

JKSS의 공저자 및 핵심어에 대한 사회네트워크 분석과 응용

김종건^a, 최순국^b, 최용석^{1, b}

^a부산과학기술대학교 컴퓨터정보계열 ^b부산대학교 통계학과

요약

사회네트워크분석은 사회네트워크 구조를 시각화한 네트워크 그림에서 중요한 위치를 차지하는 노드(사람, 기업, 기관 등)와 노드들 간의 관계와 특징을 찾는 분석기법이다. 그러나 네트워크 그림에서 노드들의 특징을 찾는 것은 쉽지 않다. 따라서 노드들 간의 연관네트워크에 대한 대응분석을 적용하여 그들의 관계에 대한 특징을 보여주려 한다. 본 연구에서는 이러한 사회네트워크 분석과 대응분석을 이용하여 한국통계학회의 학술지인 JKSS(Journal of the Korean Statistical Society)의 공저자와 핵심어를 중심으로 게재 논문의 특징을 살펴보려 한다

주요용어: 사회네트워크 분석, 중심성, 대응분석, 한국통계학회 학술지.

1. 서론

우리는 처음 만나는 사람인데 알고 보니 친구의 친구처럼 가까운 사이일 경우 세상이 참 좁다는 말을 사용한다(허명희, 2010). 실제로 1967년에 사회학자 Stanley Milgram이 주장한 6단계 분리(six degrees of separation)를 거치면 이 세상의 모두가 서로 아는 사람으로 연결될 수 있다고 했다. 현대사회는 이전 사회보다 인터넷과 스마트폰 등과 같은 각종 통신기술의 발달로 인하여 이 이론에서 주장하는 것보다 사람들의 관계가 더욱더 긴밀하게 이루어지고 있다. 오늘날에는 사회학을 비롯하여 정치학, 경영학, 공학 등의 분야에서 관계의 중요성이 인식되면서 이에 대한 분석이 필요하게 되었다.

사회네트워크 분석(social network analysis; SNA)은 Barnes(1954)에 의해 처음 사용된 용어로, 사회네트워크 현상에서 관계 구조를 시각화하여 중요한 위치에 있는 사람, 기업, 기관 등을 찾아내어 다양한 영역에 적용시킬 수 있는 분석기법이다(손동원, 2002). 최근 이 분석을 응용한 사례를 살펴보면, 정민수와 정동준(2008)은 대한예방의학회지 게재논문을 분석하여 공저자·핵심어 네트워크와 군집 양상을 살펴보았고, 김재경 외 3인(2009)은 충성고객과 이탈고객의 구매 특성을 비교 연구 하였다. 그리고 최승배 외 3인(2011)은 축구경기 분석을 하였다.

한국통계학회(2011)은 한국통계학회 학술지인 JKSS를 일정 주제로 분류하여 각 주제에 따른 논문과 저자에 대한 빈도를 살펴보았다. 그러나 본 연구에서는 이러한 사회네트워크 분석을 활용하여 단순히 빈도뿐만 아니라, JKSS의 SCIE(Science Citation Index Expanded)급 전환 이전과 이후의 특징을 살펴보려 한다. 2절에서는 사회네트워크 분석과 이 분석에서 중요한 대상의 인접행렬(adjacency matrix)을 다변량 그래프적으로 분석하기 위한 대응분석(correspondence analysis)의 기초 이론에 대해 설명하고, 3절에서는 2절의 기법을 활용하여 JKSS의 특징을 살펴본 후, 끝으로 4절에서 본 연구를 정리·요약하려 한다.

¹ 교신저자: (609-735) 부산 금정구 장전동 산30, 부산대학교 통계학과 교수. E-mail: yschoi@pusan.ac.kr

2. 사회네트워크 분석 및 대응분석

2.1절에서는 손동원(2002, 5장)과 허명희(2010, 1장과 2장)를 참고하여 사회네트워크 분석의 정의와 자료의 구조에 대해서 알아보고, 2.2절에서는 이 분석을 통하여 형상화된 네트워크의 특징을 파악하기 위한 기법인 대응분석을 최용석(2001, 2장)을 참고하여 소개하려 한다.

2.1. 사회네트워크 분석

사회네트워크 분석에서는 사회적 구조를 간단하게 노드(node)와 선(link)으로 표현하는데, 노드는 개인, 사회, 집단, 조직 등으로 정의되고, 선은 이들의 정보거래, 친구관계, 직무거래 등 다양한 관계로 정의된다. 네트워크에서는 중요한 노드 또는 네트워크 흐름에 중요한 역할을 하는 노드를 찾는 것이 중요하며 이를 나타내는 척도로 중심성(centrality)이 사용된다.

각각의 노드는 인접행렬을 통해 행과 열로 나타낼 수 있고, 행과 열이 교차되는 셀은 관계(선)가 있으면 1, 없으면 0의 값을 가질 수 있다. 즉 q 개의 노드를 가진 인접행렬을 $A = \{a_{ij}\}$, $i, j = 1, 2, \dots, q$ 라 할때, $a_{ij} = 1$ 이면 노드 i 에서 노드 j 로 가는 관계가 있음을 의미한다. 인접행렬은 행과 열이 동일 대상을 나타내지만 다른 대상일 경우 n 개의 행 노드와 m 개의 열 노드의 연결로 교차자료가 구성되고, 이를 $n \times m$ 연관행렬(affiliation matrix)이라고 한다.

