Expert Systems with Applications 38 (2011) 15159–15171

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Finding research trend of convergence technology based on Korean R&D network

Jae Hoon Jin [a], Sang Chan Park [b,1], Chong Un Pyon [c,*]

[a] Department of Industrial and Systems Engineering, KAIST, 373-1 Kusung-Dong, Yuseong-Gu, Daejeon 305-701, South Korea
[b] College of Business Administration, Kyung Hee University, 1 Hoegi-dong, Dongdaemun-gu, Seoul, South Korea
[c] Department of R&D Policy, Korea Institute of Machinery and Materials, 156, Gajungbukno, Yuseong-Gu, Daejeon 305-343, South Korea

ARTICLE INFO

Keywords:
Social network
Convergence technology
R&D
Network construction

ABSTRACT

Traditionally, collaboration network or citation network is used to answer the old question how scientists or engineers interact with each other. This paper introduces a R&D network to make up the missing aspect of the traditional approaches about using multi-sources and to find out the trend of convergence technology R&D in Korea. We collect data about human resources and national R&D projects from Korean national R&D databases, and then construct a weighted network between experts by using meta-data mapping and the network folding technique. And we apply Newman's grouping algorithm that is generalized to a weighted network for detecting the community structure of the network. Gathering data from multi-sources is useful to reveal the structure of network rather than to use only one database. Lastly, we perform a network analysis to examine important experts. The result shows significant information about research trend and core experts in Korea. We expect this study will be helpful in three ways: (1) how to make a network from heterogeneous multi-sources, (2) how to figure out the current situation of convergence technology R&D, (3) how to discover who are important people in Korean convergence technology R&D network. And this paper is just a cornerstone of the work to investigate the current situation of national R&D projects in Korea.

1. Introduction

A social network is a set of individuals connected through socially meaningful relationships, such as friendship, co-working, or information exchange (Wasserman, Faust, Iacobucci, & Granovetter, 1994; Wellman, 1996). The Social network theory traditionally views social relationships in terms of nodes and links (Wasserman et al., 1994), where the nodes are the individual actors within the networks and the links are the relationships between the actors. The strength of links between nodes in a real-world social network is an important theoretical issue. The probability of two friends of an individual knowing each another is much greater than the probability of two people chosen randomly from the population knowing each another (Guare, 1990; Newman, Watts, & Strogatz, 2002; Watts & Strogatz, 1998). Actors with strong links usually have some sort of common grounds on which they establish their relationships (Preece, 2002; Wellman & Gulia, 1997), and thus often constitute a subgroup. Because of the common grounds, actors with strong links – and hence representing a subgroup – often share common interests, needs, or services; that is the necessity of subgrouping (Preece, 2002; Schwartz & Wood, 1993; Wellman & Gulia, 1997).

One part of social network researches is about research and development (R&D) networks, especially about how scientists or engineers interact with each other. These works are represented in the form of the collaboration network or the citation network that is the network of scientific collaboration, as documented in the papers scientists write. A collaboration network is a kind of affiliation network of scientists in which a link between two scientists is established by their coauthorship of one or more scientific papers (Newman, 2001a, 2001b). For the construction of network, both the existence and the times of coauthoring (or citation, acknowledgement) are important. And the times of coauthoring, for sure, implies some information about "how close they are and how easily networked with each other". So to fully characterize the interactions in real networks, the weights of links should be considered.

In spite of many researchers' studies on many types of collaboration networks, most of them are concerned about just some groups of people, a number of journals, and a few affiliations. There are practical difficulties to cover more large area, such as a country level, at once: (1) data are distributed in many other databases, (2) types of data are heterogeneous. Generally, previous researchers use one database to get data for the construction of network. For example, Newman (2001a, 2001b) used The Los Alamos E-print

* Corresponding author. Tel.: +82 42 868 7828; fax: +82 42 868 7051.
E-mail addresses: jinjae@kaist.ac.kr (J.H. Jin), sangchan@khu.ac.kr (S.C. Park), pcu@kimm.re.kr (C.U. Pyon).
[1] Tel.: +82 2 961 0494; fax: +82 2 961 0515.

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.05.088

Archive, Medline, Stanford Public Information Retrieval System, and Networked Computer Science Technical Reference Library. Huang et al. (2003) and Singh (2005) used United States Patent and Trademark Office database. These databases are so huge, even they are single databases, that sometimes researcher could not deal with all data, but only a part of them. Nevertheless, we can assume that if we could use rich data and information collected from many databases, we will be able to construct more useful network to show the structure and the effects underlying a specific phenomenon. However, the traditional methods to make a collaboration network and citation network are hard to support to construct a network from multi-sources.

Technological convergence is the tendency for different technological systems to evolve towards performing similar tasks (Wikipedia). Convergence can refer to previously separate technologies such as voices, data and videos that now share resources and interact with each other, synergistically creating new efficiencies. The phenomenon of convergence occurs when innovations emerge at the intersection of established and clearly defined industry boundaries, thereby sparking off an evolutionary development with much broader impact (Hacklin, Marxt, & Fahrni, 2009). In recent industry developments within information technology (IT), bio-technology (BT) and nano-technology (NT), the convergence of technologies and knowledge bases has induced a variety of industrial points of inflection. Hence, industry boundaries have become blurred, and the innovation does not take place within previously existing industrial silos anymore, but rather between them (Hacklin et al., 2009).

In the case of USA, government agencies including National Nanotechnology Initiative (NNI) and National Science Foundation (NSF) invest about $130 billion per year in NBIC (NT, BT, IT, and Cognitive Technology) convergence technology R&D. EU makes intensive investment in IT, BT, NT related technologies through FP7 (EU Seventh Framework Programe) program, except the individual investment plan of each member country; 69.9% of total R&D budget is assigned to them. The Japanese government administers development projects including Protein 3000, Millennium Research for Advanced Information Technology (MIRAI), and Exploratory Research for Advanced Technology (ERATO) for IT, BT, NT, and convergence technology R&D. Likewise, in Korea, the government chose convergent industry based on IT, BT, NT as growth engine for next generation, and started to make an effort to foster them. Now relevant authorities, including MEST (Ministry of Education, Science, and Technology), MKE (Ministry of Knowledge Economy), drive programs assigned as the role. The Korean government has a plan to invest 1,500 billion for later 10 years for acquiring core cutting link technologies and fostering growth potential. As a consequence, during past several years, thousands of national R&D projects related with convergence technology have carried out in Korea according to the own technical roadmaps. It is no wonder that the performance of each project has been reported, but there lacks the view of accomplished convergence technology R&D by now and explanations from the researchers' viewpoint.

In this paper, we present a study of the construction of Korean R&D network from multi-sources and various analyses about the network. (The term, R&D network, is a little ambiguous, but now we use it to mean a social network about experts who participate in Korean national R&D projects.) We collected data, especially related with IT, BT, and NT, from three Korean national databases. The National R&D human resource database gives the information about researcher, who involved in national R&D project, including his/her name, age, affiliation, and major fields. The National R&D project information database gives details of projects ordered by government authorities. The National Digital Science Links database gives research results of scientists in the form of papers and reports. Then we integrated them as a kind of scientist-research field matrix using meta-data mapping by National Standard Classification of Science and Technology. A network folding technique was used to convert expert-research filed matrix into expert-expert matrix. This expert-expert matrix was used as an adjacency matrix of the social network of experts. According to these consecutive steps, Korean convergence technology R&D network to analyze could be acquired. We expect this study will be helpful in three ways: (1) how to make a network from heterogeneous multi-sources, (2) how to figure out the current situation of convergence technology R&D (what convergence technologies are developing or not in Korea), (3) how to find important people in Korean convergence technology R&D network.

In the next section, we will briefly review related works of R&D network, including collaboration and citation network, and network clustering. Methodologies used to construct R&D network are presented in Section 3. Section 4 is devoted to a description of the empirical study. In Section 5, the findings of constructed network and clustering are introduced. Finally, Section 6 shows the conclusion of the study and research directions for the future.

2. Literature review

2.1. Research about R&D network

There are many researches about individual networks about science and technology. These networks lead scientific and technical progress. First, about scientific network, Moed et al. investigate the results of a study of "bibliometric" (publication and citation) for university research policy (Moed, Burger, Frankfort, & Van Raan, 1985). Nederhof and van Raan examine the productivity and impact of major economics research groups with the influence of key scientists by sensitivity analysis and its contribution (Nederhof & van Raan, 1993). Secondly, about the technological networks Eric von Hippel explores a novel type of cooperative R&D: the informal trading of proprietary know-how between rival (and non-rival) firms (Eric von von von von Hippel, 1987). He explores whether and when technology trading between direct competitors is an economically advantageous form of cooperative R&D. Hagedoorn and Schakenraad study the major international networks of inter-firm alliances with identification of major players within information technology and its sub-fields (Hagedoorn & Schakenraad, 1992). Almeida and Kogut investigate the technological networks about its regional variations in the localization of spillovers (Almeida & Kogut, 1999). They examine the relationship between the mobility of major patent holders and the localization of technological knowledge by the analysis of paper citations and show the influence of local transfer of knowledge by the interfirm mobility of engineers.

