

소셜 네트워크 분석을 활용한 통계학회 논문집과 응용통계연구 공저자 네트워크 비교[†]

전희주¹

¹동덕여자대학교 정보통계학과

접수 2015년 1월 2일, 수정 2015년 1월 23일, 게재확정 2015년 2월 10일

요약

본 연구의 목적은 한국통계학회가 출판하는 2개 학술지 한국통계학논문집과 응용통계연구를 가지고 소셜 네트워크 분석을 통해 개별 연구자들의 공저자 네트워크 영향력 분석뿐만 아니라 두 학술지가 가지고 있는 공저자 네트워크 형태와 특성을 조사하여 비교하는 데 있다. 그 결과, 공저자 네트워크의 형태를 나타내는 밀도, 포괄성, 상호연결성, 군집계수와 추이성은 거의 동일한 값을 보였으며, 응용통계연구가 한국통계학회논문집보다 노드의 수가 많은 이유로 평균연결정도, 평균거리, 직경은 더 높게 나타났다. 결국 한국통계학회논문집과 응용통계연구지 공저자 네트워크의 형태는 매우 유사한 모습을 보였다. 이는 두 논문집의 이용자가 유사하거나 동일하기 때문인 것으로 추정된다. 두 학술지 공저자 네트워크의 중심성 변수들에 대한 비교는 통계적 유의수준 0.05에서 응용통계연구 공저자 네트워크가 한국통계학회논문집보다 근접중심성과 매개중심성 측면에서 높은 것으로 나타났다. 응용통계연구 공저자 네트워크가 한국통계학회논문집 공저자 네트워크보다 근접중심성이 더 높아 공저자들 간에 서로 정보가 더 빠르게 전달되고, 매개중심성 또한 더 높게 나타나 응용통계연구 공저자들이 한국통계학회논문집 공저자들보다 매개성이 더 높은 결과를 보였다.

주요용어: 공저자, 공저자 네트워크, 소셜 네트워크, 소셜 네트워크 분석.

1. 서론

인간은 비슷한 경향을 가진 부류끼리 군집을 형성하여 서로 심리적인 안정을 가져오며 공동의 노력으로 원하는 성공적인 결과를 가져온다. 최근 이러한 비슷한 사람들끼리의 연결은 페이스북, 트위터, 카카오톡과 밴드 등의 소셜 네트워크 서비스 (social network service; SNS)의 확산과 스마트폰의 급속한 보급으로 사람들 간의 사회적 관계성과 소셜 네트워크의 역할은 기업에서 마케팅 측면뿐만 아니라 학자들 간의 학문연구에서도 갈수록 중요해지고 있다. 이러한 공동의 협력은 집단창의성 (swarm creativity)을 통해 혁신을 가져올 수 있을 뿐만 아니라 성공의 가능성도 더욱 커지게 하는 결과를 가져온다.

소셜 네트워크 분석 (social network analysis; SNA)은 개인들 간의 관계를 형성하는 한 쌍 사이에 존재하는 정보인 관계 데이터를 통해 노드 (예: 사람)들로 구성된 소셜 네트워크의 구조 형태와 각 사람들의 소셜 네트워크에서의 역할을 분석하는 것을 목적으로 한다. 결국 소셜 네트워크 분석을 통해 개인들로 이루어진 전체 네트워크뿐만 아니라 내부의 작은 네트워크들의 형태 및 구조를 파악할 수 있다. 또한 네트워크에서 누가 친구가 가장 많은지? 누가 전체에 영향력을 줄 수 있는 핵심 인물인지? 누가

[†] 이 논문은 2012년도 동덕여자대학교 신입교원정착연구비의 지원을 받아 연구되었음.

¹ (136-714) 서울시 성북구 화랑로 13길, 동덕여자대학교 정보통계학과, 조교수.
Email: hjchun@dongduk.ac.kr

네트워크와 네트워크를 연결할 수 있는지? 등을 소셜 네트워크 분석을 통해 알 수 있다. 소셜 네트워크 분석을 활용한 주제로 통신회사 통화 네트워크 (Chun, 2011), 보험회사 직원들의 네트워크 (Chun과 Leem, 2014), 축구경기 중 선수들의 공을 주고받는 네트워크 (Choi 등, 2011), 학생들의 복수정보 수강 네트워크 (Cho, 2012) 등의 다양한 네트워크 분석이 있다.

저자들 간의 학술 공동연구 또한 이러한 각 공저자를 하나의 네트워크로 표현하는 공저자 네트워크로 볼 수 있으며 소셜 네트워크 분석을 통해 공저자들의 네트워크의 특성과 구조뿐만 아니라 각 공저자들의 역할을 분석할 수 있다. 최근 학술 연구는 이공계뿐만 아니라 인문사회과학 분야 모든 분야에서 연구자들 간의 공동연구와 같은 협력은 갈수록 증가하고 있으며, 특히 이러한 공저현상은 2000년대 들어 급격히 증가하는 추세이다 (Velden 등, 2010). 학술연구에서 이러한 협력의 증가는 학문분화, 기회시간비용, 연구의 질, 위험분산 등에 의한 요인들이 작용하는 것으로 연구되었다 (Choi와 Lee, 2009). 공동연구로 대표되는 연구자들의 협력 현상은 각 연구자들을 하나의 노드로 보는 소셜 네트워크로 표현될 수 있으며 이러한 소셜 네트워크는 공저자 네트워크, 공동 네트워크 또는 협력 네트워크로 불린다. 공저자 네트워크는 소셜 네트워크 분석을 통하여 개별 연구자의 사회적/학술적 관계와 학술적 영향력 정도와 공저자 네트워크의 형태와 특성들을 파악될 수 있다. 학술지 논문에서 나타나는 공저 네트워크에 대한 연구는 Kretschmer (1994)가 초기 사례에 해당되며 이후 다양한 협력 네트워크 관련 연구들은 주로 개별 연구자들 간의 사회적 관계와 학술적인 영향력 분석이 주로 다루어져 왔다. 지금까지의 공저자 네트워크에 대한 연구는 공저자 네트워크에 나타난 일반적인 기초통계, 네트워크의 구조적 속성, 학술적 영향력 지표로 중심 영향력지표인 연결정도중심성, 근접중심성, 매개중심성, 위세영향력을 나타내는 고유 벡터중심성, 신뢰 영향력을 나타내는 페이지랭크 등을 제시하고 그 변수들 간의 상관관계를 제시하는 데 있었다.

