Title: Academic Research Activities and their Co-author and Keyword Network in Epidemiology Fields : Analysis of Papers in the Korean Journal of Epidemiology, 1991~2006

Author: Minsoo Jung, PhD Candidate, MPH

Position: Graduate School of Public Health, Seoul National University

Corresponding Author : Minsoo Jung

Address : Department of Health Policy and Management 414, School of Public Health, Seoul National University. 28 Yeongeon-dong, Jongno-gu. Seoul 110-799, South KOREA

Tel: +82-2-16-674-6449 Fax: +82-2-745-9104 E-mail: mins.jung@gmail.com

Running head: Co-author and keyword network

Abstract

Objectives: This research analyzed knowledge structure and its effect factor by evaluation of coauthor and keyword network in Korea's Epidemiology sector.

Methods: The data was extracted from 318 papers listed in *the Korean Journal of Epidemiology*, and was transformed into 643 coauthors and 131 keywords matrix. In this matrix a link was judged by impact factors which were calculated by the weight value of what the role was and the rate of how many authors participated. We verified that the research achievement was dependent upon the author's status and network index.

Results: The results showed a small world effect according to the development of a random network in the center of a few high productivity researchers. In particular, degree centrality was more developed than closeness centrality. Also, power law distribution was discovered in impact factor and research productivity by college affiliation. In multiple regression, the effect of the author status was significant in both the impact factor calculated by the participatory rate and the number of listed articles. Moreover, a small group of researchers with outstanding research productivity carried out many of the core academic activities in the Journal of Preventive Medicine and Public Health.

Conclusions: This study shows that the small world phenomenon exists in coauthor and keyword networks in the unit of journal like as citation networks. However, the coauthor networks in the field of epidemiology was more differentiated than preventive medicine field.

Key words: Epidemiology, Preventive Medicine, Sociology, Interdisciplinary Communication, Korea

INTRODUCTION

Science and Technology Studies are a discipline that studies how social factors intervene in the production and exchange of scientific knowledge [1-3]. It critically interprets the process of creation of scientific knowledge and analyzes the relationship between science and society or studies the social structure of science. This paper uses one of the methods of analyzing the social structure of science, the article coauthor network and keyword network, to examine the scientist community in the field of epidemiology.

Currently the top academic research is published in SCI/E and SSCI selected by ISI (Institute of Science Information). This is done by JCR (Journal Citation Report) conducting IF (Impact Factor) computation of the influential effect index of each journal and article every year. To take 2007 for example, IF = (the number of papers published in relevant journals in 2005 and 2006 which were cited in 2007) / (the number of papers published in relevant journals in 2007). However, scales such as cited index or cited half-life cannot consider the qualitative side on citations. Instead, by providing inducements on cites, it can at times create distorted behavior. This paper obtained IF of contributors in a single journal based on STS theories and network analysis. By studying the papers submitted to a journal during a set period of time, this method adds to the understanding of the coauthor network and research subjects of scientists who are active in that specific field. Not only can a knowledge structure map of the relevant journal then be provided but also the distorted inducement on citations can be eliminated and research behavior reorganized.

A coauthor network is the relationship of scientists who have authored a paper together. And, a keyword network is a network that reveals structured knowledge through keywords presented in one paper. When these research behaviors of scientists are analyzed through network analysis, the way knowledge is being created and shared in the scientist community can be dynamically discovered. Accordingly, studies on coauthor, keyword, and citation network have been utilized as tools to measure knowledge structuralization in Korean journals [4-8], and developed especially in the field of bibliometrics [9].

From the viewpoint of network theory, coauthor network shows the characteristics of power law distribution [10,11]. This refers to the phenomenon in which scientists freely produce knowledge through joint research but only a few researchers and keywords become prominent in journals and many others do not [12]. This stratification phenomenon of science is, as Merton points out, due to individual characteristics and simultaneously to relational characteristics emerging according to the social structure of the relevant discipline [13]. It is conjectured that the scientist community within KSE (the Korean Society of Epidemiology) have been forming a specific structure much like an epidemiological transition of infectious disease [14]. This paper is a follow-up research of the one that investigated JPMPH¡s (the Journal of Preventive Medicine and Public Health) coauthor network [15], and it examined papers, coauthors, and keywords published in KJE (the Korean Journal of Epidemiology) for the past 16 years, analyzing elements that affect major researcher clusters and research productivity. Through such evaluation work, major characteristics and subjects in the field of epidemiology in Korea were discussed.

OBJECTS AND METHODS

I. Objects

The research object is KJE from 1991 (Vol. 13, No. 1) to 2006 (Vol. 28, No. 2). The reason for setting this period was because, in the preliminary study, the publication time period of the first articles by the top 20 IF people was generally in 1990-1993. A period of 16 years allows the generation shift and keyword changes in the scholars coming in and those retiring in the coauthor network to be examined. The total number of papers submitted during this period was 318 including reviews and symposiums, and there were 643 coauthors. Although review and symposium articles have different characteristics from research articles, they did not affect the IF ranking in the analysis. However, the IF of core researchers increased somewhat.