Besides to the research about the individual networks, many researchers examine the co-evolution of science and technology. Tijssen's co-classification analysis yields quantitative measure of the level of interdisciplinarity, the strength of interdisciplinary relations between fields, and the graphical representation (that is a map) of the interdisciplinary structure in the single field by the co-occurrence of different subject-classification headings assigned to research publications (Tijssen, 1992). Dosi and Kogut describe co-evolution of organizations and technologies to explain how firm-specific and nation-specific capacities are related to international competitiveness (Dosi & Kogut, 1993). McKelvey addresses the distinctions between 'basic science' and 'technological development' to conceptualize the environments for the evolution of science-based technologies (McKelvey, 1997). Narin et al. exhibit the contribution of public science to industrial technology by tracing citation linkage between US patents and scientific research papers (Narin, Hamilton, & Olivastro, 1997).

Up to now, these researchers use the databank of paper citation of journals and the patent information for the study of co-evolution. However, when the co-evolution of science and technology is taken for granted, the R&D network constructed based on these databank should be improved about its elaboration and information. So we are willing to use multi data sources for the construction of R&D network. To our knowledge, no one has explored the R&D network with multi sources.

2.2. Methodologies of network grouping

There are many approaches to construct networks and detect community structures which are some division of the network. They are broadly classified into two categories according to the addition or deletion of edges to or from the network; the agglomerative approach and divisive approach (Scott, 2000).

In an agglomerative approach, after calculation of similarities between vertex pairs, the edges are iteratively added to an initially empty network starting with the vertex pairs with highest similarity. The similarities are measured through such measurements as distance, density, frequency, centrality. Centrality can be measured in terms of degree, closeness, density, betweenness (Hanneman, 1998; Scott, 2002). Degree of centrality is the measure of point centrality by counting the number of nodes one node is adjacent to (Nieminen, 1974). Closeness is the measurement of global centrality in term of the "closeness" of all the nodes in the group or network by measuring the path distance (Freeman, 1979; Freeman, 1980). The betweenness is the measure of influence of a node over the flow of information between other nodes (Freeman, 1977).

In the divisive approach, which tries to overcome the poor results of agglomerative approach, the edges that are most "between" communities are focused rather than the edges that are most "central" to communities. Girvan and Newman propose an alternative approach by generalization of Freeman's betweenness centrality to edges and define the edge betweenness of an edge (Girvan & Newman, 2002). They apply the method to a collaboration network of scientists and found two types of communities; scientists grouped by similarity either of research topic or methodology. And then, they define a measure of the quality of a particular division of network to distinguish the better structure, which is the modularity (Newman & Girvan, 2004).

Usually, the agglomerative approach has the tendency to find only the cores of communities and leave out the peripheries, although they have an obvious community membership. So we choose the divisive approach of Girvan and Newman for networking community structures.

2.3. Binary network vs. weighted network

The method of Girvan and Newman that we choose for networking community structures assigns binary memberships to nodes. The most group detecting algorithms as well as Girvan and Newman's assign a binary membership. It is not compatible with the nature where the degrees of the strength for the links are various. Under the binary membership, the information about the intensity of the link cannot be expressed.

In order to complement this drawback, many researchers propose a weighted network on behalf of the binary network. Newman addresses the problem by mapping weighted networks onto multi-graphs with little or no modification of unweighted networks (Newman, 2004). He derives an appropriate generalization of the algorithm of Girvan and Newman to weighted networks. Li. et al. introduce a clustering coefficient of weighted networks (Li et al., 2005). Davis and Carley propose a stochastic model and group detecting algorithm for social networks, namely FOG, which uses fuzzy, overlapping groups (Davis & Carley, 2008). It generates grouping from link data using a stochastic model of link emission from group entities and a maximum-likelihood clustering method. In this paper, we choose the Newman's method to weighted network because his results show various examples from the very simple case to the more complex cases for application.

3. Methodology

3.1. Data collection

Korean Ministry of Education, Science, and Technology provides open national R&D information and knowledge portal services named NTIS (National Science and Technology Information Service, http://www.ntis.go.kr) to enhance the efficiency during entire R&D cycles from planning to application. This service is connected with 15 authorities to execute national R&D projects internally, so that provides information about finished or on-going R&D projects including projects themselves, human resources, equipments/materials, and performances.

This system consists of several sub-systems operated independently and we used two of them, human resource and project. As stated in Section 1, national R&D human resource database gives the information such as project participant's name, age, affiliation, and major fields; among them, we focused major fields. The database keeps each expert's one major field and two or three sub-major fields. (The project participants belong to various organizations, for example, university, national laboratory, company laboratory, and so on. So we call them experts briefly from now on.) Generally, sub-major fields may follow a taxonomy system, in this case, Korean National Standard Classification of Science and Technology. Moreover, the sub-major field is able to work as a medium between experts because two experts, if they have the same sub-major field, may be closer than the case when they do not have any same sub-major field. If two have perfectly the same sub-major fields, they seem to be interested in the same research theme. So, there is more possibility for them to cowork together and have a close relationship.

In order to investigate who is related with the convergence technology, we are interested in the experts whose sub-major fields cover interdisciplinary areas, especially IT, BT, and NT. If an expert has one IT related sub-major field and one BT related sub-major field, it reveals he/she is a professional about two different fields and has a high fitness to research BIT convergence technology. We use a category-based approach to select a subset of the national R&D human resource database available online. First, we select IT, BT, NT related categories from Korean National Standard Classification of Science and Technology. This taxonomy system, established in 2005, has a three-depth; the number of categories for each level is 19, 178, 1235 from the highest to the lowest. We choose 29 mid level categories such as 'Genetic Engineering', 'Nano Materials', 'Wireless and Mobile Communication', then perform a cross search within 9 high level categories to which mid level categories belong to. Within the initial search results, an expert whose performance is under 10 papers or under 5 projects is eliminated for the practical reason. Finally, we get data of 817 experts with the form of following:

$$\begin{aligned}&\mathit{Expert}(\mathit{Name}, \mathit{Age}, \mathit{Affiliation}, \mathit{Major\ Field}, \mathit{Sub\text{-}major}\ 1,\\ &\quad \mathit{Sub\text{-}major}\ 2, \mathit{Sub\text{-}major}\ 3)\end{aligned}$$

As a next step, we collect data of projects performed by selected experts from 2003 to 2008 by searching national R&D project information database. For the data is collected in October 2009, and the information of 2009 projects that are not finished yet is abandoned. Total 5967 projects are found and arranged with the form of following:

$$Project(Supervisor, Project\ Name, Serial\ Number, Standard\ Technical\ Classification)$$

The database is incomplete, so it has many missing values and half-descriptive classifications. We have to fix faults using meta-data mapping with some reasonable assumptions based on the information we can get from keywords in a project name or other classifications included.

Lastly, we collect the information of papers each expert wrote from National Digital Science Links database. Unfortunately, direct access to the database was not available, same as human resource and project case; we have to search online individually. Therefore, we do not use this information to make a typical collaboration network, but we add an attribute of expert for chasing one's research flow. For the sake of convenience, we limit the range of search to a number of important experts who are revealed after a network construction.

3.2. Meta-data mapping

Metadata is "data about data". An item of metadata may describe an individual datum, or a content item, or a collection of data including multiple content items and hierarchical levels, such as a database schema. In data processing, metadata provides information about other data managed within an application. This commonly defines the structure or schema of the primary data.

A meta-data mapping tool that is able to define the structure is needed to integrate data collected from multi-source. In this case, Korean National Standard Classification of Science and Technology is the tool to integrate human resource data and project data from the view point of technology. We define an expert-technology matrix which gives the strength between an expert and his/her interested or skilled technology fields.

$$\text{ET Matrix} = [w_{ij}], \quad \text{where } w_{ij} \text{ is a strength between the expert } i \text{ and the technology field } j$$

The row of matrix is the ID of an expert, and the column is the ID of technology based on Korean National Standard Classification of Science and Technology. A cell_{ij} indicates the extent that how much an expert has specialty about each technology field. We give 1 as a default value at the cell when its technology field matches with the expert's sub-major field and added the count of technology field matching project that he/she involved in. Obviously, there are many ways to do under the given information. This is one way to value a cell of ET matrix. Varying the default value or introducing a linear and nonlinear modeling function is also considerable. The details why we choose this method will be explained in Section 4.

3.3. Network folding

Defined ET matrix is a kind of task assignment matrix identifying which set of individuals should be coordinated with their activities. Now, if we multiply the ET matrix and the transpose of the ET matrix, we obtain an expert-expert matrix. This technique, called network folding, is used to generate social network by using other people related with network as a medium:

$$\text{EE Matrix} = ET \times ET^T = E'E.$$

This product results in a people by people matrix where a $cell_{ij}$ (or $cell_{ji}$) indicates the extent to how much $expert_i$ is close with $expert_j$. In other words, the resulting matrix represents the extent to which each pair of expert coworks together. R&D Network is a weighted network constructed by using EE matrix as an adjacency matrix. Additionally, diagonal cells of EE matrix have turned into 0 to eliminate self-link.