본 연구는 한국통계학회가 출판하는 2개 학술지 한국통계학논문집과 응용통계를 가지고 소셜 네트워크 분석을 통해 개별 연구자들의 소셜 네트워크 영향력 분석뿐만 아니라 두 학술지가 가지고 있는 공저자 네트워크 형태와 특성을 조사하여 비교하고자 한다.

2. 이론적 배경

2.1. 소셜 네트워크 분석의 의미

소셜 네트워크는 개인이나 조직이 친구관계, 거래관계, 공동 소속관계 등과 같은 사회적 관계에 의해 형성된 구조이다. 이 네트워크는 행위자 (actor)인 노드 (node)와 관계 (tie) 혹은 링크 (link)에 의해 표현된다. 본 연구에서 공저자 네트워크는 연구자가 노드가 되며 연구자들 간의 공동연구의 관계가 링크로 표현된 소셜 네트워크이다. 이 소셜 네트워크를 나타내는 인접 행렬 (adjacency matrix)은 저자와 저자 간 공저자 관계가 존재할 경우 그 행렬의 요소는 1의 값을 가지고, 존재하지 않을 경우는 0의 값을 갖는다. 이런 공저자 네트워크는 소셜 네트워크의 하나로 볼 수 있으며 관계적 속성 (relational properties)과 구조적 속성 (structural properties) 변수들을 중심으로 연구되어 왔다. 구조적 속성의 지표 변수들은 사회 연구의 다양한 추세에서 소셜 네트워크가 가지고 있는 구조적 형태와 특성을 묘사하기 위해 사용된다. 반면에 관계적 속성의 지표 변수들은 네트워크 구성원 (노드)들 간 관계의 형태나 내용 (content)에 중점을 둔다. 일반적으로 소셜 네트워크 분석은 개별 구성원, 하부 그룹 (subgroup), 전체 네트워크의 세 가지 측면에서 분석이 이루어진다.

논문 공저자 네트워크는 연구 목적에 따라 1) 복잡한 네트워크의 본원적 특징과 역동성에 대한 연구와 2) 시간에 따른 네트워크 유형의 위상 변화를 이해하기 위한 연구로 구분된다. Barabasi 등 (2002), Newman (2001a, 2001b, 2001c), Watts와 Strogatz (1998), Kretschmer (1994), Liu 등 (2005), Nascimento 등 (2003), Otte와 Rousseau (2002), Rodriguez와 Pepe (2008); Li-chun 등 (2006), Lee (2010);

Huang 등 (2008), Nam과 Seol (2007); Lee 등 (2011)가 첫 번째 연구에 해당되고 Barabasi 등 (2002)과 Borner 등 (2005)가 두 번째에 속한다고 할 수 있다. 특히 Newman은 공저자 네트워크 사례를 이용하여 일반적인 네트워크 특성을 처음으로 연구한 학자로 알려져 있다.

공저자 네트워크의 소셜 네트워크 분석은 정보의 수집, 분류, 검색 및 확산과 관련된 학제 간 연구 영역인 정보과학에서 많이 활용된다. 또 공저자 네트워크는 큰 정보 네트워크의 일반적인 위상, 본원적 특징과 역동성을 이해하기 위해서 연구되고 있다. 정보과학을 연구하는 연구자들은 공저자 네트워크 분석을 통해 새롭게 나타나는 학문분야를 찾아내고, 핵심 학술지 및 연구자들을 정하고, 투고자의 지리적 분포뿐만 아니라 데이터베이스 설계 원칙을 추론하기 위하여 개별적인 학술지의 특징에 대해 결론을 도출해 내고 있다. Li-chun 등 (2006)과 Yan과 Ding (2009)의 연구는 미시적 수준에서 공저자 네트워크를 연구하여 개별적인 저자들의 특징을 찾아내고 영향력 분석을 주로 하고 있다.

공저자 네트워크 분석은 또한 사회과학에서 사회적 현상을 이해하기 위한 목적으로 활용되며, 그 예로 Fafchamps 등 (2006)은 경제학의 공동저자 네트워크 연구로 발달 경제학 분야에서 중요한 협업을 통한 지식 생산에 대한 외부효과에 대해 논하고 있다. Menezes 등 (2009)는 컴퓨터 과학에서 글로벌 공동 저자 네트워크의 소셜 네트워크 분석을 이용하여 학문 하위분야의 형성을 설명하고 있다.

