

Investigating the Behavior of College Students in Class Registration using Social Network Analysis

Sung-Min Lim^a · Jinheum Kim^{b,1}

^aKorea Forestry Promotion Institute; ^bDepartment of Applied Statistics, University of Suwon

(Received May 6, 2013; Revised July 13, 2013; Accepted July 25, 2013)

Abstract

We investigate the class registration behavior of college students with data collected from the spring semester of 2012. The network originally consists of a 2-mode network between a student and class subjects. We first transform the 2-mode network into 1-mode network of students. We also subdivide it into a smaller network such as components and communities to understand the detailed behavior of students in class registration. Finally, the relationship between GPA and the centrality measures, gender, and grade of students is analyzed using a multiple regression analysis.

Keywords: 1-mode network, centrality measures, class registration, communities, components, social network analysis.

1. 서론

인터넷의 급속한 발전으로 인해 다양한 콘텐츠가 개발되고 있는데 그 중 하나가 SNS이다. SNS란 Social Networking Service의 약자로 온라인상에서 불특정 다수의 사람들과 관계를 맺을 수 있는 서비스를 말한다. SNS의 주된 특징은 첫째, 사용자 프로파일을 공개한다. SNS는 크게 페이스북과 같이 사용자들끼리 친구추가를 하고 상대방이 수락을 해주면 친구가 되는 폐쇄적 서비스와, 트위터와 같이 자신이 정보를 올리면 불특정 다수의 사람이 내가 올린 정보를 같이 공유할 수 있는 개방적 서비스로 나누어진다. 하지만 이 두 가지 서비스 모두 사용자의 ID, 나이, 성별 등과 같은 개인 프로파일의 공개를 원칙으로 한다. 둘째, 다른 사용자와 정보를 공유한다. 여기서 정보란 주로 의사소통 목적의 텍스트와 가치교환 목적의 이미지, 동영상 등을 말한다. 셋째, 커뮤니티 기능을 가진다. 이는 공통된 관심사를 가진 사용자들끼리 정보공유 및 가치교환이 이루어지기 때문이다.

1967년에 미국의 사회심리학자 J. Travers와 S. Milgram은 ‘큰 사회에서 두 사람을 연결하기 위해 필요한 중개자의 수는 매우 적다’라는 큰 사회에서의 ‘작은 세상’ 이론을 증명하고자 우편물을 전달하는 실험을 실시하였다 (Travers와 Milgram, 1969). 출발점에 있는 사람을 무작위로 선택하여 도착점에 있는 사람들의 정보를 주고, 그 사람을 알지 못할 경우 그 사람을 알 것 같은 자기가 아는 사람에게 우편물을 전달하도록 하는 실험이었다. 296개 우편물 중에서 64개가 도착했고, 평균 경로의 길이는 5.5명으로 두 사람을 연결하는 데 필요한 중개자가 6명임을 증명한 것이다. 비록 표본의 수가 적다는 연구의 취약점

¹Corresponding author: Professor, Department of Applied Statistics, University of Suwon, 17 Wauan-Gil, Bongdam-Eup, Hwaseong 445-743, Korea. E-mail: jkimdt65@gmail.com

은 있지만 큰 사회가 소수의 중개자로 연결되어 있다는 이론이 증명된 것이다. 이러한 연구는 이후로 계속 진행되어 왔는데, 특히 Huh와 Lee (2011)는 ‘작은 세상’ 이론을 가상의 네트워크에서 몬테칼로 방법을 활용해서 구현하였다.

인터넷의 보급과 스마트폰의 급속한 확산으로 우리 사회는 구성원들끼리 밀접하게 연결되고 정보와 가치를 공유하며 살아가고 있다. 이러한 복잡한 사회 네트워크를 계량적으로 분석하는 방법론이 사회 네트워크 분석(social network analysis; SNA)이다. 다시 말해 SNA는 복잡한 네트워크의 지식의 흐름 및 네트워크의 진화·확산을 살펴보는 데 용이한 분석 방법론이다. SNA는 관계를 형성하는 한 쌍 사이에 발생하는 정보인 관계 데이터를 통해 노드들로 구성된 네트워크의 구조와 각 사람들의 네트워크에서의 역할을 분석하는 것을 말한다 (Chun, 2011). 이와 같은 SNA는 현재 정치, 기업, 스포츠, 사회학, 물리학, 생물학 등 분야에서 널리 이용되고 있다.