3.4. Network grouping

Centrality is the measurement adapted from the mathematic graph theory to analyze the relation status and dynamic of the network and group. Centralization is the structure indicator for the network, group, and the individual/node position status related to other nodes (Wang & Chiu, 2008). Degree, closeness, and betweenness are widely used centralities.

Degree centrality is defined as the number of links incident upon a node:

$$C_D(V) = \frac{\deg(v)}{n-1},$$

where n is the number of node in network. Degree is often interpreted in terms of the immediate possibility of node for catching whatever is flowing through the network (such as a virus, or some information).

Closeness centrality is defined as the reciprocal of the sum of geodesic distance (i.e., the shortest path) between a node v and all other nodes reachable from it:

$$C_C(V) = \frac{1}{\sum_{t \in V \setminus v} d_G(v,t)}$$

Closeness can be regarded as a measure of how long it will take information to spread from a given node to other reachable nodes in the network.

Betweenness is a centrality measure of a node within a network. Nodes that occur on many shortest paths between other nodes have higher betweenness than those that do not:

$$C_B(V) = \sum_{\substack{s \neq v \neq t \in V \\ s \neq t}} \frac{\sigma_{st}(V)}{\sigma_{st}},$$

where σ_{st} is the number of shortest paths from s to t, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through a node v. This may be normalized by dividing through the number of pairs of vertices not including v, which is $(n-1)(n-2)$ for directed graphs and $(n-1)(n-2)/2$ for undirected graphs.

A boundary spanner is a node which, if removed from a network, creates a new component, often called a Gate Keeper. A node that has high betweenness centrality, but low degree centrality, potentially acts as a link between groups.

Network grouping, in other word, detecting community structure in network is fundamental for uncovering the links between structure and function in complex networks and for practical applications in many disciplines such as biology and sociology (Fortunato & Barthélemy, 2007). The traditional method for detecting community structure in networks is hierarchical clustering. One first calculates a weight, including some centralities explained above, W_{ij} for every pair i, j of nodes in the network, which represents in some sense how closely nodes are connected. Then one takes the n nodes in the network, with no links between them, and adds links between pairs one by one in order of their weights, starting with the pair with the strongest weight and progressing to the weakest. As links are added, the resulting graph shows a nested set of increasingly large components (connected subsets of nodes), which are taken to be the communities. Because the components are properly nested, they all can be represented by using a tree of the type, in which the lowest level at which two nodes are connected represents the strength of the link that resulted in their first candidate members of the same community. A slice through this tree at any level gives the communities that existed just before a link of the corresponding weight was added. Trees of this type are sometimes called dendrograms in the sociological literature.

Instead of trying to construct a measure that tells us which links are most central to communities, Girvan and Newman(2004) focus the links that are most "between" communities instead on those links that are most central. Node betweenness has been studied in the past as a measure of the centrality and influence of nodes in networks. Freeman defines the link betweenness of a link as the number of shortest paths between pairs of nodes that run along it. If there is more than one shortest path between a pair of nodes, each path is given equal weights such that the total weights of all of the paths are unity. If a network contains communities or groups which are only loosely connected by a few intergroup links, then all shortest paths between different communities must go along one of these few links. Thus, the links connecting communities will have high link betweenness. By removing these links, we separate groups from one another and so reveal the underlying community structure of the network.

Most of the networks that have been studied have been binary in nature; that is, the links between nodes are either present or not. Such networks can be represented by $(0,1)$ or binary matrices. However, many networks are intrinsically weighted with, their links having differing strengths. In a social network there may be stronger or weaker social ties between individuals. A weighted network is more complex than a binary one, so it has received relatively little attention. And many useful measures and tools are suitable for only binary network. To deal with a weighted network, researchers have typically introduced an arbitrary cut-off level of the weight, and then dichotomized the network by removing links with weights that are below the cut-off, and then setting the weights of the remaining links equal to one (Wasserman et al., 1994). The outcome of this procedure is a binary network consisting of links that are either present or absent (Scott, 2000; Wasserman et al., 1994). But, to ignore weights is to throw out a lot of data that, in theory at least, could help us to understand phenomena better.

Certainly some network grouping algorithm can be applied to such networks by simply ignoring link weights, but to do so is to discard useful information contained in the weights, information that could help us to make a more accurate determination of the communities. Newman (2004) generalized his algorithm applicable to a weighted network and it appeared to work excellently. In other words, the extra information contained in the link weights does indeed help us enormously to discern the community structure in the network, and the generalized Newman grouping algorithm does a good job for finding structure. For values of w greater than 2, the algorithm classifies essentially all nodes correctly.

But, how do we know when the communities found by the algorithm are good ones? A popular method now widely used relies on the optimization of a quantity called modularity, which is a quality index for a partition of a network into communities. This quantity measures the fraction of the links in the network that connect nodes of the same type (i.e., within-community links) minus the expected value of the same quantity in a network with the same community divisions but random connections between the nodes (Clauset et al., 2004). When A_{vw} is the adjacency matrix of the network, the modularity Q is defined as:

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(c_v, c_w),$$

where $m = \frac{1}{2}\sum_{vw}A_{vw}$ is the number of links in the network, $k_v = \sum_w A_{vw}$ is the degree of node v, and δ function $\delta(c_v, c_w)$ is 1 if $i = j$ and 0 otherwise. If the number of within-community links is no better than random, we will get $Q = 0$. Values approaching $Q = 1$, which is the maximum, indicate networks with strong community structure. In practice, values for such networks typically fall in the range from about 0.3 to 0.7.

We use various centrality measures to find out who has important role in R&D network and apply Newman's generalized algorithm to discover convergence technology R&D groups.

4. Empirical study

4.1. R&D network construct

One may distinguish two types of weights. Weights may represent some form of similarity or a form of dissimilarity in the co-authorship case; for example, how often two authors have collaborated over a certain time span, or weights may represent a form of dissimilarity. Particularly, this is true when weights represent distances. And the way to measure the weight for weighted networks has been introduced in three types. The first type is converting some quantities in non-weighted network into the weight of link; the weight of an link is measured by the point degrees k_i and k_j (e.g. $w_{ij} = k_i k_j$) of its two ends (Macdonald, Almaas, & Barabási, 2005). The second type, in some networks, typically natural measurement of weight is already given by the phenomena and event investigated by the network. In the scientific collaboration network, the times of co-authorship are regarded as the weight of links (Newman, 2001). The third type is in the works about modeling weighted networks; the weight w_{ij} of a link l_{ij} connecting a pair of nodes (i and j) is defined as $w_{ij} = (w_i + w_j)/2$; where w_i is defined as i node's assigned number (from 1 to N) divided by N; the weight w is assigned to the link when it is created, which is drawn from a certain distribution $\rho(w)$ (Antal & Krapivsky, 2005; Goh, Noh, Kahng, & Kim, 2005).

The R&D network is a blended case of the second type and the third type in the sense that expert's sub-major information and the number of participated project are given, but there are many ways to formulate weight using given information. We think that the weight of R&D network would have a characteristic to converge as the number of participated project increases. At certain level, small difference of absolute value in weight will be meaningless to distinguish one expert's ability from another's. Intuitively, the more times, the closer is the relationship, and the less contribution that one new event can provide to the relationship. That means the contribution of a new event to the relationship should decrease on marginal.

We try three methods to weight ET matrix, comparing which made good result. The first method, mentioned in Section 3, is setting default value (e.g., 1) according to expert's sub-major field and adding the count of involved project. This is a normal approach typically used in task assignment type matrix. The second method is a kind of discretization approach using a scree plot that discretizes a range of numeric attributes in the dataset into nominal attributes. We divide the range of weight in ET matrix into 4 levels by the result of scree analysis: (1) level 1: from 1 to 2, (2) level 2: from 3 to 6, (3) level 3: from 7 to 10, (4) level 4: more than 11. Then, we use the level value as the weight of ET matrix. The third method is using a tanh function to convert the counts into weight. Tanh function starts from $\tan h(0) = 0$; and increases up to 1 when the independent variable is large enough (Lia et al., 2005). But, tanh function tends to increase too fast, so that the weight divided by its standard deviation is used as an input value to compensate it.

