR	0.000103
8	SONG, ZN	0.000103
9	XU, LJ	0.000103
10	XUE, X	0.000103
11	ZHANG, T	0.000103
12	ZHANG, WD	0.000103
13	GAO, YF	0.000103
14	GE, J	0.000103
15	TANG, XM	0.000103
16	XU, JM	0.000103
17	XU, TF	0.000103
18	YANG, SC	0.000103
19	ZHOU, Q	0.000103
20	AKBARI, H	0.000103
21	BRETZ, S	0.000103
22	KONOPACKI, S	0.000103

Table 15.3 Closeness (top 22) from Web of Science database

ID	Author	Closeness
1	LEVINSON, R	0.000103
2	JI, WZ	0.000103
3	LI, YW	0.000103
4	QIN, J	0.000103
5	QU, J	0.000103
6	SHI, YX	0.000103
7	SONG, JR	0.000103
8	SONG, ZN	0.000103
9	XU, LJ	0.000103
10	XUE, X	0.000103
11	ZHANG, T	0.000103
12	ZHANG, WD	0.000103
13	GAO, YF	0.000103
14	GE, J	0.000103
15	TANG, XM	0.000103
16	XU, JM	0.000103
17	XU, TF	0.000103
18	YANG, SC	0.000103
19	ZHOU, Q	0.000103
20	AKBARI, H	0.000103
21	BRETZ, S	0.000103
22	KONOPACKI, S	0.000103

data, we could see that all of these 22 people have the highest value of closeness (0.000103). While this value is not so high, these people would be the ones to connect to establish collaborations.

15.4.1.3 Betweenness

Table 15.4 shows the betweenness values from Web of Science database. This data set is ranked from the highest value of betweenness to the lowest one. According to these data, we could see that *LEVINSON-R, KOLOKOTSA-D, YANO-A, COTANA-F, PISELLO-AL* have the highest value of closeness (21, 20, 14, 4, 4). It means that they have number of shortest paths passing through them. These people seem to play an important role of connecting different groups or they might have higher power in communication, communication control, and communication flow than others in the same network.

By comparison of Tables 15.2, 15.3, and 15.4, it can be seen that *LEVINSON-R* is the only one that has the highest rank in every table. According to these data, we could assume that he might be the most influential expert in this network because he stays close to other experts, he knows many experts, and he also has the shortest

Table 15.4 Betweenness results (top 16) from Web of Science database

ID	Author	Betweenness
1	LEVINSON, R	21
2	KOLOKOTSA, D	20
3	YANO, A	14
4	COTANA, F	4
5	PISELLO, AL	4
6	SANTAMOURIS, M	2
7	GOBAKIS, K	0.667
8	KARLESSI, T	0.667
9	MASTRAPOSTOLI, E	0.667
10	PANTAZARAS, A	0.667
11	ZEREFOS, SC	0.667
12	BRINCHI, L	0.2
13	NICOLINI, A	0.2
14	CASTALDO, VL	0.2
15	PIGNATTA, G	0.2
16	ROSSI, F	0.2

Table 15.5 Comparing data among degree centrality, closeness, and betweenness

ID	Author	Degree	ID	Author	Closeness	ID	Author	Betweenness
1	**LEVINSON, R**	10	**1**	**LEVINSON, R**	0.000103	**1**	**LEVINSON, R**	21
2	JI, WZ	10	2	JI, WZ	0.000103	**2**	**KOLOKOTSA, D**	20
3	LI, YW	10	3	LI, YW	0.000103	**3**	**YANO, A**	14
4	QIN, J	10	4	QIN, J	0.000103	4	COTANA, F	4
5	QU, J	10	5	QU, J	0.000103	5	PISELLO, AL	4
6	SHI, YX	10	6	SHI, YX	0.000103	6	SANTAMOURIS, M	2
7	SONG, JR	10	7	SONG, JR	0.000103	7	GOBAKIS, K	0.667
8	SONG, ZN	10	8	SONG, ZN	0.000103	8	KARLESSI, T	0.667
9	XU, LJ	10	9	XU, LJ	0.000103	9	MASTRAPOSTOLI, E	0.667
10	XUE, X	10	10	XUE, X	0.000103	10	PANTAZARAS, A	0.667
11	ZHANG, T	10	11	ZHANG, T	0.000103	11	ZEREFOS, SC	0.667
12	ZHANG, WD	10	12	ZHANG, WD	0.000103	12	BRINCHI, L	0.2
13	**KOLOKOTSA, D**	9	13	GAO, YF	0.000103	13	NICOLINI, A	0.2
14	**YANO, A**	9	14	GE, J	0.000103	14	CASTALDO, VL	0.2
15	FURUE, A	7	15	TANG, XM	0.000103	15	PIGNATTA, G	0.2
16	HIRAKI, E	7	16	XU, JM	0.000103	16	ROSSI, F	0.2

paths passing. Moreover, KOLOKOTSA-D and YANO-A are other influential experts, because they got higher values of betweenness and degree centrality. This can be seen in Table 15.5.