일반적인 수치자료에서 평균 등의 중심위치 파악이 중요한 것처럼 사회네트워크 분석에서도 중심노드를 이용하여 각 네트워크의 중심을 파악한다. 중심노드는 중심성이 높은것을 말하며 네트워크에서 핵심적인 위치에 있으면서 네트워크의 흐름에 주요한 역할을 하며 정보 교류의 핵심 역할을 한다. 중심성을 나타내는 척도는 근접 중심성(closeness centrality)과 매개 중심성(betweenness centrality)으로 나눌 수 있다.

먼저 근접 중심성은 하나의 노드에서 직접 연결된 노드뿐만 아니라 간접적으로 연결된 모든 노드들의 거리를 산출하여 각 노드마다 전체 네트워크에서 중심에 얼마나 위치하고 있는지를 측정하는 방법이다. n 개의 노드로 구성된 네트워크에서 $d(i, j)$ 를 노드 i 에서 노드 j 에 이르는 거리라고 할 때, i 번째 노드에 대한 일반화 근접 중심성 $C_C(i)$ 는 다음과 같이 표현된다.

$$C_C(i) = \frac{\sum_{j \neq i} (1/d(i, j))}{n-1}, \quad i = 1, 2, \dots, n. \quad (2.1)$$

다음으로 매개 중심성은 네트워크 내에서 한 노드가 담당하는 매개자 혹은 중재자 역할의 정도를 측정한 방법이다. 예를 들어 노드 A와 B가 오직 노드 C를 통해서만 관계를 맺을 수 있는 경우 노드 C는 높은 매개 중심성을 갖는다고 할 수 있다. 네트워크 내에서 g_{ij} 를 노드 i 와 노드 j 사이에 존재하는 최단 거리 경로의 경우의 수라 하고, g_{ikj} 를 노드 i 와 노드 j ($i \neq j$) 사이에 존재하는 노드 k 를 경유하는 횟수라 할 때, 노드 k 에 대한 일반화 매개 중심성 $C_B(k)$ 는

$$C_B(k) = \frac{\sum_{j \neq k} \sum_{i \neq k, i} (g_{ikj}/g_{ij})}{[(n-2)(n-1)/2]}, \quad k = 1, 2, \dots, n. \quad (2.2)$$

와 같고, 매개 중심성의 범위는 0과 1 사이이다.

2.2. 대응분석

사회네트워크 분석은 네트워크 그림을 통하여 노드들 간의 연결적 관계만을 파악할 수 있으나, 그들간의 특징을 구체적으로 파악하기 힘들다는 단점을 가지고 있다. 이를 보완하기 위하여

다변량 기법인 대응분석을 적용하고자 한다. 대응분석이란 분할표 자료의 행과 열을 저차원 공간상의 점들로 동시에 나타내어 그들의 관계를 탐구하려는 탐색적 자료분석기법이다.

행의 수가 p 이고 열의 수가 $q(\leq p)$ 인 이원분할표 자료행렬 $\mathbf{O} = \{o_{ij}\}$, $i = 1, 2, \dots, p; j = 1, 2, \dots, q$ 를 2.1절에서 정의한 연관행렬이라 하자. 연관행렬 $\mathbf{O}$ 로부터 행렬곱적 $\mathbf{O}'\mathbf{O}$ 는 열에 관한 인접행렬이고, $\mathbf{O}\mathbf{O}'$ 은 행에 관한 인접행렬이 된다. 이와 같은 연관행렬 $\mathbf{O}$ 를 이용하여 대응분석을 적용하기 위한 대응행렬(correspondence matrix)은

$$\mathbf{F} = \{f_{ij}\}, f_{ij} = o_{ij}/o_{++} \quad (2.3)$$

이며 o_{++} 는 $\mathbf{O}$ 의 모든 원소들의 전체합이 된다.

대응분석에서 다차원의 정보를 저차원으로 표현하기 위한 대응분석도의 좌표행렬을 얻기 위해 중심화(centred) 대응행렬 $\mathbf{F} - \mathbf{rc}'$ 의 일반화 비정칙치분해

$$\mathbf{F} - \mathbf{rc}' = \mathbf{U}\mathbf{D}_\lambda\mathbf{V}' \quad (2.4)$$

를 이용한다. 여기서 $\mathbf{r} = (f_{1+}, \dots, f_{p+})'$ 과 $\mathbf{c} = (f_{+1}, \dots, f_{+q})'$ 는 크기가 $p \times 1$ 과 $q \times 1$ 인 벡터이며, $f_{i+} = \sum_{j=1}^q f_{ij}$ 과 $f_{+j} = \sum_{i=1}^p f_{ij}$ 이다. 그리고 $\mathbf{D}_\lambda = \text{diag}(\lambda_1, \dots, \lambda_{q-1})$ 는 비정칙치를 대각원소로 하는 대각행렬이고, $\mathbf{U}'\mathbf{D}_r^{-1}\mathbf{U} = \mathbf{V}'\mathbf{D}_c^{-1}\mathbf{V} = \mathbf{I}$ 를 만족하며, $\mathbf{D}_r = \text{diag}(f_{1+}, \dots, f_{p+})$ 과 $\mathbf{D}_c = \text{diag}(f_{+1}, \dots, f_{+q})$ 는 크기가 $p \times p$ 와 $q \times q$ 인 대각행렬이다.