And the modularity, a quality index for a partition of a network into communities, is used as a measure of comparison. Fundamentally, it is very difficult to judge which network is better, then, it would be rather to make a decision from an applicative viewpoint. In our case, the network which shows a good result of grouping is

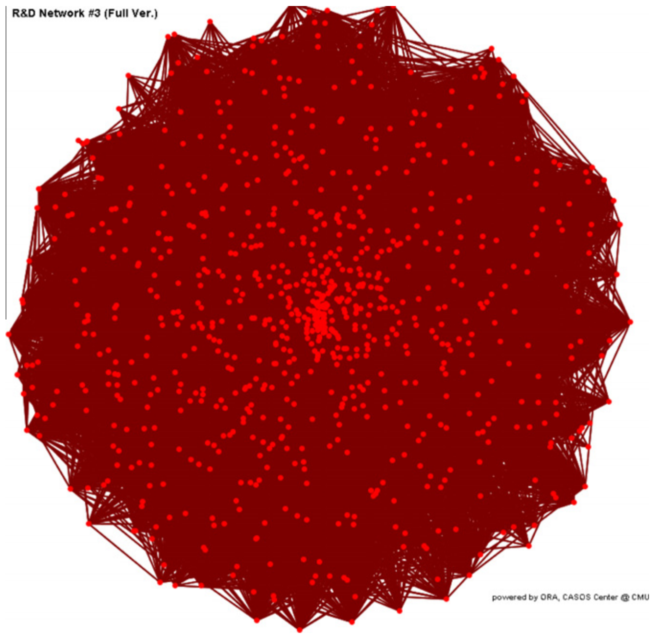

Fig. 1. Initial network.

essential to study a current state of R&D network. Therefore, the method shows higher modularity is preferred. The modularity of Newman grouping for the network made by the first, the second, and the third method is 0.5384, 0.5054, and 0.4593, respectively. We think that though the first method does not have the convergence characteristic, but now the absolute value of weight is not too high to affect negative effects, so it reveals the structure of R&D network best. However, the introduction of transform function is required later, when more data is available and the size of network is very huge. Fig. 1 shows the initial network constructed by the first method.

The constructed network has 817 nodes and 40,714 links. Considering various network measures, average distance (2.717), connectedness (0.9951), density (0.0611), efficiency (0.9409), this is a well-connected and sparse network that most nodes are connected by a few links.

The degree histogram also shows that R&D network is a typical scale free network that contains many nodes with a few links and a few nodes with many links (see Fig. 2). As Derek de Solla Price (1965) described, the network of citations between scientific papers typically has power-law distribution, so it is a scale free network. The characteristic of scale free network, "the rich get richer", fits well for that it is probable a new expert has relationship with an existing expert who has a good research performance rather than a poor one. And there are just three components in the network, whereas two of them are isolated. This means almost all nodes in the network are connected loosely, so that it is a difficult to group the network.

To find out who has important role in the network, we investigate upper 1% expert on various centralities. About degree centrality, important experts are Kim, Young Ran, Won, Mi-Sun (0.1887), Kim, Young Ho, Lee, Hyeong-Kyu, Cho, Jung-Hyuk (0.185), Choi, Soo-Young (0.1814), Kang, Sung Man, Kim, Ji Young, Kim, Hyung-Kee, Kim, Hee-Sun, Park, Hyun-Sung, Lee, Jin Hwa, Rhim, Hyang-Shuk, and Cho, Eun-Jung (0.1801). About betweenness, centrality important experts are Lee, Sung-Hoon (0.0445), Lee, Jang-Hee (0.0333), Mun, Han-Seo (0.0309), Park, Kwang Kyun (0.0249), Kim, Hyun Soo (0.0211), Lee, Dong-Choon, and Park, Jung-Hee (0.0208). About closeness centrality, important experts are Lee, Sung-Hoon (0.2411), Mun, Han-Seo (0.2409), Choi,

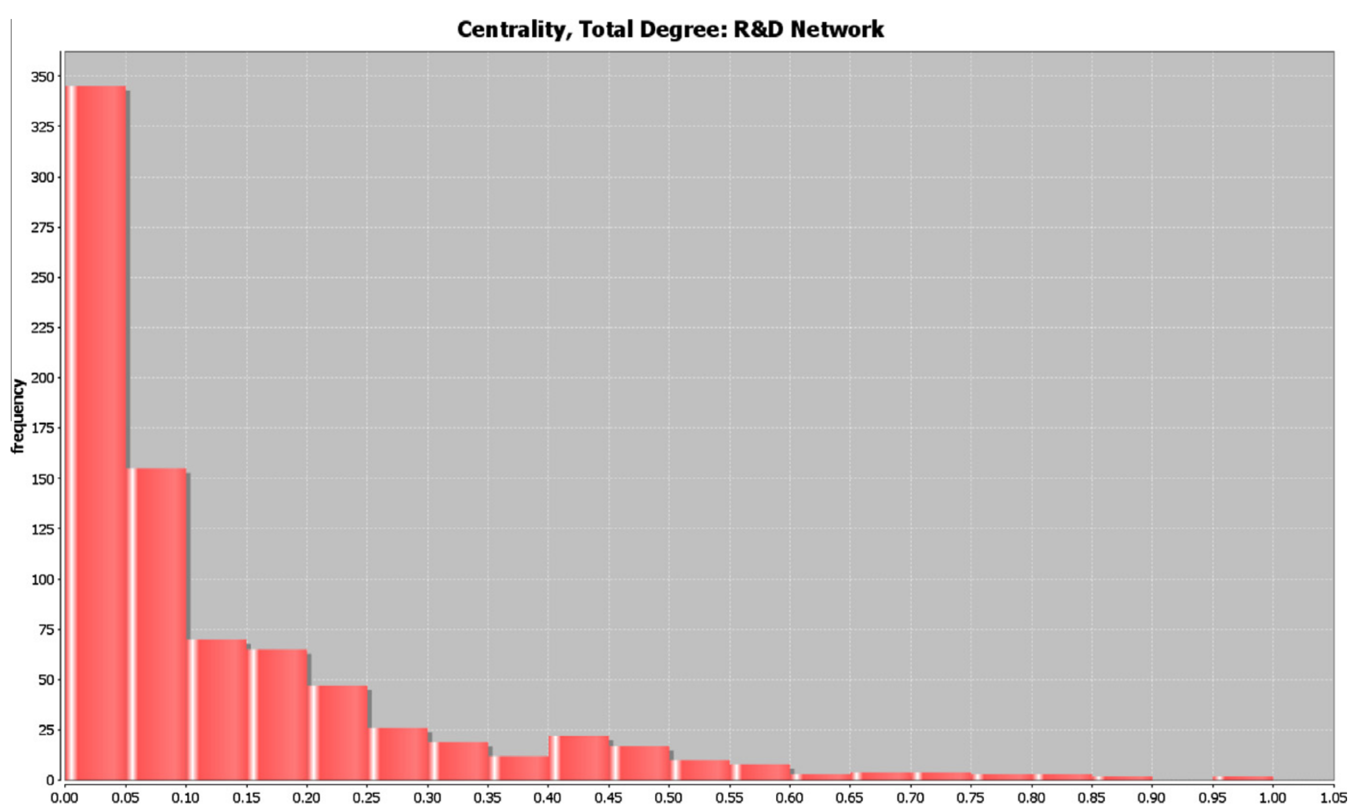

Fig. 2. Degree histogram.

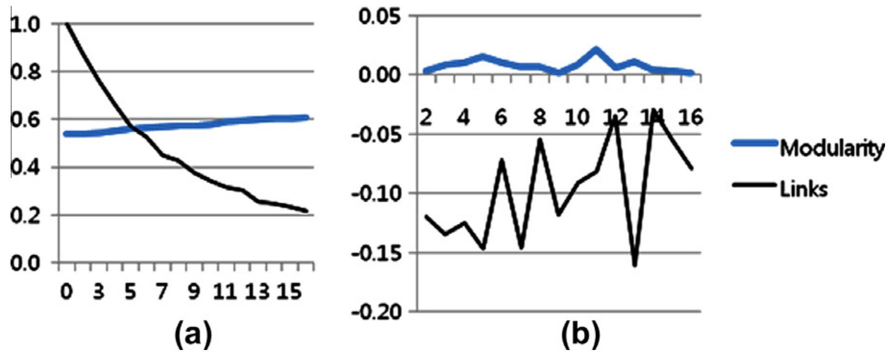

Fig. 3. Network cut-off by the comparison of modularity and number of links.

Soo-Young (0.2399), Jung, Hee Jung, Kim, Young Ran (0.2386), Lee, Yu Mi (0.2382), Kim, Young Ho, Lee, Hyeong-Kyu, and Cho, Jung-Hyuk (0.2376). About eigenvector centrality, important experts are Kang, Sung Man, Kim, Ji Young, Kim, Hyung-Kee, Kim, Hee-Sun, Park, Hyun-Sung, Lee, Jin Hwa, Rhim, Hyang-Shuk, and Cho, Eun-Jung (1.0). A boundary spanner is Choi, Soo Mi, and potentials are Lee, Jang-Hee (0.0183), Kim, Dong Ik (0.0132), Park, Jung-Hee (0.0127), Lee, Dong-Choon (0.0115), Han, Pyeong Leem (0.0112), Choi, Soo Mi (0.0111), Han, Kyung-Sook (0.0106), and Choi, Yong Doo (0.0102). In result, experts who have many links and a position in the center of the network are Kim, Young Ran, Kim, Young Ho, Cho, Jung-Hyuk, and Choi, Soo-Young. And experts who connect the network and their absence will collapse the network are Lee, Sung-Hoon, Lee, Jang-Hee, Mun, Han-Seo, Lee, Dong-Choon, and Park, Jung-Hee. But in our study, they are more meaningful who are important in each subgroup rather than who are important in whole network.