15.4.1.4 Number of Citations

Table 15.6 shows the number of citations from Web of Science database. This data set is ranked from the highest value of citation number to the lowest one. According to these data, we could see that *COTANA-F and PISELLO-AL* have the highest values of citations number (41). It means that their articles have been cited by other experts for many times. We can assume that their work is significant and also related to other experts' work. They might be the ones developing the general theory and could be used as the basic knowledge or reference for other experts; or their work might be easy to apply to different fields.

By comparison of Tables 15.5 and 15.6, it can be seen that *LEVINSON-R* seems to be the most important export in this network too because his work has been cited by many experts (Table 15.7).

15.4.1.5 Overall

According to all information we got (degree centrality, closeness, betweenness, and citations number), we could come up with the ranking in Table 15.8.

Table 15.6 Number of citations from Web of Science database

ID	Author	Total citations
1	COTANA, F	41
2	PISELLO, AL	41
3	AYOMPE, LM	38
4	CONLON, M	38
5	DUFFY, A	38
6	MCCORMACK, SJ	38
7	LEVINSON, R	33
8	AKBARI, H	32
9	BRETZ, S	32
10	KONOPACKI, S	32
11	KOLOKOTSA, D	31
12	CUCCHIELLA, F	29
13	D'ADAMO, I	29
14	SANTAMOURIS, M	17
15	GIRIDHARAN, R	15
16	GOWREESUNKER, BL	15
17	KOLOKOTRONI, M	15
18	DIAKAKI, C	14
19	PAPANTONIOU, S	14
20	VLISSIDIS, A	14

Table 15.7 Comparison of data among degree centrality, closeness, betweenness, and citations number

ID	Author	Degree	ID	Author	Closeness	ID	Author	Betweenness	ID	Author	# Cited
1	**LEVINSON, R**	10	**1**	**LEVINSON, R**	0.000103	**1**	**LEVINSON, R**	21	1	COTANA, F	41
2	JI, WZ	10	2	JI, WZ	0.000103	2	KOLOKOTSA, D	20	2	PISELLO, AL	41
3	LI, YW	10	3	LI, YW	0.000103	3	YANO, A	14	3	AYOMPE, LM	38
4	QIN, J	10	4	QIN, J	0.000103	4	COTANA, F	4	4	CONLON, M	38
5	QU, J	10	5	QU, J	0.000103	5	PISELLO, AL	4	5	DUFFY, A	38
6	SHI, YX	10	6	SHI, YX	0.000103	6	SANTAMOURIS, M	2	6	MCCORMACK, SJ	38
7	SONG, JR	10	7	SONG, JR	0.000103	7	GOBAKIS, K	0.667	7	**LEVINSON, R**	33
8	SONG, ZN	10	8	SONG, ZN	0.000103	8	KARLESSI, T	0.667	8	AKBARI, H	32
9	XU, LJ	10	9	XU, LJ	0.000103	9	MASTRAPOSTOLI, E	0.667	9	BRETZ, S	32
10	XUE, X	10	10	XUE, X	0.000103	10	PANTAZARAS, A	0.667	10	KONOPACKI, S	32
11	ZHANG, T	10	11	ZHANG, T	0.000103	11	ZEREFOS, SC	0.667	11	KOLOKOTSA, D	31
12	ZHANG, WD	10	12	ZHANG, WD	0.000103	12	BRINCHI, L	0.2	12	CUCCHIELLA, F	29
13	KOLOKOTSA, D	9	13	GAO, YF	0.000103	13	NICOLINI, A	0.2	13	D'ADAMO, I	29
14	YANO, A	9	14	GE, J	0.000103	14	CASTALDO, VL	0.2	14	SANTAMOURIS, M	17
15	FURUE, A	7	15	TANG, XM	0.000103	15	PIGNATTA, G	0.2	15	GIRIDHARAN, R	15
16	HIRAKI, E	7	16	XU, JM	0.000103	16	ROSSI, F	0.2	16	GOWREESUNKER, BL	15
17	ISHIZU, F	7	17	XU, TF	0.000103				17	KOLOKOTRONI, M	15
18	KADOWAKI, M	7	18	YANG, SC	0.000103				18	DIAKAKI, C	14
19	MIYAMOTO, M	7	19	ZHOU, Q	0.000103				19	PAPANTONIOU, S	14
20	NODA, S	7	20	AKBARI, H	0.000103				20	VLISSIDIS, A	14
			21	BRETZ, S	0.000103						
			22	KONOPACKI, S	0.000103						

Basic centrality elements show that Levinson-R is the expert who has the highest value of every centrality element except total citations number. From this analysis, we could say that Levinson-R is the most important expert in this network. He might be the key of this network. Moreover, Kolokotsa-D and Yano-A could be the other two of the most important and influential experts in this network due to the high values of degree centrality, citations number, and betweenness.