식 (2.4)로부터 대응분석도를 위한 행 좌표행렬은 $\mathbf{D}_r^{-1}\mathbf{U}\mathbf{D}_\lambda$ 이고, 열 좌표행렬은 $\mathbf{D}_c^{-1}\mathbf{U}\mathbf{D}_\lambda$ 이며, $s(\leq q-1)$ 차원의 대응분석도의 설명력은 $\left(\sum_{k=1}^s \lambda_k^2 / \sum_{k=1}^{q-1} \lambda_k^2 \right) \times 100\%$ 이다.

3. 활용 사례

3.1. 연구 자료 및 방법

본 연구에서는 JKSS의 게재 논문들의 특징을 살펴보기 위해 2003년(32권 2호)부터 2010년(39권 4호)까지의 논문들의 정보를 사용하였다. 이 중 2008년 37권 1호의 경우 핵심어가 명시되지 않아 분석 자료에서 제외하였다. 이는 2003년 이전의 자료에서는 각 논문의 핵심어(key-word)가 명시되지 않았고, 2008년 이후부터 JKSS가 SCIE급의 학술지로 변경되면서 SCIE급 전환 이전과 이후의 변화를 살펴보기 용이할 것이라 판단되어 고려하였다. 따라서 연구 자료는 표 1과 같이 278편의 논문, 413명의 저자, 1,059개의 핵심어로 구성되어 있다.

노드는 논문, 핵심어, 저자이고 선은 논문에 핵심어가 포함되어 있으면 관계가 있다고 보고 1, 포함되어있지 않으면 관계가 없다고 보고 0으로 표현하였다. 일반적으로 한 편의 논문은 여러 가지 핵심어들을 이용하여 작성되고, 핵심어들의 중요도는 파악하기 힘들기 때문에 일정하게 가중치를 1로 부여하였다.

3.2. SCIE급 전환 이전의 JKSS

표 1의 SCIE급 전환 이전의 JKSS를 살펴보면 167편의 논문, 235명의 저자, 626개의 핵심어로 구성되어 있다. 그림 1의 (가)는 논문과 핵심어에 대한 167×626 의 전체 연관네트워크 그림이다. 그림을 살펴보면 방대한 양의 핵심어의 사용으로 인하여 구체적인 정보를 확인하기 힘들기 때문에 626개의 핵심어 중에서 최소 2편 이상의 논문에서 제시된 핵심어를 선별하였고, 93편의 논문,

표 1: 연구 자료

구분	발행년도	논문 수	저자	핵심어
SCIE급 전환 이전	2003 - 2007	167	235	626
SCIE급 전환 이후	2008 - 2010	111	198	494
합계(저자 및 핵심어 중복제거)		278	413	1,059

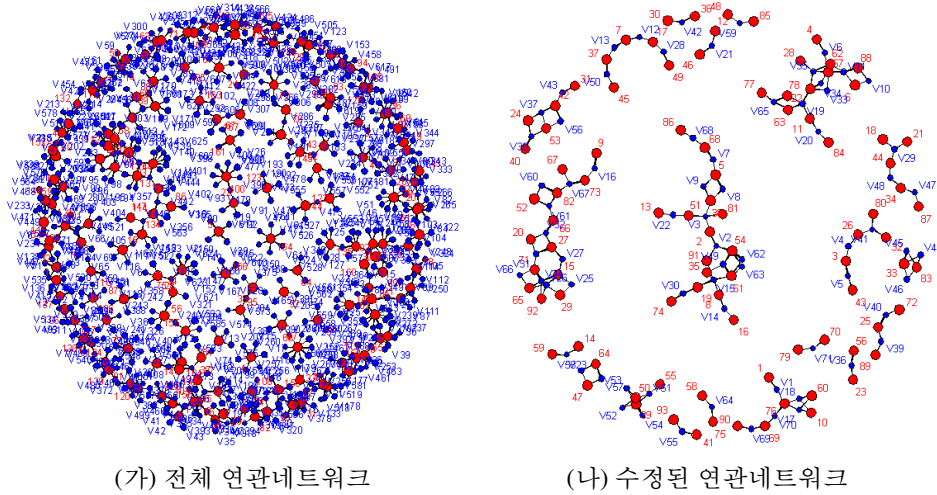


그림 1: SCIE급 전환 이전의 논문과 핵심어 연관네트워크

표 2: SCIE급 전환 이전의 핵심어 네트워크 중심성

핵심어	노드	근접 중심성	그룹	핵심어	노드	매개 중심성	그룹
reference prior	V26	0.1701	B	efficiency	V3	0.0323	A
matching prior	V24	0.1606	B	consistency	V2	0.0298	A
extreme value index	V49	0.1469	A	metropolis-hastings algorithm	V32	0.0199	B
rate of convergence	V35	0.1452	C	asymptotic normality	V15	0.0174	A
asymptotic normality	V15	0.1279	A	markov chain monte carlo	V61	0.174	B
second order parameter	V62	0.1279	A	weak convergence	V19	0.0108	C
extreme value distribution	V63	0.1279	A	coefficient of variation	V7	0.0091	A
jump size	V33	0.1262	C	fractional bayes factor	V27	0.0086	B
one-side kernel	V34	0.1262	C	gibbs sampler	V31	0.0086	B

139명의 저자, 71개의 핵심어들로 수정된 자료에 대한 분석을 실시하였다. 그림 1의 (나)는 수정된 93×71의 연관네트워크 그림이다.