4.2. Grouping

As explained in Section 3, we tried to apply Newman grouping to constructed R&D network. Though the modularity is reasonable value (0.5384), but the number of group found is only four. The reason of poor result is that the network is too sparse and links are too abundant to perform a successful grouping. To solve this problem, a network cut-off is adopted. The deletion of abundant links helps to find out useful relationship between experts. Besides, it is true that cut-off means the loss of information. Therefore, to

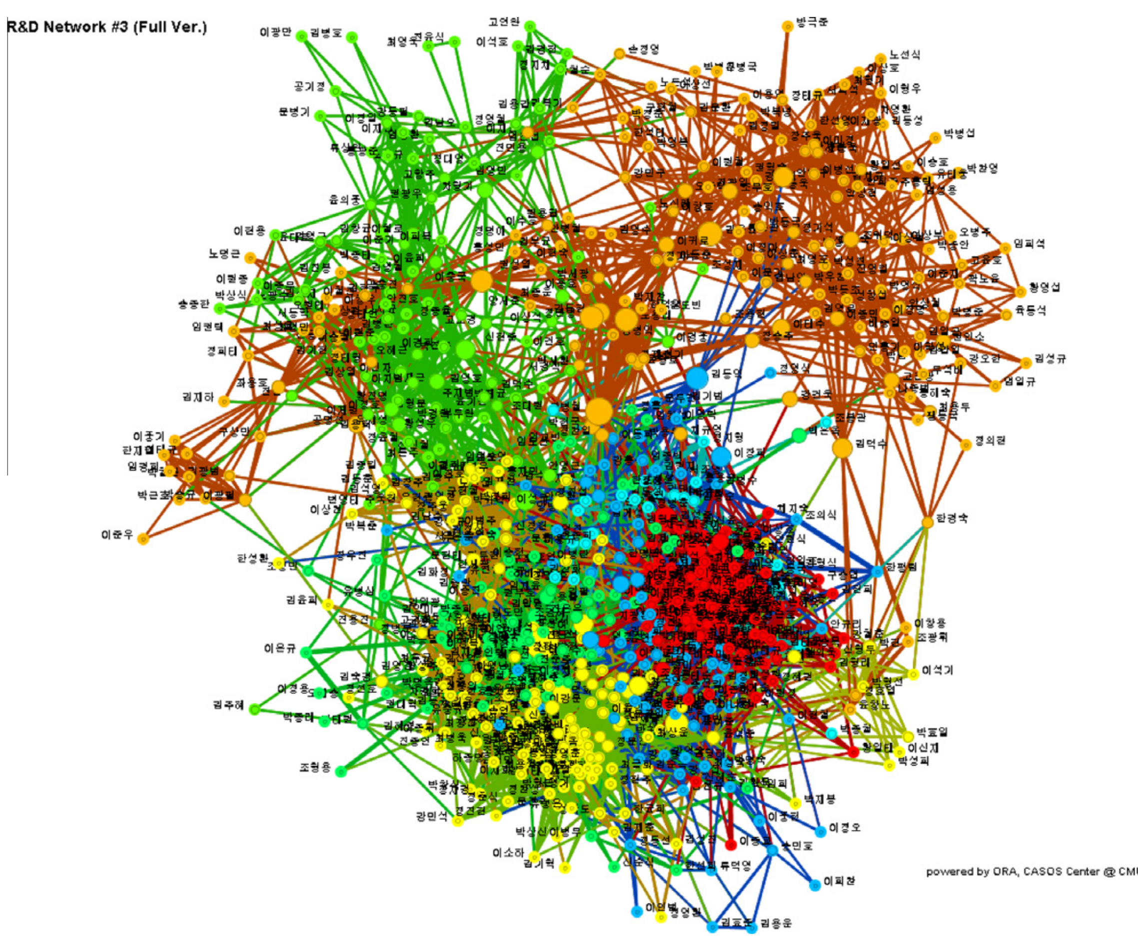

Fig. 4. Treated network and grouping result.

Table 1
Basic information of 9 groups.

Group	Member	Leading affiliation of members	Leading sub-major of members
Group 1	191	ETRI (5.24%), KAIST (4.71%), KIST (4.19%), Hanyang University (4.19%), Kyungpook National University (3.14%), Yonsei University (3.14%)	Physiological instrumentation (3.68%), computer networking (3.16%), mobile communication system (2.98%), embeded software (2.81%), sensor network (2.63%), electromagnetic nano-materials (2.63%), application software (2.46%), integrated circuit design (2.28%), telecommunications protocol (2.11%), molecular and nano-composites preparation (1.93%), mobile communication network (1.93%), Internet information consumer electronics terminal (1.93%), control and automation system (1.93%)
Group 2	106	Korea University (7.55%), Hanyang University (7.55%), KERI (6.60%), KIST (5.66%)	Semiconductor processing (4.73%), semiconductor materials (4.42%), discrete semiconductor (4.10%), semiconductor (3.79%), compound semiconductor (3.79%), semiconductor materials (3.47%)
Group 3	58	Ewha Womans University (8.62%), KIST (6.90%)	Biomaterial and device (12.14%), formulation development (11.56%), biological ceramics (4.62%), gene therapy technology (4.62%)
Group 4	132	Seoul National University (9.85%), Catholic University (4.55%), Kyungpook National University (4.55%), Ulsan University (4.55%)	Biochemistry and molecular biology (14.58%), gene expression and regulation (14.58%), molecular and cellular biology (10.23%), functional genomics (3.32%), structural genomics (2.81%)
Group 5	97	KIST (7.22%), KRICT (6.19%), Kunghee University (4.12%)	New drug discovery (23.69%), natural products chemistry (9.06%), medicinal and combinatorial chemistry (8.01%), pharmaceutical materials manufacturing process (4.88%), protein and enzyme molecule biochemistry (2.09%), plant resources (2.09%), synthetic methodology (2.09%)
Group 6	90	Seoul National University (7.78%), Chonnam National University (6.67%), Chonbuk National University (6.67%)	Microbiology, parasitology and immunology (8.68%), recombinant DNA technology (5.66%), artificial cell and organ development (4.15%), stem cell engineering (4.15%), immunology (3.40%)
Group 7	24	Pukyong National University (8.33%), Seoul National University (8.33%), Seoul National University Hospital (8.33%), Inje University (8.33%)	Diseases of psychiatry and neuroscience (15.28%), aging and geriatric medicine (6.94%), physiology (6.94%), lipid molecule biochemistry (6.94%), diseases of respiration and circulation (6.94%)
Group 8	85	Catholic University (4.71%), Sungkyunkwan University (4.71%), Chosun University (4.71%)	Functional food and food bioactive components (20.47%), natural product chemistry (3.94%), cosmetic ingredient discovery (3.54%), clinical medical science (2.76%), anatomy, pathology and legal medicine (2.76%)
Group 9	10	Seoul National University Hospital (20.00%)	Biological databases and managements (26.67%), game and animation (6.67%), biomedical informatics (6.67%)

determine a cut-off level, we plot the change of modularity upon the change of link cut-off level in two ways, the absolute .value and the increase rate. The number of link is normalized from 0 to 1.

Considering the crossing point and increase rate, to cut-off link weight is under 5 is the best (see Fig. 3). The result applying Newman grouping to cut-off network is that 12 groups are found and the modularity is 0.5589. Three groups among 12 are meaningless, because they are a group of isolated nodes and pendants. Fig. 4 shows a cut-off network and grouping result.

5. Findings

5.1. Sub-group characteristics

By the result of grouping, there are 9 meaningful groups in our R&D network. Table 1 shows basic information about each group. The leading features are upper 5% of each group.

Based on a set of each group members' attributes, we assume the main research theme of each group. We can find out what technologies are converged by looking the distribution of members' sub-major fields and the graph of the network divided by each group. Fig. 5 shows 9 graphs of the 9 groups that cut off links weights under 12 to see the network structure of each group easily. There are one big cluster and the other surroundings in some graph like group 4. Besides, there are two or more clusters in other graph like group 3. We guess the latter case means the convergence of technology.

And important experts of each group are figured out by centrality measures, mostly degree and betweenness. The first basic idea is an expert with high degree will be a core member of the group that has many and strong relationships with other experts within the group. The second is an expert with high betweenness will be a core member of the group that plays a role to connect many other experts within the group. The third is an expert with high degree and high betweenness will be a very important person, a core member in the group. The fourth is an expert with high potential boundary spanner will be a gate keeper of each group to connect the group he/she belonged with other group. We select important experts who belong within upper 5% in each measure from each group.