15.4.2 Visualization

15.4.2.1 Structure of the Whole Network

The overall picture of this network with 186 data entries is shown in Fig. 15.13. It is an undirected graph with 575 nodes and 1950 edges. The average degree centrality of this network is 3.391 (max 10, min 0), the network diameter is 3, the density of the network is 0.006, and the average path length is 1.084.

It is difficult to see the relationships among specific experts in Fig. 15.13. So we narrow data down by focusing only on the group of specific experts we identified through the previous step and used that data to generate the visualization again with Gephi. We did this because we wanted to understand relationships among the specific experts better. The resulting visualization is shown in Fig. 15.14.

Figure 15.14 is an undirected graph with 62 nodes and 88 edges. The average degree centrality of this network is 2.839 (max 10, min 0), the network diameter is 4, the density of the network is 0.047, and the average path length is 3.198.

From the above graph, we could see that the nodes have four different colors. These colors are determined by modularity, which is another measure of the structure of networks. It is used to represent the strength of division of a network into nodes. We could call them clusters or communities of the network. We could see that there are four clusters or modularity within this network: blue, red, yellow, and green. The green one is the biggest cluster, whereas the yellow one is the smallest cluster in this network. It means that there are many experts in the green cluster and may be these people are connected to more people than experts in other clusters.

Moreover, we could see that there are six nodes that have strong edges (*Cotana-F, Pisello-AL, Yano-A, Kolokatsa-D, Levenson-R, and Zhang-WD*). These people seem to have higher power than others, and they might be able to influence other people because they are in the center or core of the network. They are connected to too many experts, and also they stay close to other strong nodes. This could imply that the most important experts in this network are the nodes that have strong edges. In order to see the picture clearly, we use Fig. 15.15.

Table 15.8 The most important experts in this network analyzed by author

ID	Author	Degree	Total citations	Betweenness	Local cluster
1	LEVINSON, R	10	33	21	0.533
2	KOLOKOTSA, D	9	31	20	0.389
3	YANO, A	9	13	14	0.611
4	COTANA, F	6	41	4	0.467
5	PISELLO, AL	6	41	4	0.467
6	SANTAMOURIS, M	6	17	2	0.733
7	ZHANG, T	10	8	0	1
8	ZHANG, WD	10	8	0	1
9	AKBARI, H	3	32	0	1
10	BRETZ, S	3	32	0	1
11	KONOPACKI, S	3	32	0	1