2.2. 소셜 네트워크 구조 특성 분석

네트워크 구조의 특성을 나타내는 지표로는 네트워크 크기 (network size)인 노드의 수와 링크의 수, 네트워크 밀도 (density), 연결정도 (degree), 연결거리 (distance), 평균거리, 직경 (diameter), 컴포넌트 (component)의 수, 상대적 포괄성 (inclusiveness), 절대적 포괄성, 추이성 (transitivity), 상호연결성 (reciprocity)과 군집계수 (clustering coefficient)을 들 수 있다.

네트워크 밀도는 네트워크에서 노드들 사이에 연결된 정도를 말하며, 전체 노드들이 연결된 개수를 말하며 수식으로는 $\frac{k}{n(n-1)/2}$ 로 표현된다. 여기서 n 는 노드의 수이며, k 는 링크의 수이다. 결국 밀도는 네트워크를 형성하는 관계가 얼마만큼 응집되어 있는지를 설명한다. 네트워크 밀도가 높다는 것은 노드의 연결 관계가 그만큼 많으며 서로 긴밀하게 연결되어 상호간 도움이나 교류가 많음을 내포한다. 밀도는 0과 1의 범위에서 값을 가지며, 밀도가 0이라는 것은 연결선이 하나도 없는 네트워크이고, 1은 모든 노드들이 서로 연결되어 있다는 의미이다. 밀도가 높은 네트워크일수록 그 네트워크 노드들 간에 신뢰와 협력의 정도를 더욱 높은 경향을 보인다 (Coleman, 1988).

연결정도는 네트워크 내에서 각 노드의 특징을 설명할 때 활용하는 가장 기본적인 지표로 각 노드에 직접 연결되어 있는 노드들의 개수 (또는 연결선의 개수)로 표현되며, 특정 노드의 영향력 또는 활동력을 인식하는 지표로 인식된다. 연결정도가 높다는 것은 그 노드가 전체 네트워크에서 영향력이 높다는 것을 의미한다. 연결거리는 두 노드 간에는 다양한 연결경로 중에서 가장 짧은 거리 (the shortest path)를 나타내며 특정 노드의 연결성 단계 (또는 차수)의 의미를 가진다. 연결거리가 짧다는 것은 두 노드 간에는 가깝게 연결되어 있고 연결성이 높다고 할 수 있다.

평균거리는 모든 노드 쌍의 평균 최단거리로 네트워크의 크기를 파악할 수 있는 지표이다. 평균거리는 $L = \frac{2}{n(n-1)} \sum_{i>j} d_{ij}$ 로 표현된다. 여기서 n 는 노드의 수이며, d_{ij} 는 노드 i 에서 j 에 이르는 가장 짧은 경로에 대한 링크의 수이다. 직경은 네트워크의 크기를 파악하기 위한 지표로 네트워크 내의 노드들 간의 거리 중에서 가장 긴 것이 얼마나 되는지를 나타낸다.

컴포넌트의 수는 한 소셜 네트워크 안에 나타나는 컴포넌트의 개수를 말하며, 상대적 포괄성은 컴포넌트에 포함되어 있는 노드의 비율이고 절대적 포괄성은 전체노드 수에서 고립된 노드의 수를 뺀 것을 말한다. 추이성은 관계의 추이성이 성립되는 상대적 빈도로 친구의 친구가 되고 싶은 비율이며, $(B \Rightarrow C)$ 의 패턴의 수를 $(B \Rightarrow A \Rightarrow C)$ 패턴의 수로 나눈 값으로 나타낸다. 상호연결성은 네트워크에서 두

노드가 서로 연결을 맺고 있는 확률로 한 쪽에서 정보를 주는 경우 상대 쪽에서 되갚는 경우의 비율을 의미하며 다이아드 (Dyad)와 아크 (Arc) 두 가지로 측정이 된다.

$$\text{reciprocity(Dyad)} = \frac{\text{mutual}}{\text{mutual} + \text{asymetric}}$$

$$\text{reciprocity(Arc)} = \frac{2 \times \text{mutual}}{2 \times \text{mutual} + \text{asymetric}}$$

여기서 mutual은 양방향 링크의 수이며 asymmetric은 일방향 링크의 수를 말한다. 군집계수는 한 노드의 이웃 노드들이 서로 연결을 맺고 있는 확률을 계산한 지표로 자기 자신 (Ego) 노드 이외의 다른 (Alter) 노드들이 연결되어 있는 정도를 나타낸다. 군집계수 C 는 다음과 같이 계산된다.

$$C = \frac{1}{n} \sum_{i=1}^n C_i, \quad C_i = \frac{1}{n_A(n_A - 1)} \sum_{i \geq j} d_{A_{ij}}$$

여기서 n 는 노드의 수, n_A 는 자기 자신 이외의 다른 노드의 수, $d_{A_{ij}}$ 는 자기 자신 이외의 다른 노드들끼리 연결된 링크의 수이다.

2.3. 중심성 분석

중심성은 한 행위자가 전체 네트워크에서 영향력을 나타내는 정도를 나타내는 지표로, 연결정도중심성 (degree centrality), 근접중심성 (closeness centrality), 매개중심성 (betweenness centrality)와 위세중심성 (prestige centrality) 등으로 측정될 수 있다.

연결정도중심성은 한 노드가 다른 노드들과의 관계의 정도를 나타내며 한 노드에 직접적으로 연결되어 있는 다른 노드들의 합으로 얻어 진다. 이는 그 노드가 네트워크의 중심에 위치하는 정도를 계량화한 것으로 높은 연결정도를 가지는 노드는 네트워크에서 중심적인 역할을 하고 있음을 의미한다 (Bonacich, 1987). 연결정도중심성은 다음과 같이 표현된다.

$$C_D(p_k) = \sum_{i=1}^n d(p_k, p_i),$$

여기서 $d(p_k, p_i) = \begin{cases} 1, & \text{노드 } p_k \text{와 } p_i \text{가 연결되어 있으면} \\ 0, & \text{그렇지 않으면} \end{cases}$

이고 n 은 노드의 수, $d(p_i, p_j)$ 는 노드 p_i 와 p_j 간 존재하는 경로의 수이다.