Additionally, we can think that in the case of the group convergence occurred, high degree members and high betweenness members are important to understand the situation, what technologies are converging by whom. And in the case of the group that no convergence appeared, boundary spanners are meaningful to seek the opportunity to connect with other groups:

- Group 1: The main research theme of this group seems to be ubiquitous mobile devices and smartwears with a physiological instrumentation function, namely IT, BT, and NT convergence technology. This group is very complex than the others, means a remarkable dominant technology does not exist and there are many technologies with similar priority. But, it may imply this group is a true convergence technology group that many technologies are mixed. Members with high degree are Yang, Choong Jin, Yoon, Hyung Ro, Lee, Mee Jeong, Lee, Cheol Jin, and Jung, Chun Ki. Members with high betweenness are Gong, Myoung Seon, Kim, Deok Su, Kim, Yong Jin, An, Beong-Ku, and Lee, Tae Soo. Members with high degree and betweenness are Kim, Dong-Hyun, Roh, Yong-Rae, Yi, Choong-Kook, Chang, Byung Chul, and Chung, Wan-young. Members with high boundary spanner are Koh, Jin-Gwang, Gwon, In So, Son, Jeong Young, Lee, Gun Ho, Lee, Byung Sun, Lee, Jae Jin, and Jung, Ho Yeol.

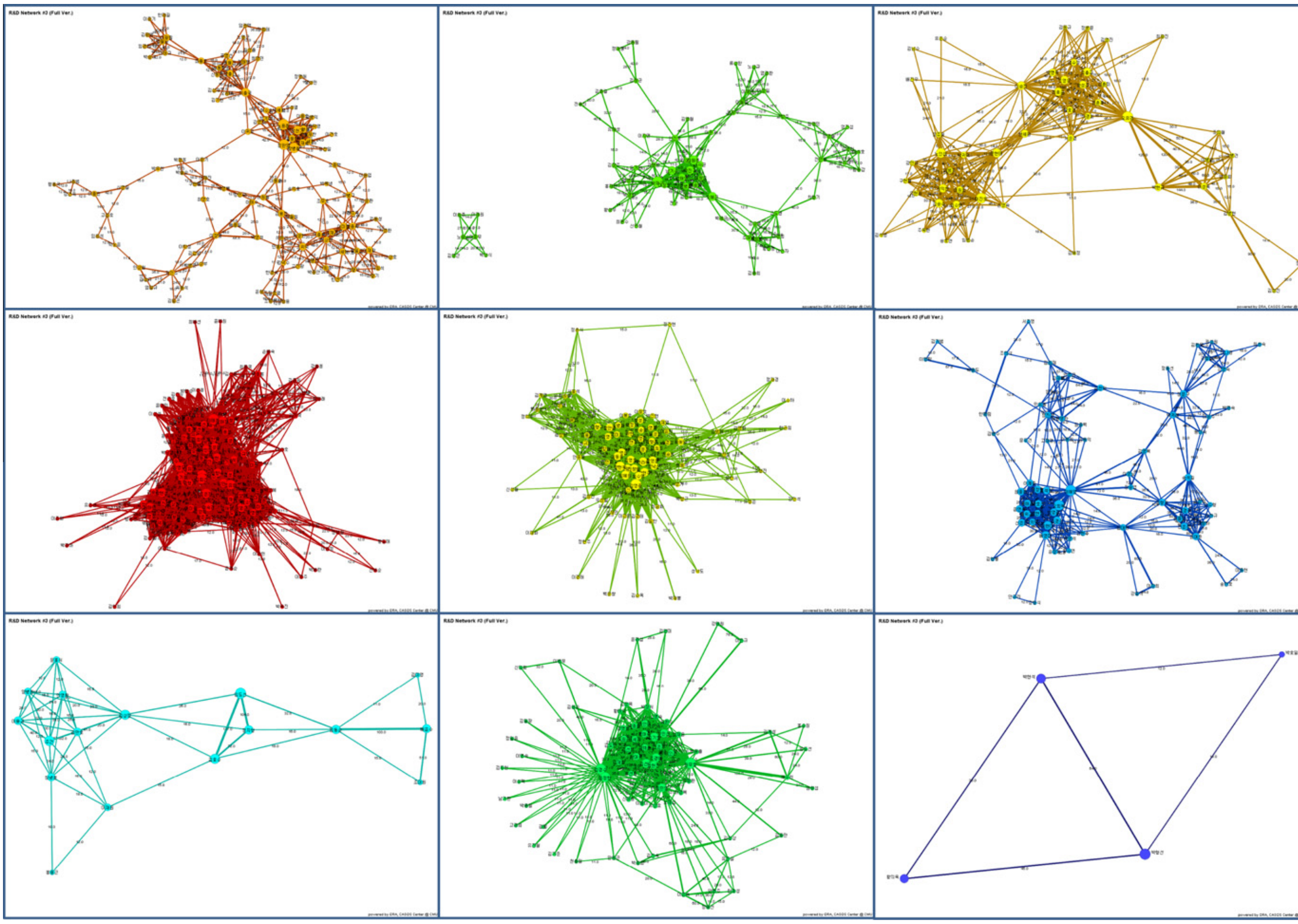

Fig. 5. Network graph of each group (brief version).

As a result of network analysis, Roh, Yong-Rae, Chang, Byung Chul, and Chung, Wan-young are the representatives of cluster related with sensor technology; Yi, Choong-Kook is the representative of cluster related with material technology including some nanotechnologies; Lee, Mee Jeong is the representative of cluster related with network and telecommunication technology.

- Group 2: The main research theme of this group seems to be multi-purpose and organic semiconductor, namely IT and BT convergence technology. This group does not have a dominant technology, but most sub-major technologies are related with a semiconductor. Members with high degree are Nam, Ki Su, Seong, Tae-Yeon, Oh, Hye Gun, and Lee, Won-Jun. Members with high betweenness are Park, Jung-Hee, Yu, Jae-Sung, Lee, Jeong Ho, and Cha, Hyung Kee. Members with high degree and betweenness are Park, Jae Keun and Park, Jin Koo. Members with high boundary spanner are Kim, Jong Gil, Kim, Ju-Hye, Lee, Kyoung Joung, and Jeon, Keun.
  As a result of network analysis, Park, Jae Keun and Park, Jin Koo are the representatives of both cluster related with semiconductor materials and processing technology. This means semiconductor materials and processing are already combined inseparably.
- Group 3: The main research theme of this group seems to be biocompatible material, namely BT and NT convergence technology. This group is dominated by two technologies, Biomaterial and device and Formulation development. Members with high degree are Oh, Yu-Kyoung and Lee, Jin-Ho. Members with high betweenness are Moon, Hyun Tae and Lee, Seong Wook. A member with high degree and betweenness is Roh, In Sup. Members with high boundary spanner are Kim, Jae Hoon, Lee, Sang-Cheon, Joo, Wook Hyun, and Hwang, In-Taek.
  As a result of network analysis, Roh, In Sup is the representative of cluster related with biomaterial and device technology and biological ceramics technology; Oh, Yu-Kyoung is the representative of cluster related with formulation development and gene therapy technology.
- Group 4: The main research theme of this group seems to be therapeutics for an incurable disease, namely BT. Almost technologies are related with genetic engineering and drug delivery. Members with high degree are Kang, Sung Man, Kim, Hee-Sun, Yu, Young Do, Lee, Tae Hoon, Cho, You-Hee and Cho, Eun-Jung. Members with high betweenness are Kim, Dong Seon, Kim, Seong-Jun, Park, Eui Kyun, Song, Ki Won, Won, Mi-Sun, and Chung, Bong-Chul. A member with high degree and betweenness is Kwon, Young-Geun. Members with high boundary spanner are Kim, Min-Sun, Kim, Young Ho, Yoon, Gae Soon, Yoon, Hye Won, and Choi, Hei-Sun.
  As a result of network analysis, Kwon, Young-Geun is the representative of this group. And Kim, Young Ho has a strong connection with group 6. This means there are a possibility and a demand to cowork group 4 and 6 together.
- Group 5: The main research theme of this group seems to be new drug development, namely BT. This group is highly dominated by new drug development technology Members with high degree are Kim, Young Choong, Yang, Hyun-Ok, and Lim,

Hye Won. Members with high betweenness are Mun, Han-Seo, Park, Kwang Kyun, and Shin, Jae Min. Members with high degree and betweenness are Kim, Young Ho and Choi, Soo-Young. Members with high boundary spanner are Kim, Suk Kyung and Jeong, Moon-Jin.

As a result of network analysis, Kim, Young Ho and Choi, Soo-Young are the representatives of this group. And Jeong, Moon-Jin has a strong connection with group 3 and 6. This means there are a possibility and a demand for group 5 to cowork with group 3 and 6.

- Group 6: The main research theme of this group seems to be stem cell and artificial tissue and organ, namely BT and NT convergence technology. This group does not have a remarkable dominant technology, but Microbiology, parasitology and immunology has the highest priority. Members with high degree are Kim, Kyung-Sik, Oh, Hee-bok, Lee, Hyun-Ah and Pyo, Suhk-Neung. Members with high betweenness are Kim, Dong Ik, Lee, Jang Hee, Lee, Chung Yul, and Jun, Eon Suk. A member with high degree and betweenness is Lee, Aeh Young. Members with high boundary spanner are Song, Min-Ho and Lee, Dong-Hee.