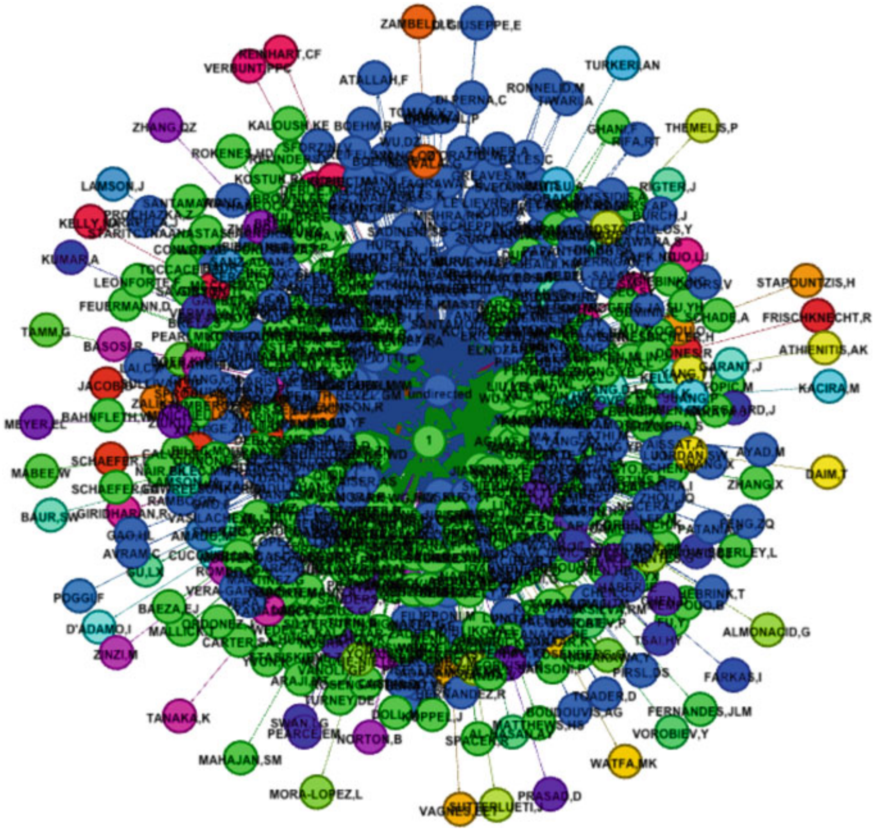

Fig. 15.13 The overall picture of the whole network with 186 data entries (all data)

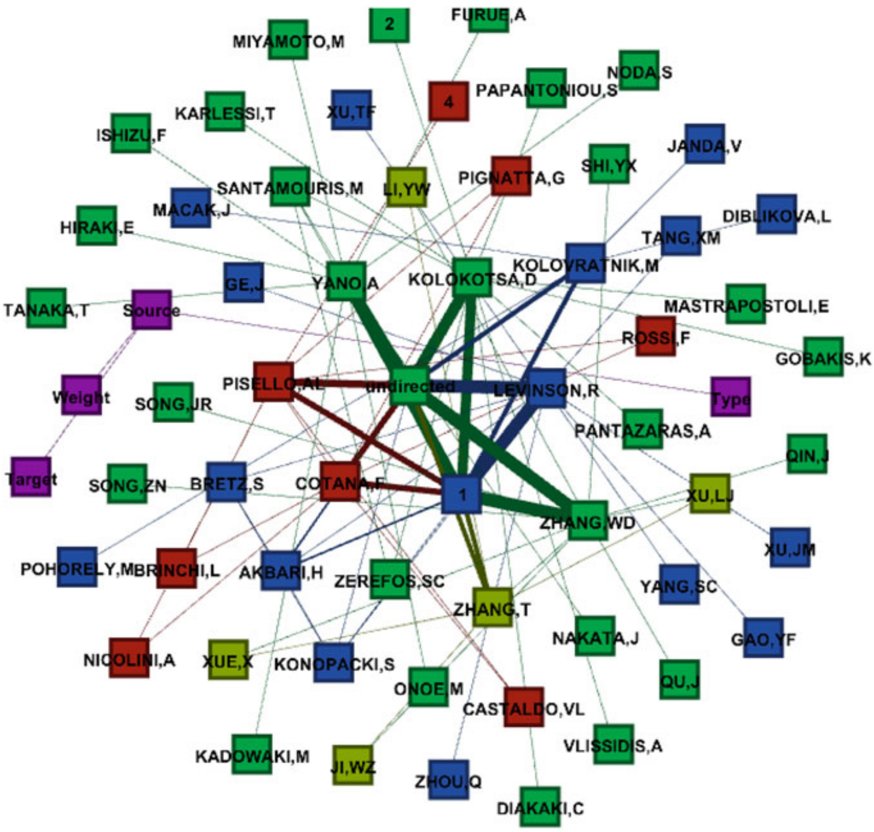

Fig. 15.14 The overall picture of the key experts

15.4.2.2 Structure of Each Clusters Within Network

As we mentioned above, there are four different clusters of this network: green, red, blue, and yellow (divided by modularity value).

- ***Cluster with green nodes***. The green nodes are the highest in number in this network. If we look deeply into the graph, we could see that there are three small clusters embedded in this green cluster (see in Figs. 15.16, 15.17, and 15.18).

From these above three graphs, we could see that there is a core or central node inside every green cluster. The core node of the 1st green cluster is *Kolokatsa-D*, the core node of the 2nd green cluster is *Zhang-WD,* and the core node of the 3rd green cluster is *Yano-A*. Without these core nodes, experts in green nodes cannot connect to each other and form a cluster. That is why the core or central node is very important because they help to form the network by connecting other nodes.