S 대학교에서는 130학점 이상을 취득해야만 학사 학위를 받을 수 있다. 학과별로 약간 다르지만 교양, 계열기초, 전공으로 세분하여 해당 교과목을 이수해야 졸업이 가능하다. 본 논문에서는 교양선택 과목을 중심으로 S 대학교 재학생들의 수강신청 행태가 단과대학별로 어떻게 서로 다른지를 살펴보고자 한다. 여기서 수강신청 행태란 어떤 학생들이 수강신청을 공유하는지를 파악하는 것을 의미한다. 가령, 주로 학과 내 친구들과 무리 지어 수강하는지 아니면 과목의 특성에 따라 독자적으로 수강신청이 이루어지고 있는지를 파악하는 것이다. 또한 수강신청 네트워크에서 구한 네트워크 특성값과 학점과의 관계를 살펴보고자 한다. 이 연구를 위해 ㈜사이랩에서 개발한 네트워크 분석 전문 소프트웨어인 ‘Netminer4’를 사용하였다 (Lee, 2011).

2절에서는 분석 데이터를 소개하고 네트워크를 정의한다. 3절에서는 네트워크 전체의 특성을 나타내는 지표와 개별 노드들의 중심성 정도를 나타내는 지표 및 전체 네트워크의 모듈화(modularity) 정도를 나타내는 지수를 소개한다. 4절에서는 경상대학 학생들의 수강신청 행태, 5절에서는 경상대학 이외 단과대학 학생들의 수강신청 행태를 분석한다. 6절에서는 네트워크 분석을 통해 얻은 중심성 지표가 학점에 미치는 영향을 살펴본다. 7절에서는 SNA를 통해 살펴본 S 대학교 재학생들의 수강신청 행태에 대해 결론을 맺는다.

2. 분석 데이터 소개 및 네트워크 정의

본 논문에서는 S 대학교 재학생 11,198명의 2012년 1학기 수강신청 데이터셋을 분석자료로 이용하였다. 데이터셋은 학생 속성과 과목 속성에 대한 변수로 이루어져 있는데, ID, 소속 단과대학 및 학과, 성별, 학년은 전자에 해당하고, 과목코드, 과목명, 분반, 이수구분, 해당 과목의 취득학점은 후자에 해당한다. Figure 2.1은 학생과 과목으로 이루어진 2-모드 네트워크의 일부이다. 단, 원은 학생을 나타내고, 사각형은 과목을 나타낸다. 2-모드 네트워크에서는 한 학생이 수강한 과목을 알 수 있을 뿐만 아니라 어떤 과목을 수강한 학생들도 알 수 있다. 전자의 한 예로, ‘a2298’ 학생이 수강한 과목은 ‘글로벌 시대의 매너와 자기관리’, ‘법과 현대생활’, ‘영미 문학과 영화’ 임을 알 수 있다. 본 논문에서는 학생들의 수강신청 행태를 살펴보는 것이 목적이기 때문에 학생과 학생으로 이루어진 1-모드 네트워크로 변환하고자 한다. 이때 동일한 과목을 수강한 학생들은 링크로 정의하였고 링크의 가중값은 동일 과목수로 정의하였다. 한 예로 ‘a2298’ 학생과 ‘a2292’ 학생의 링크의 가중값은 2인데 Figure 2.1를 보면, 두 학생은 모두 ‘법과 현대생활’, ‘영미 문학과 영화’ 과목을 수강했음을 알 수 있다. 한편, 2-모드 네트워크를 1-모드 네트워크로 변환할 때 다음 두 가지를 가정하였다. 첫째, 학과 특성을 배제하기 위해 교양필수, 계열기초, 전공필수, 전공선택 등과 같은 과목은 제외하고 교양선택 과목만을 가지고 네트워크를 구성하였다. 둘째, 동일한 과목을 2개 이상 수강할 때 링크로 정의하였다. 이는 수강신청 행태가 서로 유사하지 않아