  As a result of network analysis, Lee, Aeh Young is the representative of cluster related with Microbiology and Artificial cell and organ development technologies; Kim, Kyung-Sik is the representative of cluster related with Stem cell engineering technology.
- Group 7: The main research theme of this group seems to be diseases of psychiatry and geriatric medicine, namely BT. This is a small group dominated by diseases of psychiatry and neuroscience. A member with high degree is Lim, Do-Seon. A member with high betweenness is Kim, Seung Mok. A Member with high degree and betweenness is Kim, Hyo Soo. A member with high boundary spanner is none.As a result of network analysis, Kim, Hyo Soo is the representative of this group.
- Group 8: The main research theme of this group seems to be discovery and application of natural material, namely BT and NT convergence technology. This group is dominated by functional food and food bioactive components. Members with high degree are Kim, Kyung-Su, Kim, Sun Yeou, and Yoon, Jung Han. Members with high betweenness are Kim, Do-Man, Park, Chang Seo, and Baek, Eun Ok. Members with high degree and betweenness are Won, Moo Ho and Cho, Kyung Hea. A member with high boundary spanner is Park, Jong-Cheol.

  As a result of network analysis, Won, Moo Ho and Cho, Kyung Hea are the representatives of cluster related with Functional food and food bioactive components technology; Kim, Kyung-Su is the representative of cluster related with Natural product chemistry technology.
- Group 9: The main research theme of this group seems to be bioinformatics, namely IT and BT convergence technology. This group is very small and totally focuses on life and medical information research. A member with high degree is Park, Hyung Sun. A member with high betweenness is Park, Hyun Suk. A Member with high degree and betweenness is Hwang, Eui Wook. A member with high boundary spanner is Park, Hyo Il. As a result of network analysis, Hwang, Eui Wook is the representative of this group.

Fig. 6 shows the research trend of IT, BT, and NT convergence technology. By comparing our result with this figure, we will have an overlook of current situation of convergence technology R&D in Korea. There are 1 NBIT technology, 2 BIT technology, 3 NBT technologies and 3 BT technologies within 9 groups we found. R&D projects ordered by Korean government cover many portion of convergence technology trend; however, in overall view the investment to NT seems to be insufficient. In the case of IT and BT convergence, it is short of

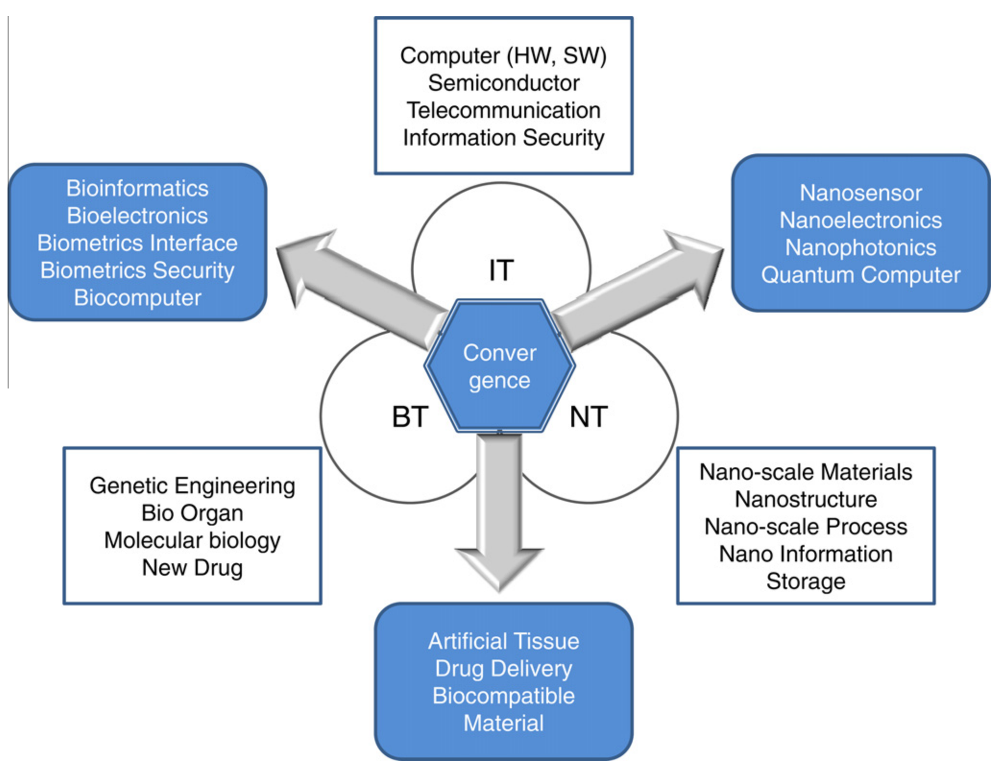

Fig. 6. The trend of convergence technology.

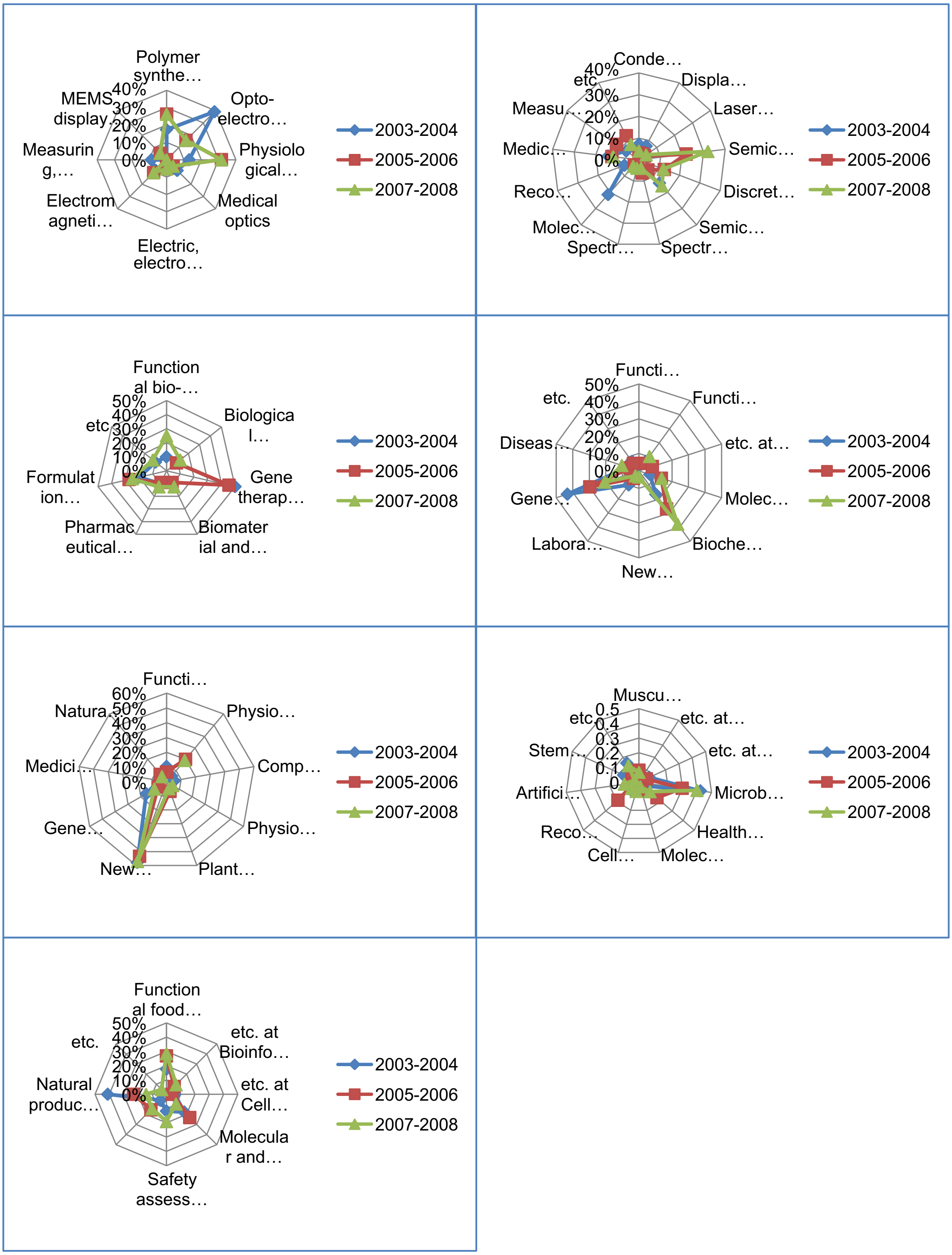

Fig. 7. Biennial research trend of each group from 2003 to 2008 (group 1–6, 8).

that biocomputer and biometrics security related R&D. In the case of IT and NT convergence, it is necessary to invest nanoelectronics, nanophotonics, and quantum computer related R&D. Considering the distribution of the convergence type and the numerical scale of experts for each group, there seems to be unbalanced in Korean convergence technology R&D; BT overwhelms other two technologies, especially NT. It can be a kind of "select and concentrate" strategy of Korean government or it is due to the fact that BT has a deep relation with medical industry.