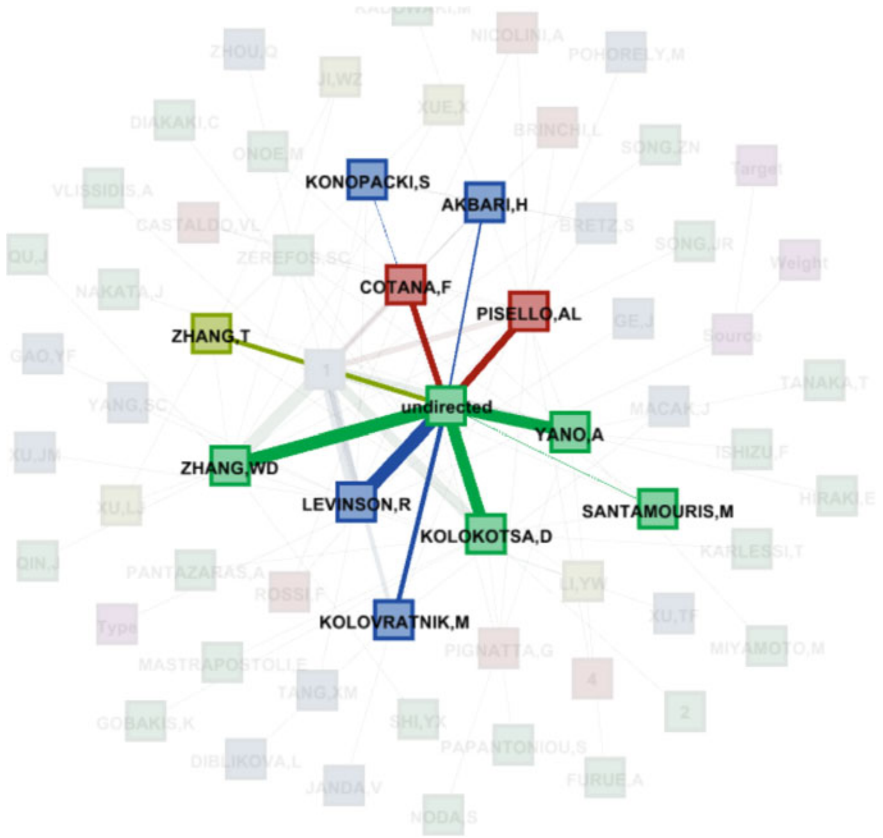

Fig. 15.15 The nodes with strong edges in the whole network

Without them, communication cannot pass to other experts and cluster cannot be formed.

In addition, *Kolokatsa-D and Yano-A* are linked to each other and these two people help to connect the 1st green cluster together with the 3rd green cluster and make the network larger than before. This linkage can benefit all in many aspects. They expand the size of the network and increase the possibility of receiving more information. However, this might slow the communication rate within the network.

- ***Cluster with blue nodes***. This cluster has medium number of nodes. If we look carefully into the graph, we could see that there are two small clusters embedded in this blue cluster (see in Figs. 15.19, and 15.20).

The core node of the 1st blue cluster is *Kolovratnik-M*, and the core node of the 2nd blue cluster is *Levnson-R*. Without these core nodes, experts in blue nodes cannot connect to each other and form a cluster. However, it seems like Levnson-R