근접중심성은 연결된 노드뿐만 아니라 네트워크 내 간접적으로 연결된 모든 행위자 간의 거리를 계산하여 중심성을 측정한다. 근접중심성은 한 노드가 다른 노드에 얼마만큼 가깝게 있는가를 보는 것으로 두 노드 사이의 거리를 측정한다. 근접중심성이 가장 큰 노드는 자신이 가진 자원을 가장 빠르게 전체 네트워크에 확산시킬 수 있는 것을 의미한다. 근접중심성은 다음과 같이 두 노드를 잇는 가장 짧은 경로 거리를 모두 더한 것에 역수를 취한 값으로 계산된다.

$$C_C(p_i) = \frac{n-1}{\sum_{j=1}^n d(p_i, p_j)}$$

매개중심성은 네트워크 내에서 한 노드가 담당하는 매개자 혹은 중재자 역할의 정도로서 중심성을 측정한다. 한 노드가 네트워크의 다른 노드들에 대한 중재역할의 중요성이 클수록, 네트워크의 흐름을 제

어할 수 있는 통제력은 그 만큼 커진다. 따라서 다른 노드들을 매개 혹은 중재할 수 있는 위치에 있는 노드는 네트워크 내에서 중요한 중심위치에 있다고 볼 수 있다. 매개중심성은 다음과 같은 식에 의해 계산된다.

$$C_B(p_k) = \sum_{j=1}^n \frac{g_{ij}(p_k)}{g_{ij}}$$

여기서 g_{ij} 는 행위자 p_i 와 p_j 간 최단 경로의 수, $g_{ij}(p_k)$ 는 행위자 p_i 와 p_j 간 p_k 를 경유하는 경로의 수를 나타낸다.

위세중심성은 파워중심성 (power centrality)과 고유벡터중심성 (eigenvector centrality) 등이 있다. 위세중심성은 노드의 중심성을 이웃하고 있는 노드의 중심성으로 계산되며 개별 행위자의 중심성과 해당 노드 (ego)와 연결된 다른 노드 (alters)의 중심성 지표를 함께 고려한 것으로, 중심성이 높은 노드와 이웃할수록 그 위세중심성은 커지게 된다. 파워중심성은 $P_d(i) = indegree/(g-1)$ 로 계산된다. 한 노드의 고유벡터중심성은 그 노드와 직접 연결된 모든 노드들의 고유벡터중심성의 합에 비례하는 원칙하에 각 노드의 중심성을 계산하는 연립방정식 풀이과정인 다음과 같은 선형대수의 고유벡터 계산방식과 같다.

$$x_i = \frac{1}{\lambda} \sum_{j < k} A_{ij} x_j$$

여기서 x_i 는 노드 i 의 고유벡터중심성이고 λ 는 고유값이다.

3. 실증분석

3.1. 소셜 네트워크 데이터 구성

한국통계학회가 출판하는 2개 학술지 한국통계학논문집 (Communications for Statistical Applications, CSAM)과 응용통계연구 (the Korean Journal of Applied Statistics)의 2006년 3월 2012년 12월 까지 7년간 발행된 논문들을 대상으로 하여, 한국통계학논문집의 저자는 616명, 응용통계연구의 저자는 723명의 저자들에 의한 메인노드 (main node) 세트와 공저자 소셜 네트워크 데이터세트를 각각 구성하였다. 공저자의 관계들은 무방향성으로 Table 3.1은 분석대상인 한국통계학논문집과 응용통계연구의 표본 특성이다. 응용통계연구가 607편으로 단독 논문을 포함한 경우에 52편이 많고 단독을 포함하지 않은 경우에는 449편으로 80편이 더 많아 응용통계연구가 한국통계학회지보다 공동논문의 비율이 더 많음을 알 수가 있다.

Table 3.1 Characteristics of sample

journal	including single authors		excluding single authors	
	# of papers	authors	# of papers	authors
CSAM	555	616	369	568
Applied Statistics	607	723	449	680

3.2. 한국통계학논문집과 응용통계연구 공저자 네트워크 특성 비교

공저자 네트워크의 고립된 노드 수는 한국통계학회지논문집이 48개, 응용통계연구가 45개로 공저자들 간의 관계를 좀 더 자세히 보기 위해 이 후의 모든 분석은 고립된 노드를 제외한 상태에서 공저자 네트워크를 분석할 것이다.