5.2. Research trend of each group

For the investigation of research trend of each group, we search the papers of important experts. Because the project data are well computerized in databases since 2003, we gather the paper information of important experts of each group from 2003 to 2008. We determine the main topic for each paper and get the tables where the numbers of papers per each topic are summarized yearly. By comparing the topic yearly, we can identify the research trend; whether the research about the topic remains, or increase or shrinks. If a student wants to study about a specific convergence technology, this trend result can help him/her to find whom to ask and what technologies to study. Fig. 7 shows the change of proportion of the topics every two years. Group 7 and 9 are omitted in the chart because they have a few number of samples. For the normalization, we use the value of proportion for the radial diagrams. We can discover that the research about opto-electronics has been reduced in group 1. And we can find out that the research about gene therapy technology has been reduced and the formulation development has been increased alternatively in group 3.

6. Discussion

This paper explores the convergence of technologies, focusing IT, BT, and NT, in Korea by using a R&D network. The R&D network is a novel approach to answer the old question how scientists or engineers interact with each other, at higher level such as a country. It can solve some limits of traditional collaboration network and citation network that researchers generally use one database to get data and they only use data itself to construct a network. There are practical difficulties to construct a social network using the information gathered from two or more heterogeneous databases. Nevertheless, we can assume that if we use rich data and information collected from many databases, we will be able to construct more useful network to show the structure and the effects underlying a specific phenomenon. And it can be done by using meta-data approach and network folding technique. The R&D network can inform our understanding of the interactions between experts who participate in national R&D projects.

For convergence technology, convergence of IT, BT, and NT is a strong and worldwide trend. We have shown that the human resource information and the project information collected from Korean nation R&D project databases can be merged and translated into a weighted network. And 9 convergence technology groups are found through the grouping by the Newman's grouping algorithm generalized to a weighted network. In this process, we can confirm the fact that the extra information contained in the link weights does indeed help us enormously to discern the community structure in the network. The network structure depends on the weights. Even though some global distribution seems to be robust, but the detailed structure has been affected by the weights.

Discovered 9 groups are named and classified to what kind of convergence technology (e.g. BIT) based upon the distribution of attribute, in this case members' sub-major technology fields, and the network structure of each group: ubiquitous mobile devices and smartwears with a physiological instrumentation function (NBIT), multi-purpose and organic semiconductor (BIT), biocompatible material (NBT), therapeutics for an incurable disease (BT), new drug development (BT), stem cell and artificial tissue and organ (NBT), diseases of psychiatry and geriatric medicine (BT), discovery and application of natural material (NBT), and bioinformatics (BIT), respectively. Generally, Korean national R&D projects follow worldwide convergence technology trend well. However, in overall view, it turns out that the investment to NT is relatively low. This can be a weak point of Korean economy or "select and concentrate" strategy of Korean government, but the knowledge about the current situation will help policy decision makers anyhow.

There are some groups that have a flat and massed structure and the other groups that have a dynamic and segmented structure within 9 groups we found. Roughly speaking, the former cases are lack of convergence and the convergence of technologies are occurring in the latter cases. Using network analysis, important people who are representatives of each group can be figured out. The most complex group among we found, for example, has three large clusters which denote sensor, material, and network and telecommunication and Roh, Yong-Rae with other two experts, Yi, Choong-Kook, and Lee, Mee Jeong are the representatives of each cluster, respectively. When a policy decision maker want to hear an advice or to start a new project, this information is very helpful.

While the empirical study is limited to cover small part of Korean national R&D projects data, its lessons can be applied to other network which is a possibility to be useful when data is collected from multi-sources. As we have seen for convergence technology, the data in many databases that has similar meaning but not exactly consistent with each other can be integrated by using meta-data mapping. If there are some people and a medium to connect them, then we can construct a network of them by introducing a schema system to transform similarity into consistency. Then, we can find out the useful result to analyze the network.

This paper is just a cornerstone of the work to investigate the current situation of national R&D projects in Korea. Therefore, further research is strongly required. First, the extension of data coverage and systematization are required. There are data of over 70,000 experts and more R&D projects in national databases. We focus the convergence technology and use a part of them. However if we enlarge the coverage, we can find out more information about the current situation of Korean national R&D projects. In that case, it has to be allowed to access to database directly or make R&D network module inside NTIS system. Second, an actual proof of the results by qualitative research is required. The result from R&D network needs to be examined whether or how it fits with the real world. Third, refining of weighting formulation is required. As we stated, if the size of network increases it will be need to introduce a modeling function. For the rest, if another database is added, then the way to formulate weight should be modified with the change.

References

Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science, 45*(7), 905–917.

Antal, T., & Krapivsky, P. L. (2005). Weight-driven growing networks. *Physical Review E, 71*(2), 026103/1–026103/6. Article No. 026103.

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*, 066111/1–066111/6.

Davis, G. B., & Carley, K. M. (2008). Clearing the FOG: Fuzzy, overlapping groups for social networks. *Social Networks, 30*.

Derek de Solla Price, J. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.

Dosi, G., & Kogut, B. (1993). National specificities and the context of change: The coevolution of organization and technology. In B. Kogut (Ed.), *Country competitiveness: Technology and the organizing of work*. New York: Oxford University Press.

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences of the United States of America, 104*(1), 36–41.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*, 35–41.

Freeman, L. C. (1979). Centrality in social network: I. Conceptual clarification. *Social Networks, 1*.

Freeman, L. C. (1980). The gatekeeper, pair dependency and structural centrality. *Quality and Quantity, 14*.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821–7826.

Goh, K.-I., Noh, J. D., Kahng, B., & Kim, D. (2005). Optimal transport in weighted complex networks. *Physical Review E, 72*(1), 1–4. Article No. 017102.

Guare, J. (1990). *Six degrees of separation: A play*. New York: Vintage.
Hacklin, F., Marxt, C., & Fahrni, F. (2009). Coevolutionary cycles of convergence: An extrapolation from the ICT industry. *Technological Forecasting and Social Change, 76*(6), 723–736.
Hagedoorn, J., & Schakenraad, J. (1992). Leading companies and networks of strategic alliances in information technologies. *Research Policy, 21*(2), 163–190.
Hanneman, R. (1998). *Introduction to social network*. <http://wizard.ucr.edu/~rhannema/networks/nettext.pdf>.
Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z.-K., et al. (2003). Longitudinal patent analysis for nanoscale science and engineering: Country, institution and technology field. *Journal of Nanoparticle Research, 5*(3–4), 333–363.
Li, M, Fan, Y., Chen, J., Gao, L., Di, Z., & Wu, J. (2005). Weighted networks of scientific communication: The measurement and topological role of weight. *Physica A: Statistical Mechanics and its Applications, 350*, 643–656.
Macdonald, P. J., Almaas, E., & Barabási, A.-L. (2005). Minimum spanning trees on weighted scale-free networks. *Europhysics Letters, 72*(2), 308–314.
McKelvey, M. (1997). Co-evolution in commercialization of genetic engineering. *Industrial and Corporate Change, 6*(3), 503–532.
Moed, H. F., Burger, W. J. M., Frankfort, J. G., & Van Raan, A. F. J. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy, 14*(3), 131–149.
Narin, F., Hamilton, K. S., & Olivastro, D. (1997). The increasing linkage between US Technology and Public Science. *Research Policy, 26*(3), 317–330.
Nederhof, A. J., & van Raan, A. F. J. (1993). A bibliometric analysis of six economics research groups: A comparison with peer review. *Research Policy, 22*(4), 353–368.
Newman, M. E. J. (2001a). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 404–409.
Newman, M. E. J. (2001b). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64*, 016131/1–016131/8.
Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E, 70*.
Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2).
Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(1), 2566–2572.
Nieminen, V. (1974). On centrality in graph. *Scandinavian Journal of Psychology, 15*, 16.
Preece, J. (2002). *Online communities designing usability, supporting sociability*. Chichester, UK: John Wiley.
Schwartz, M., & Wood, D. (1993). Discovering shared interests using graph analysis. *Communications of the ACM, 36*(8), 78–89.
Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). London: Sage Publications.
Scott, J. (2002). *Social networks: Critical concepts in sociology*. Routledge Publication Inc.
Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management Science, 51*(5), 756–770.
Tijssen, R. J. W. (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy research. *Research Policy, 21*(1), 27–44.
von Hippel, E. (1987). Cooperation between rivals: Informal know-how trading. *Research Policy, 16*(6), 291–302.
Wang, J.-C., & Chiu, C.-C. (2008). Recommending trusted online auction sellers using social network analysis. *Expert Systems with Applications, 34*(2008), 1666–1679.
Wasserman, S., Faust, K., Iacobucci, D., & Granovetter, M. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'smallworld' networks. *Nature, 393*, 440–442.
Wellman, B. (1996). For a social network analysis of computer networks: A sociological perspective on collaborative work and virtual community. In *Proceedings of the ACM SIGCPR/SIGMIS conference, Denver, USA* (pp. 11–13).
Wellman, B., & Gulia, M. (1997). Net surfers don't ride alone – Virtual communities as communities. In P. Kollock & M. Smith (Eds.), *Communities and cyberspace* (pp. 331–366). New York: Routledge.