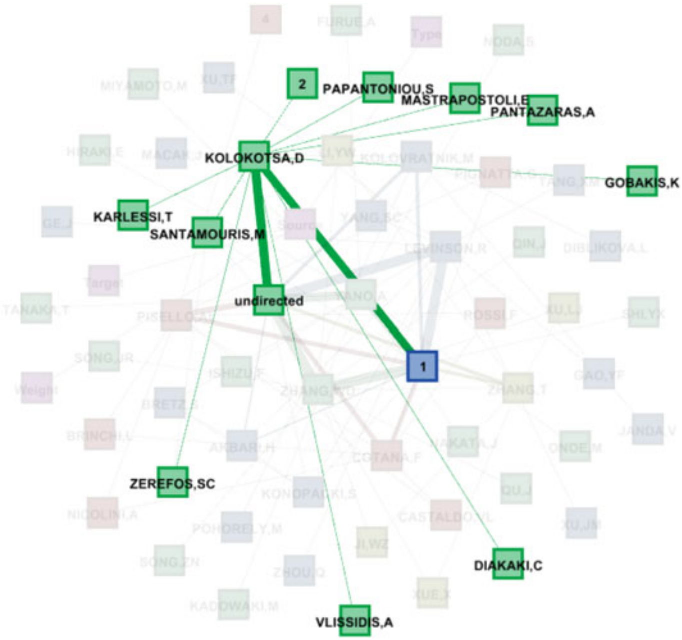

Fig. 15.16 Shows 1st small *green* cluster with Kolokotsa-D node

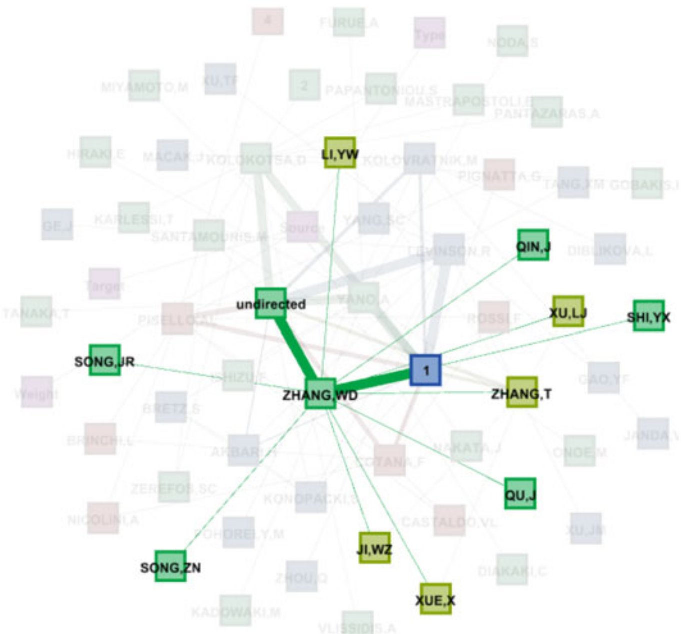

Fig. 15.17 Shows 2nd small *green* cluster with Zhang-WD node

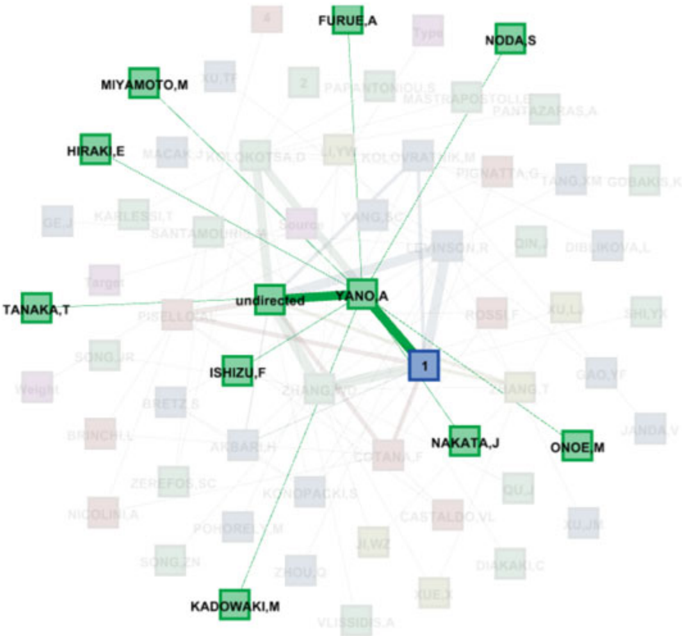

Fig. 15.18 Shows 3rd small *green* cluster with Yano-A node

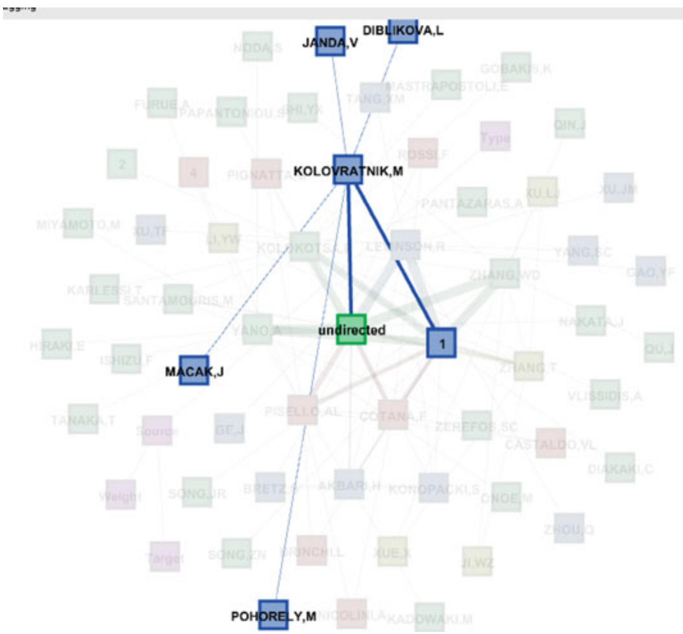

Fig. 15.19 Shows 1st small *blue* cluster with Kolovratnik-M

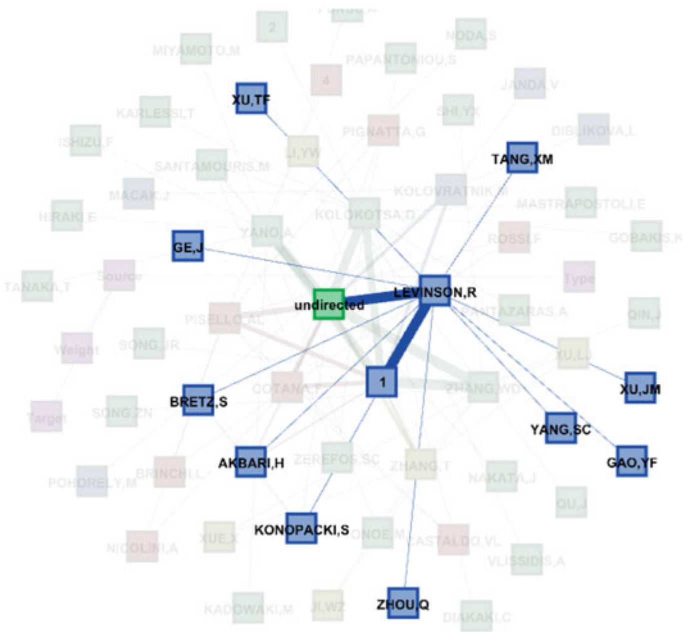

Fig. 15.20 Shows 2nd small *blue* cluster with Levnson-R

has more power than Kolovrntnik-M because he is connected to many more nodes than Kolovrntnik-M.

- ***Red cluster***

This group of cluster has a small population. There is only one red cluster. The picture of red cluster can be seen in Fig. 15.20.

From above graph, we could see that *Piesello-AL* is the core or central node of this red cluster.

- ***Yellow cluster***

This group of cluster has also a small population and has the similar pattern with red cluster. There is only one yellow cluster, which is shown in Fig. 15.21.

From the above graph, we could see that *Zhang-T* is the core or central node of this yellow cluster.

15.4.2.3 Relations Among Clusters in Network