수정된 연관네트워크로부터 핵심어간의 특징을 살펴보기 위해 행렬교적을 이용한 71×71 인접행렬의 핵심어 네트워크는 그림 2와 같다. 이 그림을 보면 크게 세 그룹의 네트워크가 형성되고 나머지 핵심어들은 독립적이거나, 작은 네트워크를 형성하는 것을 알 수 있다. 식 (2.1)에 의한 근접 중심성과 식 (2.2)에 의한 매개 중심성의 값을 기준으로 상위 20%의 노드에 대한 정보는 표 2와 같다. 가장 높은 근접 중심성을 보이는 reference prior(V26)를 중심으로 matching prior(V24) 등이 그룹 B의 핵심어 네트워크를 구성하고, Kim 외 2인(2003), Kim 외 2인(2006), Heo와 Kim(2007) 등에서 핵심어를 공유하고 있다. 그룹 B에서 매개 중심성이 가장 높은 핵심어는 metropolis-hastings algorithm(V32)이고, markov chain monte carlo(V61), fractional bayes factor(V27), gibbs sampler(V31)가

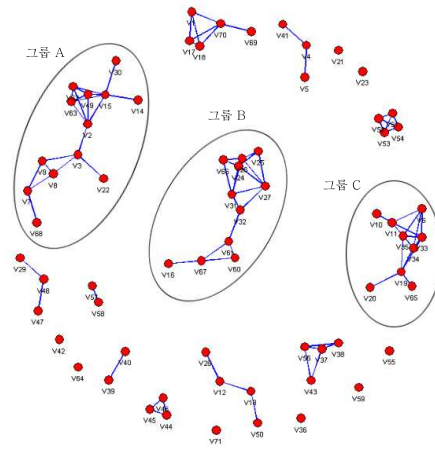


그림 2: SCIE급 전환 이전의 핵심어 네트워크

표 3: SCIE급 전환 이전의 그룹에 따른 중심 핵심어 및 논문

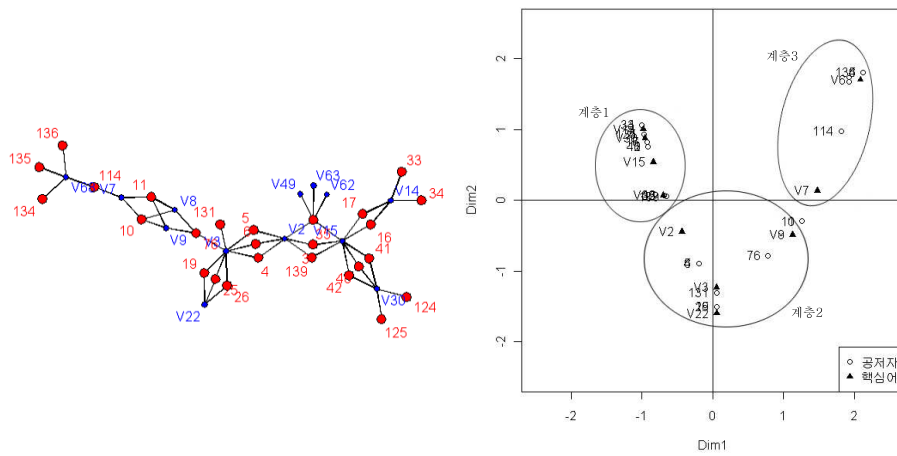
그룹	구분	핵심어 및 논문
A	근접 중심성	extreme value index(V49), asymptotic normality(V15)
	매개 중심성	efficiency(V3), consistency(V2), asymptotic normality(V15)
	공유 논문	Yun(2005), Yun(2006) 등
B	근접 중심성	reference prior(V26), matching prior(V24)
	매개 중심성	metropolis-hastings algorithm(V32), markov chain monte carlo(V61)
	공유 논문	Kim 외 2인(2003), Kim 외 2인(2006), Heo 와 Kim(2007) 등
C	근접 중심성	rate of convergence(V35), jump size(V33), one-side kernel(V34)
	매개 중심성	weak convergence(V19)
	공유 논문	Huh(2004), Kang 과 Huh(2006) 등

매개 중심성이 높게 나타났다. 이들은 그룹 B의 논문들을 이어주는 영향력 높은 핵심어가 된다.

다음으로 높은 근접 중심성을 보이는 extreme value index(V49)를 중심으로 asymptotic normality(V15), second order parameter(V62), extreme value distribution (V63) 등이 그룹 A의 핵심어 네트워크를 구성하고, Yun(2005), Yun(2006) 등에서 핵심어를 공유하고 있다. 그룹 A에서 매개 중심성이 가장 높은 핵심어는 efficiency(V3)이고, consistency(V2), asymptotic normality(V15), coefficient of variation(V7)이 매개 중심성이 높게 나타났다. 이들은 그룹 A의 여러 논문들을 이어주는 영향력 높은 핵심어가 된다.

높은 근접 중심성을 보이는 rate of convergence(V35)를 중심으로 jump size(V33), one-side kernel(V34) 등이 그룹 C의 핵심어 네트워크를 구성하고, Huh(2004), Kang과 Huh(2006) 등에서 핵심어를 공유하고 있다. 그룹 C에서 매개 중심성이 가장 높은 핵심어는 weak convergence(V19)이고, 그룹 C에서 여러 논문들을 이어주는 영향력 높은 핵심어가 된다.