After relevant data were coded through published KJE, an SAS/IML procedure was created which listed *k* number of authors in n papers as rows and columns. They were then transformed to obtain a 318*643 affiliation matrix. For attribute variables, such elements as paper¡s publication year and author¡s affiliation were also added. These attribute values were used to distinguish not the effects of the network but the individual effects of researchers.

II. Methods

It is important in network analysis to adequately define nodes and links. In coauthor analysis, nodes are clear because they are individual researchers, but links can be variously defined. This paper focused on the core researchers¡ unintended clusters in the field of epidemiology using coauthor data, and selected the method of giving weight according to author status and contribution rate. The reason for this is because, when research productivity and citation index are being obtained, the method of 'directly counting' [16] cannot reflect the coauthors¡ relative contribution ratio and thus produces distorted results, and in Western Europe, discussions of improving such calculation method have taken place since 1980 [17,18]. In order to solve the problem, this paper takes note of the fact that the contribution rate of first author and corresponding author is higher than the other authors, and proposes a model that assigns specific weights according to the author status and, at the same time, even considers how many coauthors there are.

The design of the model is as follows. First, in order to reflect author status, set values were given to first author and corresponding author, and first corresponding author; second, the contribution rate was distributed to correspond to the number of a paper¡s coauthors; third, the sum of the rows were standardized to make them regular. Because the extracted network is a two-mode network with different rows and columns, through a matrix transformation, a one-mode network composed of row and rows was created. Finally, 643*643 matrix was used to analyze.

III. Statistical Analysis

In order to analyze coauthor and keyword networks, this paper applied the following method [19]. First, the neighborhood degree of network was examined. A neighborhood degree refers to the number of links that connect two nodes that are set in a network. In other words, Z_{ijk} means the degree of relationships connected from i number of actors to j number of actors in k number of networks, or the degree of connection in the other direction. In this paper, the matrix was symmertrized to get rid of directions.

$$\deg ree_{ik} = \sum_{j=1}^{N} Z_{ijk} = Z_{ik}$$

Next, the correspondence analysis is the matrix relationship between node and node, it calculates the way a node relates to another node with different qualities. This is called a two-mode affiliation network. In this paper, the integration of author and keyword was applied here.

SAS/IML 9.1 was used as the analysis program for carrying out matrix transformation by calculating the contribution rate. NetMiner 3.1 was used to analyze the total matrix that was derived, and NetMiner 2.6 professional was used for additional analysis on core coauthor network.

RESULTS

I. Whole Network Analysis

The number of nodes in the 643*643 network was 4,830, the mean, 7.512, standard deviation, 7.437, and its range was from 0 to 67. There were 12 isolates, 34 pendants, and the network inclusiveness was 98.13%. In other words, only 1.87% of the researchers made independent submissions, and most of the researchers who made more than two submissions conducted joint studies. Also, just as regular quotation networks show strong cohesion centered round a few core authors, the coauthor network derived in this paper showed similar aspects as well.

To give an example from the results of the analysis, group A is a network connecting Lee JB, Lee HJ, Hyun BH, Bhang JH, Nam KO, Jung YE, Shin YH (Figure 1), and they conducted a joint study on a paper called "Epidemiology and Prevention Strategies of Rabies in Korea [27(1), 2005]". Among them, Lee HJ and Shin YH are also connected to a different network through another paper. In contrast, group B is separated from the total network, but forms an internally tight coauthor network. It is largely a combination of two groups, one of which is centered on Lee TY and connected with Kim SY, Lee SG, Kwon YH, Lee GH, and it is the research team belonging to Chung-Ang University, which published the papers in 1997 [19(2)], 2000 [22(2)], 2003 [25(2)], 2006 [28(1)]. The other is Choi SY's coauthor network which connected with Lee TY's network, and it is the research team belonging to Korea Cancer Center Hospital. The last group C is a group that clearly shows the structural location of network, and it shows Hong JH, who conducted joint studieswith different affiliations in 2004 [26(1)] and 2004 [26(2)], occupying a mediated position between two groups and having high network effectiveness in comparison to the results of his joint studies.

[insert Figure 1. about here]

II. Core Coauthor Network Analysis

Because it is difficult to effectively show KJE¡s core coauthor network due to the vastness of the total network data, the top 57 people with IF of over 10 were selected (node=57, link=192). The result of examining the neighborhood degree using the Jaccard coefficient showed that the IF mean was 3.368 and S.D. was 2.396. All sets that had a 2-step distance relationship in the total coauthor network were collected in order to examine the potential relationships in core coauthor network, and then only the matrix of 57 people were derived.