We tried to analyze the relationships among these four clusters but what we found is that there is none except the relation between green cluster and yellow cluster (see in Fig. 15.22). We found that Zhang-T, yellow node, and Zhang-WD, green node,

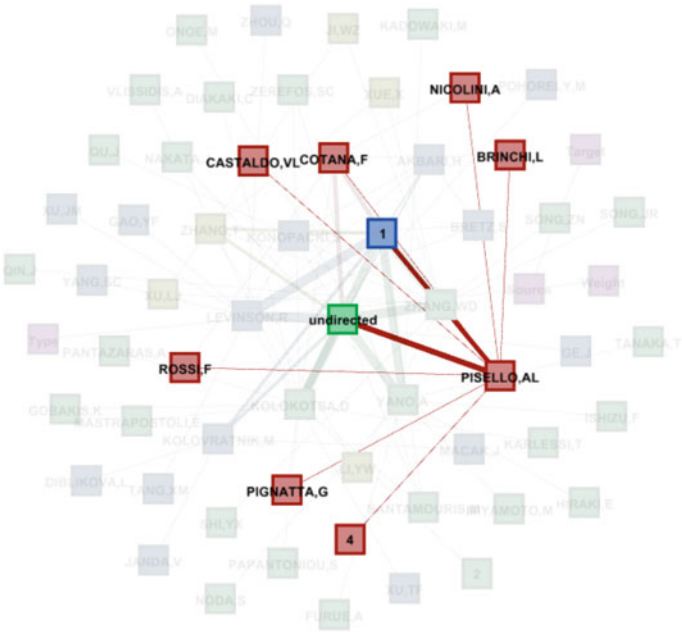

Fig. 15.21 Shows *red* cluster with Piesello-AL

are connected to each other so they bridge yellow cluster and green cluster together. They connected these two clusters together and made the clusters a bigger cluster. Without them, yellow cluster will stay separate from the network.

15.4.2.4 Group of Important Experts Identify by Using Visualization

Based on all information we got from visualization, we could identify the top ten powerful experts in Table 15.9.

And from the visualization, we could see that Levinson-R, Kolokotsa-D, and Zhang-WD are the most three powerful experts in this field because they are connected to too many experts. Moreover, Zhang-WD also helps to connect the yellow and green cluster together.

When we compared results from using centrality elements (Sect. 4.1) and visualization (Sect. 4.2), we found that they give slightly different insights.