Table 3.2 Comparison of CSAM and Applied Statistics coauthornetwork characteristics

network characteristics	CSAM	Applied Statistics
# of node	568	680
# of link	493	646
density	0.003	0.003
average degree	0.868	0.950
average distance	8.573	9.302
diameter	23	24
# of component	107	99
inclusiveness	1	1
reciprocity (arc)	1	1
reciprocity (dyad)	1	1
clustering coefficient	0.185	0.187
transitivity	0.095	0.098
# of isolated nodes	48	45

* # represents number

Table 3.2는 한국통계학회논문집과 응용통계연구 공저자 네트워크의 특성을 나타낸 것이다. Table 3.2을 보면, 응용통계연구의 저자 수 즉 노드가 더 많은 결과로 링크 수는 응용통계연구가 646으로 한국통계학회논문집의 493보다 153이 더 많은 결과를 보인다. 밀도 (density)는 한국통계학회논문집과 응용통계연구 공저자 네트워크의 밀도가 0.003으로 동일하게 나타났다. 이는 두 공저자 네트워크의 조밀한 정도는 같다고 해석할 수 있다. 평균연결정도 (average degree)는 응용통계연구가 0.95로 한국통계학회논문집의 0.868보다 높고, 평균거리 (average distance) 또한 응용통계연구가 9.302로 한국통계학회논문집의 8.573보다 높게 나타나고 있다. 직경도 응용통계연구가 24로 한국통계학회논문집의 23보다 약간 높은 경향을 보이는데 이것은 응용통계연구가 한국통계학회논문집보다 노드의 수가 평균연결정도, 평균거리, 직경에 영향을 미치지 않기 때문에 나타나는 현상이다. 포괄성 (inclusiveness), 상호연결성 (reciprocity)과 군집계수 (clustering coefficient)와 추이성 (transitivity)을 보면 한국통계학회논문집과 응용통계연구 공저자 네트워크의 형태가 매우 유사함을 알 수가 있다.

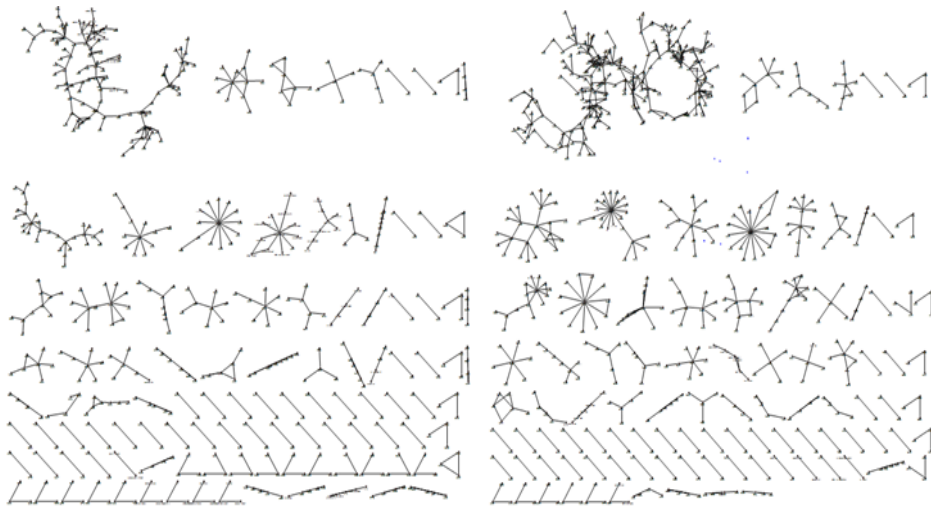
**Figure 3.1** CSAM and Applied Statistics coauthornetworks

Figure 3.1은 한국통계학회논문집과 응용통계연구 공저자 네트워크를 보여주고 있으며, 이것은 Table 3.2의 공저자 네트워크 특성을 나타내는 통계량들을 시각적으로 잘 보여주고 있다. 전체적인 네트워크 구조 측면에서 한국통계학회논문집과 응용통계연구 두 공저자 네트워크는 시각적으로 매우 유사한 구조

를 보여주고 있다. 한국통계학회논문집과 응용통계연구의 네트워크 구조와 형태가 비슷한 이유는 한국통계학회논문집과 응용통계연구는 다른 학술지와 다르게 모두 통계학회에 가입한 회원들이 주로 이용하는 학술지로 두 학술지의 이용자들이 매우 유사하거나 동일한 것으로 추정할 수 있다.

3.3. 한국통계학회논문집과 응용통계연구 공저자 네트워크의 응집성 분석 비교

Table 3.3은 한국통계학회논문집과 응용통계학회연구 공저자 네트워크에 대한 컴포넌트 분석을 한 결과이다. 한국통계학회논문집 공저자 네트워크의 컴포넌트 개수는 117, 응용통계연구 공저자 네트워크의 컴포넌트는 99개이다. 한국통계학회논문집 공저자 네트워크의 가장 큰 컴포넌트는 노드의 크기가 117이고 이는 전체 노드 수의 20.56%를 차지한 반면에 응용통계연구 공저자 네트워크의 가장 큰 컴포넌트는 노드의 크기가 216으로 이는 전체 노드 수의 31.77%를 차지하여 응용통계연구 공저자 네트워크가 한국통계학회논문집 공저자 네트워크보다 더욱 더 응집력이 높음을 알 수 있다. 이것은 응용통계연구 저자들이 한국통계학회논문집의 저자들보다 연구 네트워크가 크게 이루어지고 있음을 의미한다. Figure 3.1은 그러한 현상을 그대로 보여준다.

Table 3.3 Component comparison of CSAM and Applied Statistics coauthornetworks

	node size	# of component	% of node
CSAM	117	1	20.56%
	30	1	5.28%
	14	1	2.47%
	13	1	2.29%
	12	3	6.64%
	<12	100	62.76%
	total	107	100.00%
Applied statistics	216	1	31.77%
	24	2	7.05%
	17	1	2.50%
	16	1	2.35%
	15	1	2.21%
	13	1	1.91%
	12	1	1.77%
	<12	91	50.43%
	total	99	100.00%

3.4. 한국통계학회논문집과 응용통계연구 공저자 네트워크의 중심성 분석

Table 3.4은 한국통계학회논문집과 응용통계연구 공저자 네트워크의 연결중심성, 근접중심성, 매개중심성과 위세중심성의 한 지표인 고유벡터중심성의 값들을 구한 후 각 네트워크 내에서 영향력이 높은 상위 8번째에 속하는 공저자들을 나타낸 결과를 보여준다 (개인정보 보호를 위해 영문 성으로 표현함).