According to the results, two clusters formed with the link between Min YS and Kim MK serving as the standard, a large cluster to the left and a small cluster to the right (Figure 2). Here, five bi-components were derived, and their densities were 0.7, 1, 0.75, 0.229, and 1. In particular, people such as Chun BY, Shin EC, Lee JS, Bae GR, Kang DH, Kim MK were found to conduct joint studies comparatively easily with the core coauthors. Here, bi-components are relational clusters that remain after cutpoints or bridges have been eliminated when at least two different paths exist between two couple nodes [19]. When there are many different bi-components between core coauthors, it signifies that joint studies have become that much differentiated.

[insert Figure 2. about here]

The most important characteristic of coauthor network is the power law distribution. This does not show the normal distribution form, which focuses on the mean of the cases, but rather shows the form of diminishing with exponential growth in a small number of concentrated cases [20]. In this paper, the IF of coauthors in which contribution rates was apportioned and calculated showed a power law distribution.

The IF value was the highest in Meng KH (109.97), followed by Lim HS (103.44), Kim JS (100.81), Jee SH (83.03), Ohr HC (68.17), Choi BY (67.35), Lee WC (53.50), Shin HR (50.29), Yoo KY (47.60), Chun BY (46.06) and others. In other words, they held a firm position among the coauthors based on their high research productivity. However, high influence in network does not necessarily correspond to the number of papers. Here, the problem of efficiency is considered. According to the contribution rate calculation formula applied in this paper, it is ultimately advantageous to have a specific author status in joint studies. If the number of coauthors is small, then contribution rate rises. If this becomes extreme, however, then there are not enough links within the network, and exclusion from the core can occur.

When the characteristics of research activities of the top 20 IF people which reflect author status were examined (Table 1), the higher the published papers, the higher their ranking was comparatively. However, researchers who had the role of first author or corresponding author and had low average number of coauthors showed a relatively high position compared to the number of their papers. This was above the given contribution rate weight, and it was interpreted to be a result of their conducting many joint studies with others authors who had high research productivity. Only, the weight of the author¡s position somewhat increased when the number of coauthors smaller than when it was a fixed quantity.

[insert Table 1. about here]

Science and Technology Studies points out that research results are also affected by social position [2]. In particular, depending on the size and productivity of the group a researcher is affiliated with, as well as whether or not there are core authors, the individual coauthor¡s network can quickly grow or expand. Power law distribution occurred also in the results of the collected published papers from 97 organizations that included 643 coauthors, and the highest productivity was shown by College of Medicines, Seoul National Univ. (137 times), and the other medical schools were as followed; Catholic Univ. (88), Yonsei Univ. (84), Dongguk Univ. (78), School of Public Health, Yonsei Univ. (62), School of Public Health, Seoul National Univ. (55), Hanyang Univ. (53), Kyungpook

National Univ. (34), Korea Univ. (33). Among these organizations, universities had at least one affiliated coauthor who was in the top 30 IF ranking.

It is important, however, not only to distinguish social background element but also individual effect and network effect in coauthor network. For this, the centrality index and structural hole index of the top ten IF people who occupy the center of the network (Table 2). Here, centrality is divided into degree centrality and whole centrality. The former is proportional to the number of linked nodes, and the latter is again divided into two types, a closeness centrality that is inversely proportional to the sum of the least path distance that links two nodes, and a status centrality that shows how much linkage exists to nodes with high level of influence [19]. Generally, a closeness centrality represents a whole centrality. On the other hand, structural hole is measured by effectiveness, efficiency, aggregate, hierarchy index, and it calculates whether or not redundancy is small in researchers¡ positions in a coauthor network and constraints are few in conducting joint studies with other researchers. In other words, the centrality index is determined by the extent a researchers conducts joint studies, and the structural hole index is determined by the location a researcher occupies in the total network.

According to the results of calculation, there were considerable amount of fluctuations for each index with the top 20 IF people. Here, those who had reached the top ranks were researchers who had high research productivity and, at the same time, had various research topics. However, the ranking of centrality index and structural hole index were mostly different, and it showed that a significant part of coauthor network was determined not only by research productivity but by the relationship of coauthors.

[insert Table 2. about here]

III. Keywords Network Analysis

The keyword network analysis was carried out to supplement coauthor network, and it consisted of examining the keywords of 318 papers in total and deriving a keyword that represented each paper. Specifically, in order to raise the validity of the analysis, similar keywords were combined, and as a result, a total of 131 keywords were assigned to the papers. The 131*643 affiliation matrix of keywords and coauthors were transformed into a one-mode keyword network and classical-MDS was derived.

c-MDS relatively configures the words according to their relevance, and keywords that have higher frequency of being researched by the same coauthors have shorter distance from each other. Therefore, keywords that were jointly researched are positioned at the center of the coordinates, but keywords that were specially researched are located far from the center point. The frequency ranking of keywords aggregated in the lexicon shows smoking (58 times) at the top, followed by Hepatitis (50), Dysentery (48), Breast cancer (47), Cancer (45), Methodology (37), Cholera (30), Infectious disease (30), Measles (29), Gastric cancer (28). The results of conducting c-MDS analysis show that KJE keywords have a relatively clearer distribution than that of JPMPH. Specifically, keywords accumulated in research are divided into cancer, cardiovascular disease, health behavior and their methodology.