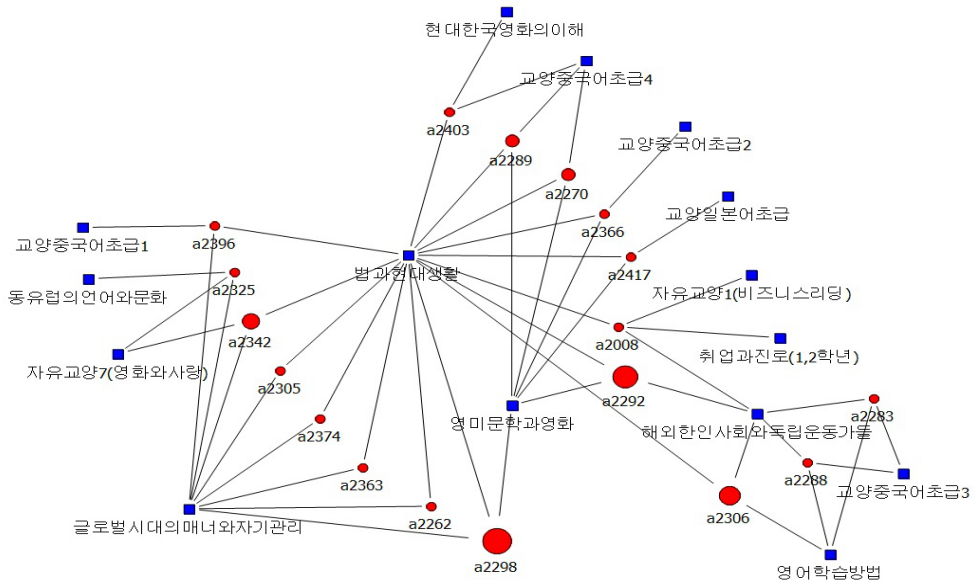


Figure 2.1. A sort of 2-mode network(Circle: student id, Rectangle: subject name)

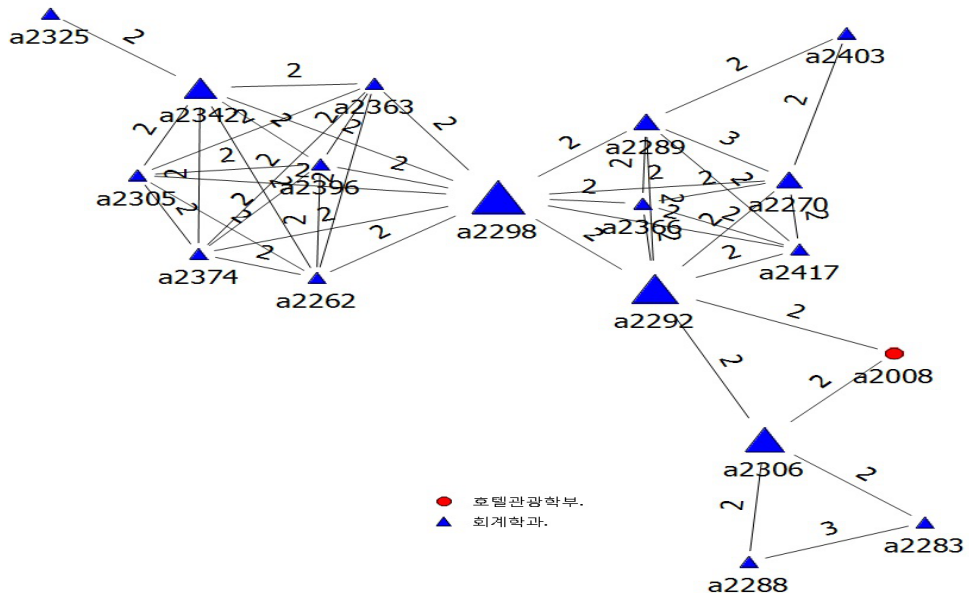


Figure 2.2. 1-mode network transformed from 2-mode network in Figure 2.1

도 우연히 한 과목은 같은 과목을 수강할 수 있다고 판단되었기 때문이다. 또한, SNA는 탐색적 자료분석을 요구하므로 가중값이 작은, 즉 한 과목만 공유하는 링크를 제외하는 것은 네트워크의 특징을 찾아내는 데 편리하기 때문이다.