핵심어 네트워크에서 형성된 세 그룹의 네트워크 특징을 자세히 살펴보기 위하여, 세 그룹의 공저자와 핵심어에 대한 대응분석을 이용하여 보았다. 먼저 그룹 A의 핵심어와 공저자 네트워크 그림 및 대응분석도는 그림 3의 (가)와 (나)이고, 그림 (가)에서 매개 중심성이 높았던 efficiency(V3)와 consistency(V2)가 두 부분을 이어주는 것을 확인할 수 있다. 그림 (나)를 살펴보면 근사적합도는 37.46%로 나타났고, 계층1(제2사분면), 계층2(제3·4사분면), 계층3(제2사분면)으로



(가) 네트워크 그림

(나) 대응분석도

그림 3: 그룹 A의 핵심어와 공저자 네트워크 그림 및 대응분석도

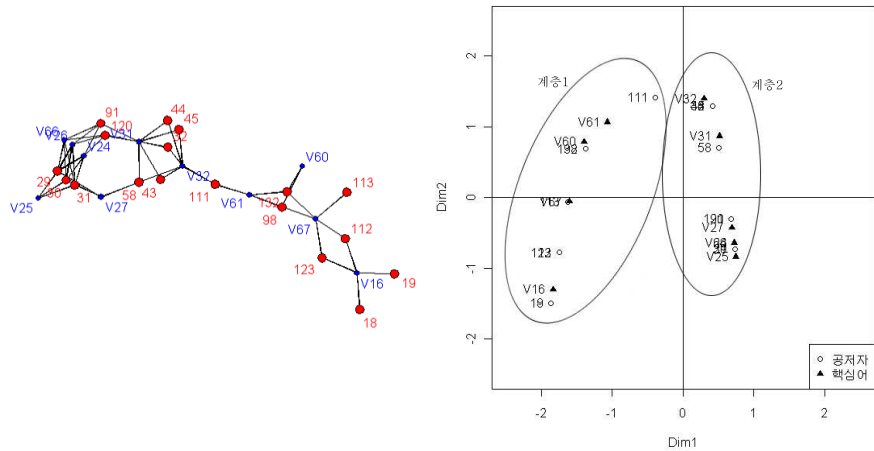
표 4: 그룹 A의 각 계층별 저자 및 핵심어

계층	구분	저자 및 핵심어
1	저자	Narinder Kumar(16), Hyo Il Park(3), A. N. Gill(34), Seok Hoon Yun(35)
	핵심어	asymptotic normality(V15), Robust(V30), extreme value index(V49)
2	저자	Hae Kyung Kim(5), Young Lee(6), N. Y. Kim(26)
	핵심어	efficiency(V3), Multi-level rotation design(V8)
3	저자	Kee Young Kim(11), Yong Hee Lee(114), Jae Won Lee(136)
	핵심어	coefficient of variation(V7), likelihood ratio test(V8), bootstrap(V68)

나누어지는 것을 확인할 수 있다. 표 4와 같이 계층1은 Narinder Kumar(16), Hyo Il Park(3), A. N. Gill(34), Seok Hoon Yun(35) 등의 저자가 asymptotic normality(V15), Robust(V30), extreme value index(V49), extreme value distribution(V63) 등의 핵심어를 이용하여 논문을 공저, 계층2는 Hae Kyung Kim(5), Young Lee(6), N. Y. Kim(26), B. M. Golam Kibria(76) 등의 저자가 efficiency(V3), Multi level rotation design(V8) 등의 핵심어를 이용하여 논문을 공저, 계층3은 Kee Young Kim(11), Yong Hee Lee(114), Jae Won Lee(136) 등의 저자가 coefficient of variation(V7), likelihood ratio test(V8), bootstrap(V68) 등의 핵심어를 이용하여 논문을 공저하였다. 특히 그림 (가)에서 매개중심성이 높았던 efficiency(V3)는 그림 (나)의 계층2에 속하는 것을 확인할 수 있다.

그룹 B의 핵심어와 공저자 네트워크 그림 및 대응분석도는 그림 4의 (가)와 (나)이고, 그림 (가)에서 매개 중심성이 높았던 metropolis-hastings algorithm(V32)과 markov chain monte carlo(V61)가 두 부분을 이어주는 것을 확인할 수 있다. 그림 (나)를 살펴보면 근사적합도는 58.09%로 나타났고, 계층1(제2·3사분면), 계층2(제1·4사분면)로 나누어지는 것을 확인할 수 있다. 각 계층의 저자 및 핵심어는 표 5와 같이 나타난다.