According to the result, these keywords were as followed: Group A was Breast cancer, Cancer, Gastic cancer, Group B was Infectious disease, Parotitis, Measles, Exposure, Rubella, Blood pressure, Group C was Methodology, Smoking, Lung cancer, Hepatitis, Obesity, Dysentery, Group D was Cardiovascular disease, Cholera, Malaria and so on. These keywords can be considered KJE¡s special research fields which are continuously researched by certain coauthors.

IV. Research Productivity Analysis

In order to verify the cluster tendency of coauthor and network examined up to this point, the total number of published papers, which is researchers¡ productivity index, and the determinants on IF were obtained through multiple regression (Table 3). The explanation variables were the numbers of first authors, corresponding authors, first corresponding authors and the first article published year. The response variables were the numbers of total articles and IF index. Because the analysis period was 1991~2006, even if one¡s first published year was before that period, it was given the same value for 1991.

The results of classifying author status form according to gender showed that the total number of papers by men were significantly higher than that by women (<.036). The first corresponding authors¡ mean (<.048) and corresponding authors' mean (<.000) were also significantly high. However, in the number of published papers by first author, there was no significant difference. One notable point is that when compared with JPMPH, the first publication year for women appeared about two years earlier.

[insert Table 3. about here]

The independent variables of a multiple regression went through the following stages due to the characteristics of coauthor network¡s power law distribution. First, they were tested to see if they satisfied the normal i.i.d. condition through the scatter plots on the residuals. Here, some outliers were excluded, and in the case of IF among the dependent variables, because the skewness was high, log was adopted. Second, the autocorrelation that originated from the characteristics of the data was examined. Because the number of papers need to be over a certain amount in order to discover the effects that follow the positional characteristics such as first author or corresponding author, cases in which IF was fewer than two were excluded, which resulted in 363 people being used for analysis. Third, the multicollinearity problem was dealt with. When the VIF of a variable was over 1.2, it was excluded from the model.

According to the results of the analysis, the frequency of first author, corresponding author, first corresponding author, first article published year, and network status index were found to influence total number of papers and IF (Table 4). These are determinants of joint studies that are distinguishable from existing studies [21]. However, for total number of papers, the frequency of first author was found to be not significant, and for IF, the corresponding author and closeness centrality were found to be not significant. These results are thought to be due to the fact that the values of IF were readjusted not only simply through paper productivity but also through the actual contribution rate in joint studies, and in this process, only closeness centrality developed in the coauthor network. Only, in the case of SCI journal, which shows a stronger power law distribution than single journals, there are reports that this degree centrality was significant [22].

[insert Table 4. about here]

V. Coauthor Agreement Analysis

Lastly, in order to find the correlation of coauthors between two journals, an agreement analysis was conducted on the top 10% of the IF in KJE and JPMPH, 64 people and

130 people, respectively, during the same time period. According to the results, using KJE as standard, 33 (51%) people among 64 people were also ranked in JPMPH. Specifically, in the top 5%, 23 (71%) people among 32 people were active in the top 10% of JPMPH¡s IF (Table 5). These results show that the academic activities of the two journals are fairly intimately related. However, for the next-to-top group in KJE, the 5-10 %, only 10 (31%) people among 32 people were in the top 10% of the JPMPH¡s IF. It therefore seems that a small group of researchers with outstanding research productivity carried out many of the core academic activities in both journals.

DISCUSSION

Coauthor network analysis appeared along with the development of randomness network, with the attention being paid to the collaboration network of the scientist community [23,24]. Because academics evolve through researcher clusters, the knowledge of how their interactions are carried out has many implications [25,26]. STS, however, goes a step further and requires an in-depth approach to the aspects of how scientific knowledge is formed [1,2]. On the one hand, this is questioning whether or not the process of producing conclusions through experiments and verifications is being influenced by social factors, and on the other hand, this is a reflection of the community in which scientific knowledge is being produced and shared perhaps becoming structuralized into a specific system type. In particular, as pointed out in ANT (Actor-Network Theory), a subfield of STS, actors and networks form a unique relationship in which each composes the other and they cannot be reduced to any one part [27]. The power law and coauthor-keyword clusters of research productivity shown in this paper reflect the workings of ANT. In other words, it shows that the scientist community is relationally composed.