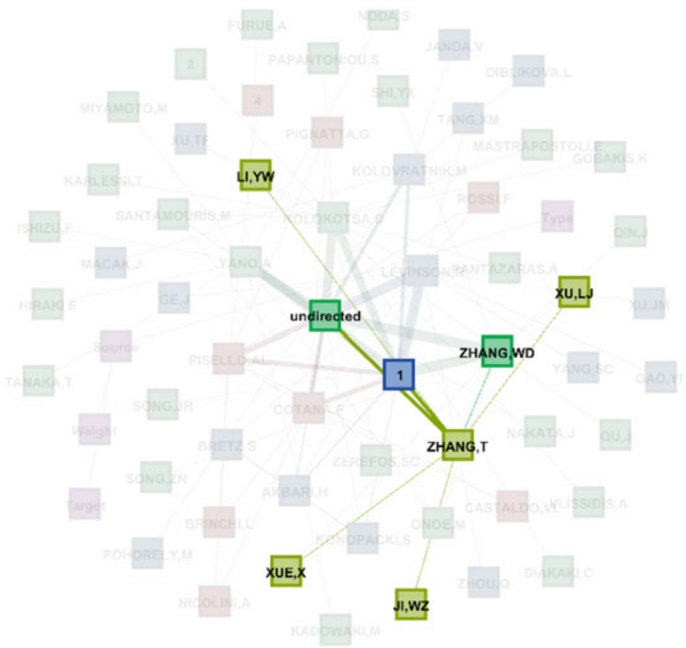

Fig. 15.22 Shows *yellow* cluster with Zhang-T

Table 15.9 Shows the most important experts using visualization analyzed by author

ID	Author	Reason
1	LEVINSON, R	He has strong edges and he is the central node of blue cluster
2	KOLOKOTSA, D	He has strong edges and he is the central node of green cluster
3	YANO, A	He has strong edges and he is the central node of green cluster
4	COTANA, F	He has strong edges
5	PISELLO, AL	He has strong edges and he is the central node of red cluster
6	ZHANG, T	He is the central node of yellow cluster and he also connect yellow cluster together with green
7	ZHANG, WD	He has strong edges. He is the central node of green cluster and he also connect green cluster together with yellow
8	KOLOVRTINI-M	He is the central node of blue cluster

15.5 Conclusion and Future Research

Social network analysis is a good analysis tool that can be applied in many fields. This chapter shows one of the specific examples of SNA application that can be used in the real world. We can identify who the important experts in the field are. With the visualization, we could explain the picture of the network, the pattern of relations, the structure of the network, and also the linkage between clusters.

References

Borgatti, S. P. (2005) Centrality and network flow. *Social Networks*, *27*(1), 55–71.

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social networks, 19*(3), 243–269.

Daim, T. U., & Oliver, T. (2008). Implementing technology roadmap process in the energy services sector: A case study of a government agency. *Technological Forecasting and Social Change, 75*(5), 687–720.

Durland, M. M., & Fredericks, K. A. (2005). An introduction to social network analysis. *New Directions for Evaluation, 2005*(107), 5–13.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215–239.

Fritsch, M., & Kauffeld-Mon, M. (2010). The impact of network structure on knowledge transfer: An application of social network analysis in the context of regional innovation networks. *The Annals of Regional Science, 44*(1), 21–38.

Hanneman, R. A., & Riddle, M. (2005) *Introduction to social network methods* (pp. 1–115).

Hansen, D. L. (2009). Do you know the way to SNA? A process model for analyzing and visualizing social media data. University of Maryland Tech Report: HCIL-2009-17, pp. 1–10.

Huisman, M., & Van Duijn, M. A. J. (2005). Software for social network analysis. *Models and Methods in Social Network Analysis*, 270–316.

Knoke, D., & Kuklinski, J. H. (1982). *Network analysis, Sage University paper series on quantitative applications in the social sciences* (pp. 7–28).

Krause, J., Croft, D. P., & James R. (2007) Social network theory in the behavioral sciences: potential applications. *Behavioral Ecology and Sociobiology*, 15–27.

Müller, M. O., Groesser, S. N., & Ulli-Bee, S. (2012). How do we know who to include in collaborative research? Toward a method for the identification of experts. *European Journal of Operational Research*, *216*(2), 495–502.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of information Science, 28*(6), 441–453.

Scott, J. (1988). Social network analysis. *Sociology, 22*(1), 109–127.

Staab, S. (2005). Social networks applied. *Intelligent Systems, IEEE, 20*(1), 80–93.

Wellman, B., & Berkowitz, S. D. (1988). *Structural analysis in the social sciences 2, social structures: A network approach* (pp. 62–82). Cambridge: Cambridge University Press.

Wetherell, C., Plakans, A., & Wellman, B. (1994). Social networks, kinship, and community in Eastern Europe. *Journal of Interdisciplinary History, 24*, 639–663.

White, H. D. (2000). Toward ego-centered citation analysis. In B. Cronin & H. B. Atkins (Eds.), *The web of knowledge* (pp. 475–496). Information Today, Medford: NJ.

Yang, K.-W., & Huh, S.-Y. (2008). Automatic expert identification using a text categorization technique in knowledge management systems. *Expert Systems with Applications, 34*(2), 1445–1455.