Table 3.4 High 10 coauthors 4 network centralities

rank	CSAM				Applied statistics			
	degree cent.	closeness cent.	betweenness ent.	eigenvector cent.	degree cent.	closeness cent.	betweenness cent.	eigenvector cent.
1	Hong	Heo	Park	Shim	Hong	Park	Choi	Hoe
2	Shim	Choi	Kim	Hwang	Song	Jung	Park	Lee
3	Lim	Park	Hoe	Hae	Yeo	Choi	Chun	Lee
4	Cho	Park	Soh	Seok	Shin	Kim	Lee	Park
5	Lee	Kim	Kim	Hong	Chun	Jung	Choi	Jang
6	Shin	Kim	Lee	Kim	Kim	Park	Kim	Kim
7	Kim	Soh	Lee	Kim	Kim	Lee	Jung	Yang
8	Kim	Choi	Kang	Kim	Kim	Choi	Kim	Lee

Table 3.5는 한국통계학회논문집과 응용통계연구 공저자 네트워크의 공저자들의 연결중심성, 근접중심성, 매개중심성과 위세중심성의 한 지표인 고유벡터중심성의 요약된 통계량 값이다. 공저자들의 연결중심성, 근접중심성, 매개중심성과 고유벡터중심성의 값들을 결과를 이용하여 한국통계학회논문집과 응용통계연구 공저자 네트워크의 중심성 간에 차이가 있는지를 살펴본 결과이다.

Table 3.5 Comparison of CSAM and Applied Statistics coauthornetwork centrality

centrality	CSAM		Applied statistics		df	t	p-value
	mean	s.d.	mean	s.d.			
degree	0.0031	0.0025	0.0028	0.0026	1246	-1.82	0.0693
closeness	0.0086	0.0074	0.0143	0.0104	1068	9.29	<.0001
betweenness	0.0007	0.0029	0.0013	0.0056	1062	2.61	0.0091
eigenvector	0.0044	0.0418	0.0031	0.0383	1163	-0.58	0.5590

Table 3.5를 보면, 한국통계학회논문집과 응용통계연구 공저자 네트워크의 중심성변수들에 대한 독립 2집단의 t -검정의 결과 통계적 유의수준 0.05에서 근접중심성과 매개중심성만이 두 네트워크 간에 차이가 있는 것으로 나타났다 (매개중심성과 고유벡터중심성은 한국통계학회논문집과 응용통계연구 공저자 네트워크의 등분산 가정이 위반되어 자유도가 수정됨). 근접중심성면에서 보면, 응용통계연구 공저자 네트워크가 한국통계학회논문집 공저자 네트워크보다 근접중심성이 더 높게 나타나 공저자들 간에 서로 정보가 더 빠르게 전달됨을 알 수 있다. 매개중심성 측면에서도 응용통계연구 공저자 네트워크가 한국통계학회논문집 공저자 네트워크보다 매개중심성이 더 높아 응용통계연구 공저자들이 응용통계연구 공저자들보다 공저자 네트워크 간을 서로 연결해 주는 공저자들 간의 매개성이 더 높음을 알 수 있다.

Figure 3.2는 117개의 노드로 구성된 한국통계학회논문집 공저자 네트워크의 가장 큰 컴포넌트 네트워크이며 Figure 3.3은 216개의 노드로 구성된 응용통계연구 공저자 네트워크의 가장 큰 컴포넌트 네트워크를 보여주고 있으며, 각 노드의 크기는 매개중심성의 크기를 나타내준다.