The significance of the coauthor network analysis on KJE is as follows. First, compared to JPMPH, KJE¡s entire network is differentiated. One of the reasons is that KJE¡s publication subjects are more limited than those of JPMPH, but the affiliations of its coauthors were relatively more homogeneous, and there was a greater tendency to consistently maintain the coauthor relationships. As a result, many bi-components were found. Second, the frequency of IF or keywords showed the characteristics of power law, with a small number of authors and research subjects representing the whole journal. However, because its researcher cluster tendency was stronger than that of JPMPH, the productivity of coauthors and the power law of keywords were relatively inactive. Third, core coauthors each had functional positions that they occupied in the network, and there was less inequality in the research subjects. As a result, relatively more groups with specialized studies were derived in the correspondence analysis. Fourth, the effects of various indexes that estimate research productivity were different. The significance of author status was different with not only gender but also productivity determinants, and these differences seem to be due to the differences in the aspects of structuralized coauthor network. Fifth, compared to JPMPH which generally has diversity as its characteristic, KJE showed a form of detailed specialized fields that were divided. Only, it could be that the differences in the yearly publication frequency known as institutional condition exerted influence in bring about such result.

KJE¡s scientist community has been seriously thinking about the specialization of research activities like JPMPH [28,29]. The coauthor and keyword network analysis can be used as evaluation data for preparing for this developmental direction, and specifically, it helps in diagnosing the current situation and obtaining a long-term developmental plan. This research compares and expands the coauthor network analysis on JPMPH to KJE, but in the future, the same model should be applied to KJHPA (Korean Journal of Health Policy and Administration), and the academic structuralization of the whole public health field should be evaluated. It was also shown that the safety of the model should also be obtained. In this paper, a bias was found that amplified IF relatively more when authors conducted a small number of independent studies rather than being first corresponding authors in many joint studies. Because this was not compensated with weight adjustments and could only be understood as a peculiarity of the case, an improvement plan needs to be found.

CONCLUSION

This paper analyzed the network form of joint studies that occurred in 643 coauthors and 131 keywords, using as subjects 318 papers published in KJE for the past 16 years. The results showed a small world effect according to the development of a random network. In particular, as degree centrality was more developed than whole centrality in KJE, a network differentiation was discovered.

REFERENCE

1. Bloor D. Knowledge and social imagery. Chicago:University of Chicago Press;1973,pp.1-15
2. Barnes B, Bloor D, Henry J. Scientific Knowledge: A Sociological Analysis. London:Athlone;1996
3. Bourdieu P. The Specificity of the scientific field and the social conditions of the progress of reason. *Soc Sci Inform* 14:19-47; Reprinted in Biagioli. Eds. The science studies reader. London:Routledge;1975.
4. Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972;178:471-479
5. Garfield E. Of nobel class: a citation perspective on high impact research authors. *Theor Med* 1992;13(2):117-135
6. Yoon B, Park Y. A text-mining-based patent network: Analytical tool for high-technology trend. *J High Technol Manage Res* 2004;15:37-50
7. Li M, Fan Y, Chen J, Gao L, Di Z, Wu J. Weighted networks of scientific communication: the measurement and topological role of weight. *Physica A* 2005;350:643-656
8. Malin B, Carley K. A longitudinal social network analysis of the editorial boards of medical informatics and bioinformatics journals. *J Am Med Inform Assoc* 2007;14:340-348
9. Narin F, Moll JK. Bibliometrics. *Ann Rev Inform Sci Tech* 1977;12:35-58
10. Watts DJ. Small worlds: the dynamics of networks between order and randomness. NJ:Princeton University Press;1999
11. Buchanan M. Nexus: small worlds and the groundbreaking science of networks. NY:W.W.Norton&Company;2002,p.235
12. Erdös P, Renyi A. On the strength of connectedness of a random graph. *Acta Math Acad Sci Hung* 1961;12:261-267
13. Merton RK. The sociology of science: theoretical and empirical investigations. Chicago:University of Chicago Press;1973,p.267-412
14. Murray JD. Mathematical biology, v.1. an introduction. NY:Springer;2002,pp.315-393
15. Jung MS, Chung DJ. Co-author and Keyword Network and their Clustering Appearance in Preventive Medicine Fields in Korea: Analysis of Papers in the Journal of Preventive Medicine and Public Health, 1991~2006. *J Prev Med Public*

Health 2008;41(1) (in press) (Korean)