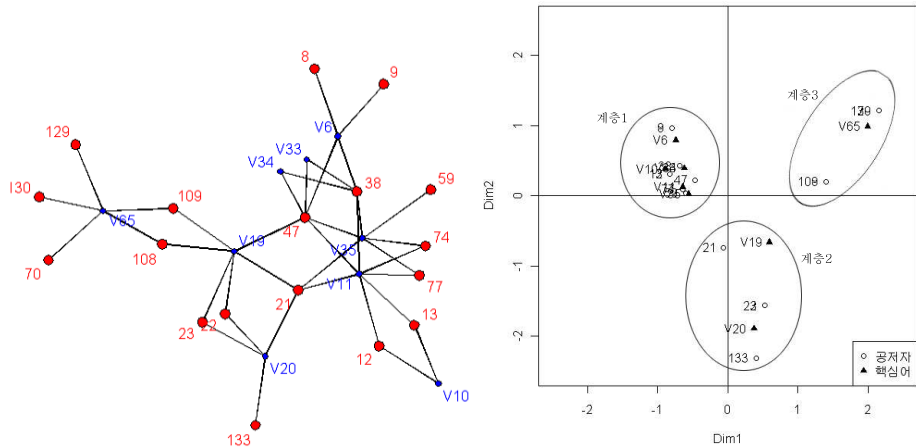
그룹 C의 핵심어와 공저자 네트워크 그림 및 대응분석도는 그림 5의 (가)와 (나)이고, 그림 (가)에서 매개 중심성이 높았던 weak convergence(V19)가 세 부분을 이어주는 것을 확인할 수 있다. 그림 (나)를 살펴보면 근사적합도는 50.31%로 나타났고, 계층1(제2사분면), 계층2(제3·4분면), 계층3(제1사분면)으로 나누어지는 것을 확인할 수 있다. 각 계층의 저자 및 핵심어는 표 6과 같이 나타난다. 특히 그림 (가)에서 매개중심성이 높았던 weak convergence(V19)는 그림 (나)의 계층2에 속



(가) 네트워크 그림 (나) 대응분석도
그림 4: 그룹 B의 핵심어와 공저자 네트워크 그림 및 대응분석도

표 5: 그룹 B의 각 계층별 저자 및 핵심어

계층	구분	저자 및 핵심어
1	저자	Seung Chun Lee(18), Jae Yong Lee(98), Jo Han Lim(112)
	핵심어	variance components(V16), proportional hazard model(V67)
2	저자	Woo Dong Lee(31), Tae Young Yang(44), Young Sook Son(58)
	핵심어	matching prior(V24), fractional bayes factor(V27), gibbs sampler(V31)

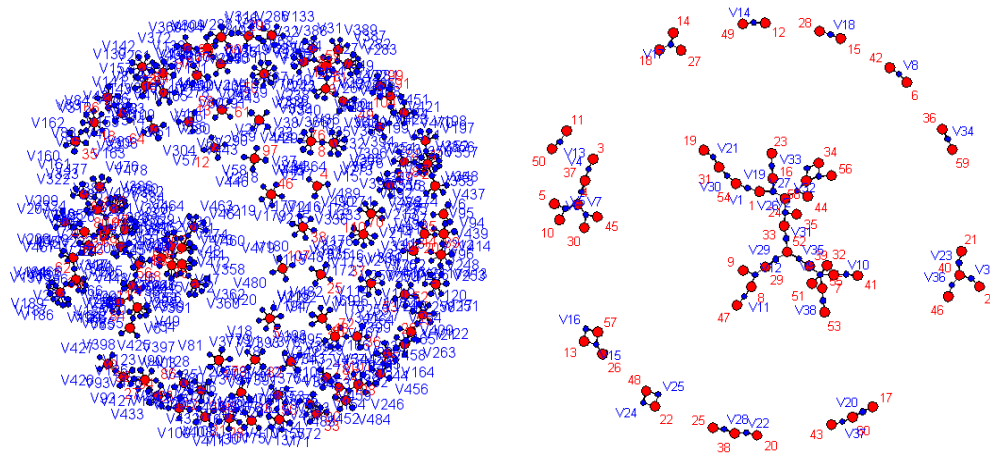


(가) 네트워크 그림 (나) 대응분석도
그림 5: 그룹 C의 핵심어와 공저자 네트워크 그림 및 대응분석도

하는 것을 확인할 수 있다.

표 6: 그룹 C의 각 계층별 저자 및 핵심어

계층	구분	저자 및 핵심어
1	저자	Mee Seon Jeong(8), Young Dai Kim(13), Kee Hoon Kang(38), Jib Huh(47)
	핵심어	nonparametric regression(V6), variance estimation(V11), rate of convergence(V35)
2	저자	Tae Yoon Kim(21), Tai Sup Lee (23), Chang Yi Park(133)
	핵심어	weak convergence(V19), empirical process(V20)
3	저자	Eui Yong Lee(70), Seong Joo Song(108)
	핵심어	compound poisson process(V65)



(가) 전체 연관네트워크 (나) 수정된 연관네트워크

그림 6: SCIE급 전환 이후의 논문과 핵심어 연관네트워크

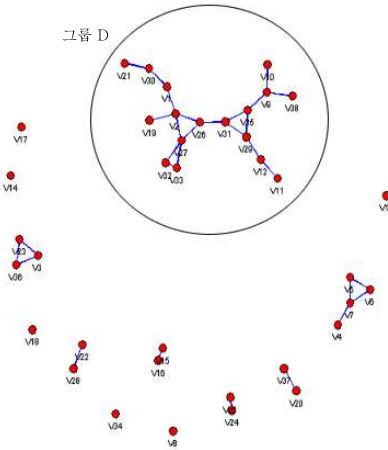


표 7: SCIE급 전환 이전의 핵심어 네트워크 중심성

핵심어	노드	근접 중심성	핵심어	노드	매개 중심성
maximum likelihood estimator	V2	0.2189	bias	V26	0.1922
bias	V26	0.2162	wavelets	V31	0.1892
confidence interval	V27	0.2122	maximum likelihood estimator	V2	0.1532
wavelets	V31	0.2104	deconvolution	V35	0.1171
deconvolution	V35	0.1991	asymptotic normality	V9	0.0871
adaptive estimation	V29	0.1946	logistic regression	V1	0.0841
			confidence interval	V27	0.0841
			adaptive estimation	V29	0.0841