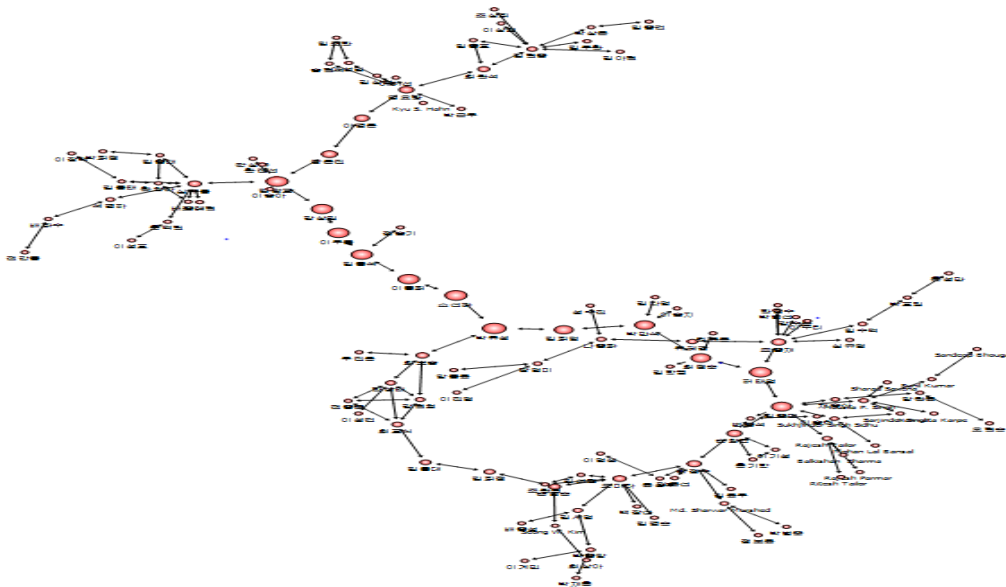


Figure 3.2 The 1st component of CSAM coauthornetwork

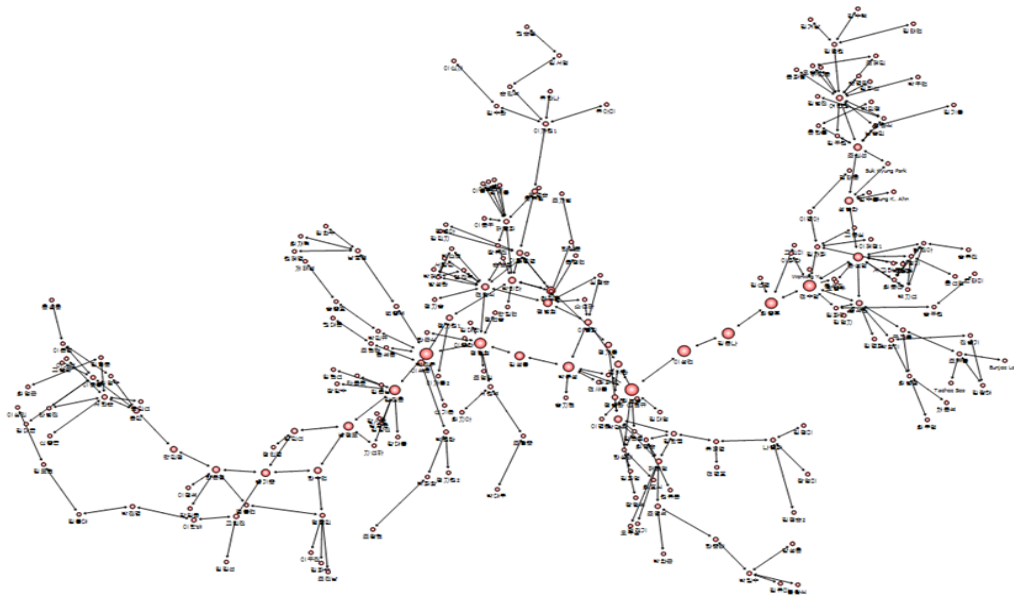


Figure 3.3 The 1st component of Applied Statistics coauthor network

4. 결론

본 연구는 한국통계학회가 출판하는 2개 학술지 한국통계학논문집과 응용통계연구를 가지고 소셜 네트워크 분석을 통해 개별 연구자들의 공저자 네트워크 영향력 분석뿐만 아니라 두 학술지가 가지고 있는 공저자 네트워크 형태와 특성을 조사하여 비교분석하였다.

한국통계학회논문집과 응용통계연구 공저자 네트워크를 비교한 결과, 공저자 네트워크의 형태를 나타내는 밀도, 포괄성, 상호연결성, 근접계수와 추이성은 거의 동일한 값을 보였으며, 응용통계연구지가 한국통계학회논문집보다 노드의 수가 많은 이유로 평균연결정도, 평균거리, 직경은 더 높게 나타났다. 결국 한국통계학회논문집과 응용통계연구지 공저자 네트워크의 형태는 매우 유사한 결과를 보였다. 이는 한국통계학회논문집과 응용통계연구의 이용자들은 다른 학술지와 다르게 모두 통계학회에 가입한 회원들로 두 학술지의 이용자들이 통계학을 기반으로 하는 유사하거나 동일한 저자들이기 때문에 네트워크의 유사한 구조를 보이는 것으로 짐작된다.

한국통계학회논문집과 응용통계연구 공저자 네트워크의 중심성변수들에 대한 비교는 통계적 유의수준 0.05에서 응용통계연구 공저자 네트워크가 한국통계학회논문집보다 근접중심성과 매개중심성 측면에서 높은 것으로 나타났다. 응용통계연구 공저자 네트워크가 한국통계학회논문집 공저자 네트워크보다 근접중심성이 더 높아 공저자들 간에 서로 정보가 더 빠르게 전달되고, 매개중심성 또한 더 높게 나타나 응용통계연구 공저자들이 한국통계학회논문집 공저자들보다 매개성이 더 높은 결과를 보였다.

본 연구는 소셜 네트워크 이론을 통계학 학술지들의 공동저자들의 네트워크에 적용해보고 향후 소셜 네트워크 이론이 다양한 영역에 응용될 수 있다는 것을 보여준다. 지금까지의 공저자 네트워크에 대한 연구는 공저자 네트워크에 나타난 일반적인 기초통계, 네트워크의 구조적 속성, 학술적 영향력 지표로 중심 영향력지표인 연결정도중심성, 근접중심성, 매개중심성과 위세중심성을 중심으로 주로 공저자들의 지위영향력을 중심으로 분석하였지만 본 연구는 한국통계학회논문집과 응용통계연구 공저자들의 지위영향력인 중심성분석 뿐만 아니라 공저자 네트워크 구조 분석과 비교를 추가하였다.