16. Cole J, Cole S. Social stratification in science. Chicago:University of Chicago Press;1973,p.33
17. Lindsey D. Production and citation measures in the sociology of science: the problem of multiple authorship. *Soc Stud Sci* 1980;10(2):145-162
18. Bremholm TL. Author productivity and citation frequency in the proceedings of the Oklahoma Academy of Science, 1921-2000. *Proc Okla Acad Sci* 2004;84:53-66
19. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge:Cambridge University Press;1994,pp.167-214
20. Barabasi AL. Linked: the new science of networks. Cambridge:Perseus;2002
21. McDowell JM, Michael M. The determinants of coauthorship: an analysis of the economics literature. *Rev Econ Stat* 1983;65:155-160
22. Lee HJ. The social structure of science and technology: network analysis for coauthorship citation and keywords[thesis]. Korea:Yonsei Univ;2003,p.99-106 (Korean)
23. Newman MEJ. Scientific collaboration networks: I network construction and fundamental result. *Phys Rev* E 2001;64:131
24. Newman MEJ. Scientific collaboration networks: II shortest paths, weighted networks, and centrality. *Phys Rev* E 2001;64:132
25. Price DJ. Little science, big science ...and beyond. NY:Columbia University Press;1986
26. Glanzel W. Coauthorship patterns and trends in the Sciences (1980-1998): a bibliometric study with implications of database indexing and search strategies. *Libr Trends* 2002;50(3):461-473
27. Callon M, Latour B. Unscrewing the big leviathan: how actors macro-structure reality and how sociology help them to do so. in Knorr-Cetina, Cicouvel A. Eds. Advances in social theory and methodology: towards an integration of micro and macro-sociology. London:Routledge;1981
28. Ahn YO, Shin MH. The role and activities of clinical epidemiologists. *Korean J Epi* 1994;16(1):20-27 (Korean)
29. Lim HS. Future of scientific research on preventive medicine in Korea. *J Prev Med Public Health* 2006;39(2):105-109 (Korean)

Table 1. Number of papers and author status characteristics among impact factor top 20 authors

	Total paper	Number of A*	Number of B*	Number of C*	Mean of co-author	Year**	Impact factor
Meng KH	22	0	1	10	3.909	1991	109.97
Lim HS	25	1	6	12	3.800	1992	103.44
Kim JS	24	1	1	7	3.375	1991	100.81
Jee SH	25	0	4	10	4.920	1991	83.03
Ohr HC	18	0	0	10	4.670	1991	68.17
Choi BY	17	0	3	9	5.120	1991	67.35
Lee WC	14	0	4	5	4.571	1993	53.50
Shin HR	14	0	1	6	6.643	1992	50.29
Yoo KY	17	0	3	5	6.059	1991	47.60
Chun BY	11	0	3	5	4.273	1992	46.06
Cheong HK	12	0	2	5	4.833	1992	45.02
Bae JM	12	0	2	5	4.417	1993	41.27
Nam CM	11	0	0	5	4.091	1991	35.39
Song YM	6	0	0	5	3.167	1992	35.00
Ahn YO	11	1	1	3	4.545	1993	33.76
Cho BM	4	0	1	3	2.000	1996	32.00
Chun BC	7	1	3	2	4.143	1997	28.41
Park BJ	7	0	2	3	4.429	1993	26.11
Suh I	11	0	0	2	5.364	1991	23.83
Kim SD	9	0	6	1	4.222	1994	23.49

*Author status: A(1st author), B(correspondant author), C(both A and B author)

**Year: Year of the first paper listed

Table 2. Network indexes among impact factor top 10 authors by individual effect and structural effect

Centrality measure			Structural Holes measure			
degree	status*	closeness	efficiency	effective	aggregate	hierachy
Chun BY(18)	Choi SY(1.4)	Lee KS(0.29)	Choi BY(1.0)	Chun BY(8.1)	Son MA(0.0)	Choi BY(1.0)
Kim SD(14)	Kim SW(1.3)	Bae GR(0.29)	Lee WC(1.0)	Kim SD(6.2)	Lee DH(0.0)	Lee WC(1.0)
Kang DH(14)	Lee CW(1.1)	Chun BY(0.29)	Shin HR(1.0)	Shin EC(5.8)	Sung JH(0.0)	Song YM(1.0)
Choi SY(14)	Cho BM(1.1)	Min YS(0.28)	Song YM(1.0)	Bae GR(5.8)	Choi SO(0.0)	Chun BC(1.0)
Shin EC(14)	Kim MK(1.0)	Shin MH(0.28)	Chun BC(1.0)	Kang DH(5.4)	Lee TY(0.0)	Park SK(1.0)
Kim MK(14)	Meng KH(1.0)	Lee JS(0.27)	Suh I(1.0)	Lee SE(5.4)	Chun BY(0.25)	Lee SK(1.0)
Lee SE(14)	Cheong HK(0.9)	Cho BM(0.27)	Park SK(1.0)	Choi SY(5.1)	Bae GR(0.26)	Lee K(1.0)
Lee JS(14)	Lee KS(0.9)	Chun JH(0.26)	Lee SK(1.0)	Kim SW(4.9)	Lee CW(0.36)	Yeh MH(1.0)
Kim SW(14)	Yoo KY(0.8)	Meng KH(0.26)	Lee K(1.0)	Kim MK(4.9)	Kim SD(0.37)	Park O(1.0)
Cheong HK(12)	Kim SD(0.8)	Jung MH(0.26)	Yeh MH(1.0)	Lee CW(4.9)	Chun JH(0.40)	Koh SB(1.0)