표 8: SCIE급 전환 이후의 그룹에 따른 중심 핵심어 및 논문

그룹	구분	핵심어 및 논문
D	근접 중심성	maximum likelihood estimator(V2), bias(V26), confidence interval(V27)
	매개 중심성	bias(V26), wavelets(V31), maximum likelihood estimator(V2), deconvolution(V35)
	공유 논문	Asgharzadeh(2009), Cavalier와 Maimondo(2010), Huang(2010)

표 9: 그룹 D의 각 계층별 저자 및 핵심어

계층	구분	저자 및 핵심어
1	저자	Jae-Hak Lim(11), Leonid Galtchouk(47), Bert van Es(65), Laurent Cavalier(91)
	핵심어	Asymptotic normality(V9), Nonparametric regression(V12), Adaptive estimation(V29)
2	저자	Sugata Sen Roy(1), Dong Wan Shin(27), Subir Ghosh(96)
	핵심어	Logistic regression(V1), Maximum likelihood estimator(V2) Micro- array(V19)
3	저자	A. Asgharzadeh (40), Guo-Liang Fan(57), Zhensheng Huang(103)
	핵심어	Confidence interval(V27), Empirical likelihood(V32), Confidence region(V33)

이를 살펴보면 SCIE급 전환 이전과 동일하게 방대한 양의 핵심어를 사용함으로 인하여 구체적인 정보를 확인하기 힘들기 때문에 494개의 핵심어 중에서 최소 2편 이상의 논문에 제시된 핵심어를 선별하였고, 이에 총 60편의 논문, 107명의 저자, 38개의 핵심어들로 수정된 자료에 대한 분석을 실시하였다. 그림 6의 (나)는 수정된 60×38의 연관네트워크 그림이다.

수정된 연관네트워크로부터 핵심어간의 특징을 살펴보기 위해 행렬교적을 이용한 38×38 인접행렬의 핵심어 네트워크는 그림 7과 같다. 식 (2.1)에 의해서 구해진 근접 중심성과 식 (2.2)에 의해서 구해진 매개 중심성의 값을 기준으로 상위 20%의 노드에 대한 정보는 표 7과 같다. 가장 높은 근접 중심성을 보이는 maximum likelihood estimator(V2)를 중심으로 bias(V26), confidence interval(V27), wavelets(V31) 등이 하나의 핵심어 네트워크를 구성하고, Asgharzadeh (2009), Cavalier와 Maimondo(2010), Huang(2010) 등에서 핵심어를 공유하고 있다. 네트워크에서 매개 중심성이 가장 높은 핵심어는 bias(V26)이고, wavelets (V31), maximum likelihood estimator(V2), deconvolution(V35)이 매개 중심성이 높게 나타났다. 이들은 네트워크에서 여러 논문들을 이어주는 영향력 높은 핵심어가 된다.

핵심어 네트워크에서 형성된 하나의 네트워크의 특징을 자세히 살펴보기 위하여, 공저자와 핵심어에 대한 대응분석을 이용하여 보았다. 그룹 D의 핵심어와 공저자 네트워크 그림 및 대응분석도는 그림 8의 (가)와 (나)와 같고, 그림 (가)에서 매개 중심성이 높았던 Bias(V26)와 Wavelets(V31)가 두 부분을 이어주는 것을 확인할 수 있다. 그림 (나)를 살펴보면 근사적합도는 20.54%로 나타났고, 계층1(제2·3사분면), 계층2(제4사분면), 계층3(제1사분면)으로 나누어지는 것을 확인할 수 있다. 각 계층의 저자 및 핵심어는 표 9와 같이 나타난다.