References

- Barabasi, A. L., Jeong, H., Neda Z., Ravasz, E., Schubert, A. and Viesek, T. (2002). Evolution of the social network of scientific collaborations. *PHYSICA*, **A311**, 590-614.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, **92**, 1170-82.
- Borner, K., Dall'Asta, L., Ke, W. and Vespignani, A. (2005). Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity*, **10**, 57-67.
- Choi, Y. and Lee, K. (2009). Analysis of types of journal paper coauthor: focused on Korean Public Administration Review (1989-2008). *Korean Public Administration Review*, **43**, 51-72.
- Choi, S., Kang C., Choi, H. and Kang, B. (2011). Social network analysis for a soccer game. *Journal of the Korean Data & Information Science Society*, **22**, 1053-1063.
- Cho, J. S. (2012). Inflow and outflow analysis of double majors using social network analysis. *Journal of the Korean Data & Information Science Society*, **23**, 693-701.
- Chun, H. (2011). Analysis and application to customers' social roles using voice network of A telecom company. *The Korean Journal of Applied Statistics*, **24**, 1237-1248.
- Chun, H. and Leem, B. (2014). Face/non-face channel fit comparison of life insurance company and non-life insurance company using social network analysis. *Journal of the Korean Data & Information Science Society*, **25**, 1207-1219.
- Coleman, J. (1988). Social capital in the creation of human capital. *The American Journal of Sociology*, **94**, S95-S120.
- Fafchamps, M., Van der Leij, M. and Goyal, S. (2006). Scientific networks and co-authorship. University of Oxford Department of Economics Discussion Paper Series, 256.
- Huang, M., Ahn, J. and Jahng, J. (2008). Patterns of Collaboration Networks : Co-authorship Analysis of MIS Quarterly from 1996 to 2004. *Journal of Society for e-Business Studies*, **13**, 193-207.
- Kretschmer, H. (1994). Coauthorship networks of invisible colleges and institutionalized communities. *Scientometrics*, **30**, 363-369.
- Lee, M., Park, M., Lee, H. and Jin, S. (2011). Analysis of Papers in the Korean Journal of Applied Statistics by Co-Author Networks Analysis. *The Korean Journal of Applied Statistics*, **24**, 1259-1270.
- Lee, S. (2010). A Preliminary Study on the Co-author Network Analysis of Korean Library & Information Science Research Community. *Journal of Korean Library and Information Science Society*, **41**, 297-315.
- Liu, X., Bollen, J., Nelson, M. L. and Sompel, H. V. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management*, **41**, 1462-4180.
- Li-chun, Y., Kretschmer, H., Hanneman, R. and Ze-yuan, L. (2006). Connection and stratification in research collaboration: An analysis of the COLNET network. *Information Processing and Management*, **42**, 1599-1613.
- Menezes, G. V., Ziviani, N., Laender, A. H. F. and Almeida, V. (2009). A geographic analysis of knowledge production in computer science. Paper presented at the International World Wide Web Conference Committee, Madrid, Spain.
- Nam, S. H. and Seol, S. (2007). Coauthorship analysis of innovation studies in Korea : A social network perspective. *Journal of Korea technology innovation society*, **10**, 605-628.
- Nascimento, M. A., Sander, J. and Pound, J. (2003). Analysis of SIGMOD's co-authorship graph. *SIGMOD Record*, **32**, 8-10.
- Newman, M. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E*, **64**, Art. No. 016131.
- Newman, M. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, **64**, Art. No. 016132.
- Newman, M. (2001c). The structure of scientific collaboration networks. *Proceedings of the National Academy of Science*, **98**, 404-409.
- Newman, M. (2003). Ego-centered networks and the ripple effect. *Social Networks*, **25**, 83-95.
- Otte, E. and Rousseau, R. (2002). Social network analysis: a powerful strategy also for the information sciences. *Journal of Information Science*, **28**, 444-453.
- Rodriguez, M. A. and Pepe, A. (2008). On the relationship between the structural and socioacademic communities of co-authorship network. *Journal of Infometrics*, **2**, 195-201.
- Velden, T., Haque, A. and Lagoze, C. (2010). A New Approach to Analyzing Patterns of Collaboration in Co-authorship Networks - Mesoscopic Analysis and Interpretation. arXiv:0911.4761.

- Watts, D. J. and Strogatz, S. H. (1998). Collectively dynamics of small-world networks. *Nature*, **393**, 440-442.
- Yan, E. and Ding, Y. (2009). Applying centrality measures to impact analysis: a co-authorship network analysis. *Journal of the American Society for Information Science and Technology*, **60**, 2107-2118.

The comparison of coauthor networks of two statistical journals of the Korean Statistical Society using social network analysis[†]

Heuiju Chun¹

¹Department of Statistics & Information, Dongduk Women's University

Received 2 January 2015, revised 23 January 2015, accepted 10 February 2015

Abstract

The purpose of this study is to compare not only network influence of individual coauthor but also the types and properties of two coauthor networks of Communications for Statistical Applications and Methods and the Korean Journal of Applied Statistics which are published by the Korean Statistical Society using social network analysis. As the result of two network structure comparison, density, inclusiveness, reciprocity and clustering coefficient which represent the type of coauthor networks show almost similar values and the Korean Journal of Applied Statistics has bigger values in average degree, average distance and diameter because it has more nodes than Communications for Statistical Applications and Methods. Finally two journals have very similar type of coauthor network. In the comparison of network centrality of two coauthor networks, closeness centrality and betweenness centrality of the Korean Journal of Applied Statistics are bigger than those of Communications for Statistical Applications and Methods at the statistical significance level 0.05. The coauthor network of the Korean Journal of Applied Statistics has faster information delivery and stronger betweenness than that of Communications for Statistical Applications.

Keywords: Coauthor, coauthor network, social network, social network analysis.

[†] This work was supported by new professor settlement grant funded by the Dongduk Women's University in 2012.

¹ Assistant professor, Department of Statistics & Information Science, Dongduk Women's University, Seoul 136-714, Korea. Email: hjchun@dongduk.ac.kr