*Kats status, attenuation factor=0.5

Table 3. Mean scores of independent variables by sex

Items	Male(n=236)	Female(N=113)	p-value
# of 1st author(A)	0.20¡0.02	0.24¡0.04	0.433
# of corr. author(B)	0.29¡0.05	0.05¡0.02	<0.000
# of A and B	0.77¡0.11	0.48¡0.09	<0.048
# of total paper	2.83¡0.24	2.11¡0.24	<0.036
impact factor	8.83¡0.94	6.30¡0.97	0.062
year of the first paper listed	1997.07¡0.30	1998.88¡0.43	<0.001

*p<0.05 by t-test

Table 4. A result of multiple regression both number of papers and impact factor

Dependant variable		Unstandardized coefficient		Standardized coefficient	t	R^2	Adj. R^2
		B	std. err.	Beta			
number of enlisted papers	(constant)	59.876	11.611		5.157		
	1st author(A)	.109	.058	.064	1.893		
	corr. author(B)*	.128	.041	.123	3.091		
	both A & B*	.070	.023	.144	3.115	0.636	0.630
	year_od**	-.030	.006	-.187	-5.165		
	degree*	.008	.003	.105	3.099		
	closeness**	.519	.042	.519	12.277		
impact factor	(constant)	42.620	11.296				
	1st author(A)**	.276	.056	.150	4.928		
	corr. author(B)	.041	.040	.036	1.012		
	both A & B**	.316	.022	.596	14.396	0.707	0.702
	year_od**	-.021	.006	-.120	-3.680		
	degree	-.000	.002	.000	-.008		
	closeness**	.262	.041	.241	6.358		

*p<0.01, **p<0.001 by multiple regression

Table 5. A Comparison with IF ranks of KJE and JPMPH

Author	5% rank of *KJE*[*]	10% rank of *JPMPH*[**]
Meng KH	1	56
Lim HS	2	1
Kim JS	3	7
Jee SH	4	81
Ohr HC	5	16
Choi BY	6	59
Lee WC	7	-
Shin HR	8	49
Yoo KY	9	61
Chun BY	10	23
Cheong HK	11	15
Bae JM	12	12
Nam CM	13	17
Song YM	14	-
Ahn YO	15	32
Cho BM	16	-
Chun BC	17	46
Park BJ	18	22
Suh I	19	26
Kim SD	20	88
Chun JH	21	27
Kang DH	22	101
Choi SY	23	-
Shin MH	24	121
Son MA	25	21
Jung MH	26	-
Park JK	27	9
Lee SY	28	11
Min YS	29	-
Shin EC	30	-
Ki MR	31	-
Kim MK	32	-

[*]The Korean Journal of Epidemiology

[**]The Journal of Preventive Medicine and Public Health

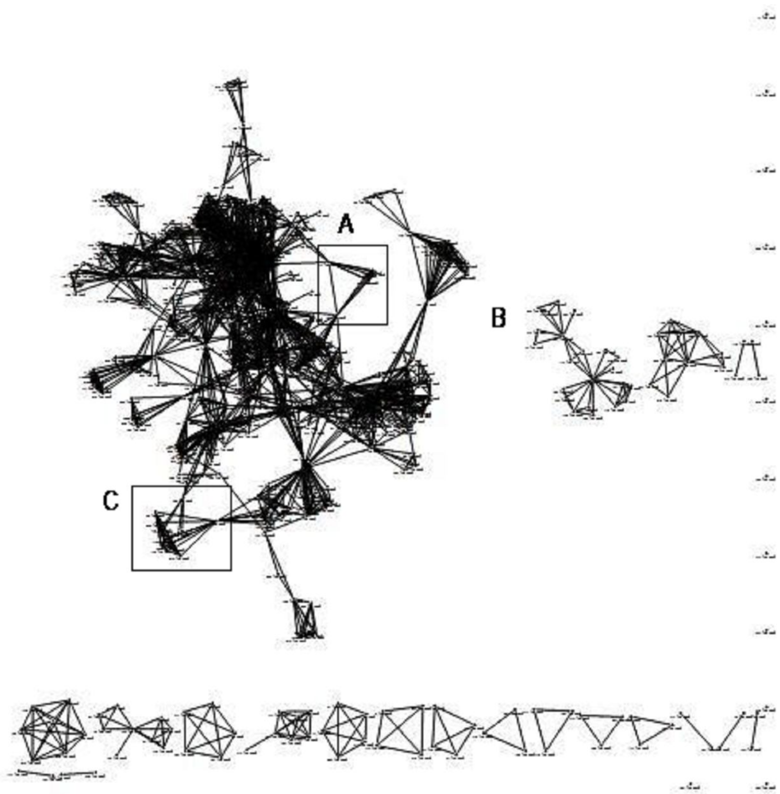

Figure 1. Co-author matrix drew by GEM (643*643)