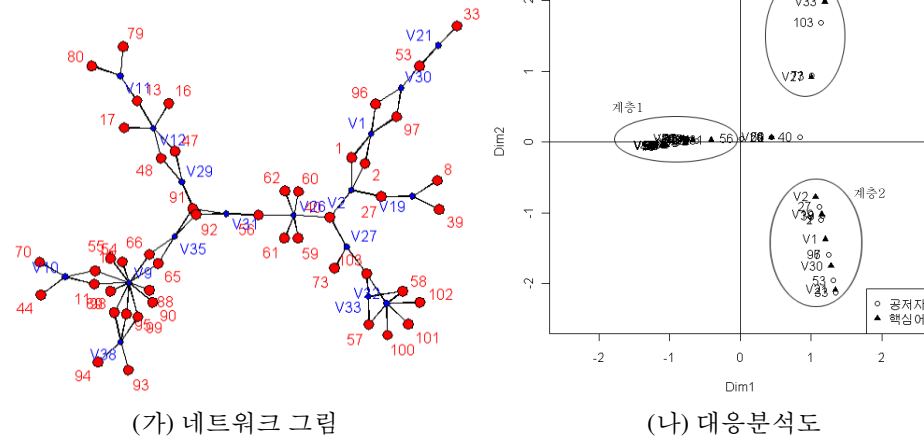


그림 8: 그룹 D의 핵심어와 공저자 네트워크 그림 및 대응분석도

4. 결론

본 연구는 사회네트워크 분석을 활용하여 JKSS에 게재된 논문으로부터 핵심어들의 네트워크를 살펴보고, 더불어 중심성이 높은 핵심어들이 네트워크에서 어떠한 특징을 갖고 있는지 알아보기 위해 대응분석을 접목시켜 보았다. 분석 결과, SCIE급 전환 이전의 논문과 핵심어 네트워크는 세 그룹의 큰 네트워크와 독립적이거나, 작은 네트워크가 형성되는 특징을 나타내어 세 분야의 연구가 활발히 이루어진 것을 확인할 수 있었다. 반면에 SCIE급 전환 이후의 논문과 핵심어 네트워크는 하나의 큰 네트워크와 여러 개의 독립된 네트워크가 형성되는 특징을 나타내어 특정 한 분야의 연구가 활발히 이루어진 것을 확인할 수 있었다. 뿐만 아니라 SCIE급 전환 이전의 경우 표 3과 같이 국내저자들의 상호간의 교류에 의한 논문투고가 많았지만, SCIE급 전환 이후의 경우 공유에 의한 국내저자의 투고보다는 표 8과 같이 개인적인 연구에 의한 해외저자의 투고가 많이 증가하였음을 확인할 수 있었다.

참고 문헌

- 김재경, 최일영, 김혜경, 김남희 (2009). 사회 네트워크 분석을 이용한 충성고객과 이탈 고객의 구매 특성 비교 연구, <한국경영과학회>, **26**(1), 183-196.
- 손동원 (2002). <사회 네트워크 분석>, 경문사, 서울.
- 정민수, 정동준 (2008). 국내 예방의학 분야의 공저자-핵심어 네트워크와 군집 양상 -대한예방의학회지(1991 2006) 게재논문의 분석-, <예방의학회지>, **41**(1), 1-9.
- 최승배, 강창완, 최형준, 강병욱 (2011). 사회네트워크분석을 통한 축구경기 분석, <한국데이터정보과학회지>, **22**(6), 1053-1063.
- 최용석 (2001). <SAS 대응분석의 이해와 응용>, 자유아카데미, 서울.
- 한국통계학회 (2011). <한국통계학회 40년사>, 한림원(주), 서울.
- 허명희 (2010). <R을 활용한 사회네트워크분석 입문>, 자유아카데미, 서울.
- Asgharzadeh, A. (2009). Approximate MLE for the scaled generalized exponential distribution under progressive type- II censoring, *Journal of the Korean Statistical Society*, **38**(3), 223-229.
- Barnes, J. (1954). Class and committees in a Norwegian island parish, *Human Relations*, **7**, 39-58.

- Cavalier, L. and Raimondo, M. (2010). Multiscale density estimation with errors in variables, *Journal of the Korean Statistical Society*, **39**(4), 417–429.
- Heo, J. E. and Kim, Y. H. (2007). Developing noninformative priors for the familial data, *Journal of the Korean Statistical Society*, **36**(1), 77–91.
- Huang, Z. (2010). Empirical likelihood for single-index varying-coefficient models with right-censored data, *Journal of the Korean Statistical Society*, **39**(4), 533–544.
- Huh, J. (2004). Nonparametric discontinuity point estimation in generalized linear model, *Journal of the Korean Statistical Society*, **33**(1), 59–78.
- Kang, K. H. and Huh, J. (2006). Nonparametric estimation of the variance function with a change point, *Journal of the Korean Statistical Society*, **35**(1), 1–23.
- Kim, D. H., Kang, S. K. and Lee, W. D. (2003). Bayesian test for the equality of the means and variances of the two normal populations with variances related to the means using noninformative priors, *Journal of the Korean Statistical Society*, **32**(3), 271–288.
- Kim, D. H., Lee, W. D. and Kang, S. G. (2006). Objective bayesian approach to step stress accelerated life tests, *Journal of the Korean Statistical Society*, **35**(3), 225–238.
- Yun, S. H. (2005). Estimation of the second order parameter characterizing the tail behavior of probability distributions: Consistency, *Journal of the Korean Statistical Society*, **34**(4), 273–280.
- Yun, S. H. (2006). Estimation of the second order parameter characterizing the tail behavior of probability distributions: Asymptotic normality, *Journal of the Korean Statistical Society*, **35**(1), 91–103.

Social Network Analysis and It's Applications for Co-authors and Key-words Networks in the JKSS

Jong Goen Kim^a, Soon-Kuek Choi^b, Yong-Seok Choi^{1, b}

^aSchool of Computer and Information, Busan Institute of Science and Technology

^bDepartment of Statistics, Pusan National University

Abstract

Social network analysis is a graphical technique to search relationships and characteristics between nodes (people, companies and organs, etc) and an important node positioning in the visualized social network figure. But it is difficult to characterize nodes in this figure. Therefore, their relationships and characteristics could be presented by applying correspondence analysis to affiliation matrix which is a kind of similarity matrix between nodes. In this study, we will provide the relationships and characteristics around co-authors and key-words of JKSS in Korean Statistical Society by using social network analysis and correspondence analysis.

Keywords: Social network analysis, centrality, correspondence analysis, Journal of the Korean Statistical Society .

¹ Corresponding Author: Professor, Department of Statistics, Pusan National University, Busan 609-735, Korea. E-mail: yschoi@pusan.ac.kr