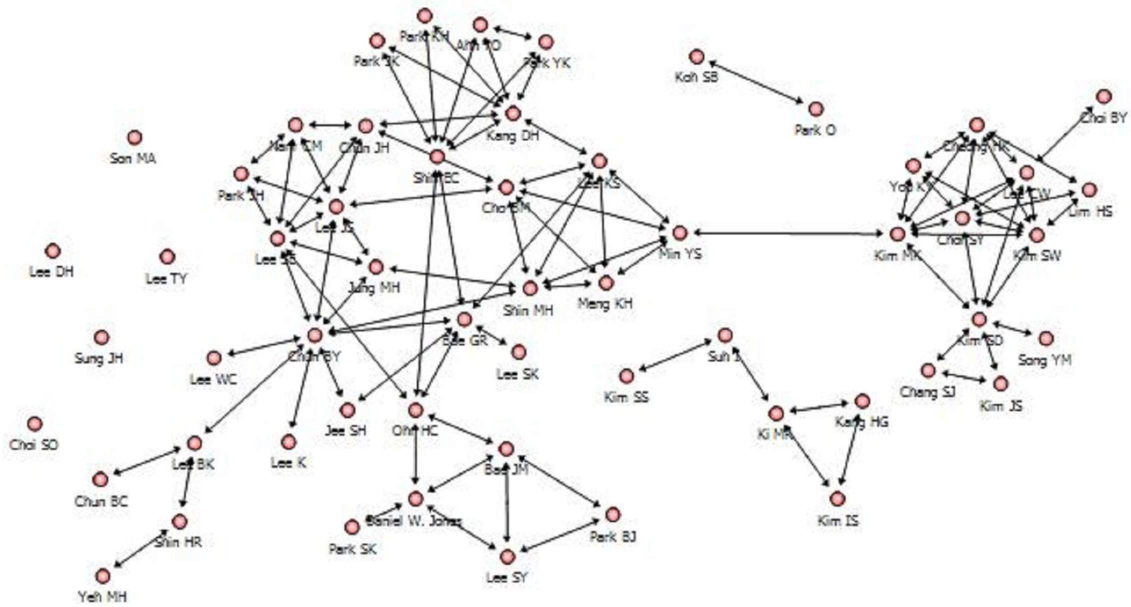

Figure 2. Core co-authorship networks using by FR* (57*57; imfact factor≧20)
*Spring-FR (Fruchterman and Reingold)

<Appendix>

An example of matrix transformation procedure is as followed.

Step 1. An example dataset of author and paper

paper	author	author status*	(1) initial weight	(2) adjust weight	(3) standardization
A	1	A and B	0.4	0.4	0.8
A	2	C	1-0.4=0.6	0.1	0.2
B	3	A	0.3	0.3	0.3
B	4	B	0.2	0.2	0.2
B	1	C	(1-0.5)/5=0.1	0.1	0.1
B	2	C	(1-0.5)/5=0.1	0.1	0.1
B	5	C	(1-0.5)/5=0.1	0.1	0.1
B	6	C	(1-0.5)/5=0.1	0.1	0.1
B	7	C	(1-0.5)/5=0.1	0.1	0.1
C	3	A and B	0.4	0.4	0.8
C	8	C	1-0.4=0.6	0.1	0.2

*A(1st author), B(correspondant author), C(none of them)

Step 2. Weighted matrix of co-author and paper

author \ paper	A	B	C	sum of weight
1	0.8	0.1	0	0.9
2	0.2	0.1	0	0.3
3	0	0.3	0.8	1.1
4	0	0.2	0	0.2
5	0	0.1	0	0.1
6	0	0.1	0	0.1
7	0	0.1	0	0.1
8	0	0	0.2	0.2

Step 3. Co-author matrix by impact factor (diagonalized=0)

	1	2	3	4	5	6	7	8
1	0	0.17	0.03	0.02	0.01	0.01	0.01	0
2	0.17	0	0.03	0.02	0.01	0.01	0.01	0
3	0.03	0.03	0	0.06	0.03	0.03	0.03	0.16
4	0.02	0.02	0.06	0	0.02	0.02	0.02	0
5	0.01	0.01	0.03	0.02	0	0.01	0.01	0
6	0.01	0.01	0.03	0.02	0.01	0	0.01	0
7	0.01	0.01	0.03	0.02	0.01	0.01	0	0
8	0	0	0.16	0	0	0	0	0