3. 기호와 용어

본 절에서는 네트워크 전체의 특성을 나타내는 지표와 개별 노드들의 중심성 정도를 나타내는 지표 및 전체 네트워크의 모듈화 된 정도를 나타내는 지수를 소개하고자 한다. 이를 위해 다음과 같이 몇 가지 기호를 정의한다.

- N : 네트워크 전체 노드 수.
- L : 네트워크 전체 링크 수.
- L_{-i} : i 번째 노드에 직접적으로 링크된 이웃 노드(alterns)들의 전체 링크 수, $i = 1, \dots, N$.
- d_i : i 번째 노드의 링크 수, $i = 1, \dots, N$.
- $d(i, j)$: i 번째 노드와 j 번째 노드 간의 최단 경로(shortest path)의 길이($i \neq j$).
- g_{jk} : j 번째 노드에서 k 번째 노드로 가는 최단 경로 수($j \neq k$).
- g_{jik} : j 번째 노드에서 $k(\neq j)$ 번째 노드로 가는 최단 경로 중에 $i(\neq j)$ 번째 노드를 거쳐 가는 경로 수.

네트워크 전체의 특성을 나타내는 지표로 밀도(density), 평균 링크 수(average degree), 군집계수(clustering coefficient) 등이 널리 쓰인다. 밀도는 네트워크 전체에서 가질 수 있는 최대 링크 수에 대한 실제 네트워크가 차지하는 링크 수의 비로 정의된다. 따라서 전체 노드들이 서로 얼마나 많이 링크되어 있는지를 알 수 있는 지표이다.

$$\text{밀도} = \frac{2L}{N(N-1)}.$$

평균 링크 수는 한 노드에 직접 링크되어 있는 노드 수의 평균으로 정의된다. 따라서 한 노드가 평균 몇 개의 노드와 링크되어 있는지를 알 수 있는 지표이다.

$$\text{평균 링크 수} = \frac{1}{N} \sum_{i=1}^N d_i.$$

군집계수는 자신을 제외한 이웃 노드들로 이루어진 네트워크의 밀도의 평균으로 정의된다. 따라서 한 노드의 이웃 노드들이 서로 링크되어 있는 정도를 계산한 지표이다. 군집계수가 높은 네트워크라면 자기 자신이 네트워크에서 빠지더라도 자신의 이웃 노드들은 잘 링크되어 있다고 해석할 수 있다.

$$\text{군집계수} = \frac{1}{N} \sum_{i=1}^N C_i, \quad C_i = \frac{2L_{-i}}{d_i(d_i - 1)}.$$

한편, 개별 노드의 중심성을 나타내는 지표로 연결정도중심성(degree centrality), 근접중심성(closeness centrality), 매개중심성(betweenness centrality), 위세중심성(eigenvector centrality) 등이 있다 (Freeman, 1979; Sohn, 2002; Huh, 2010). 연결정도중심성은 한 노드에 직접적으로 연결된 이웃 노드의 수를 네트워크 노드의 크기로 표준화 한 값으로 정의되며, 연결정도중심성은 각 노드의 직접적인 영향력을 측정하는 데 유용한 지표이다. i 번째 노드의 연결정도중심성 $C_d(i)$ 는 다음과 같다.

$$C_d(i) = \frac{d_i}{N-1}, \quad i = 1, \dots, N.$$

근접중심성은 직·간접적으로 연결된 노드들의 거리를 이용하여 계산한다. 근접중심성은 연결정도중심성의 한계를 극복하여 네트워크 전체에서의 중심성 정도를 알아볼 수 있는 장점이 있다. i 번째 노드의

근접중심성 $C_c(i)$ 는 다음과 같이 정의된다.

$$C_c(i) = \frac{N-1}{\sum_{j \neq i}^N d(i,j)}, \quad i = 1, \dots, N.$$

매개중심성은 연결정도중심성 혹은 근접중심성과 달리 역할에 중점을 둔 중심성이다. 네트워크 내에서 한 노드가 하는 역할 중에서 중개자 혹은 매개자 역할의 정도로서 중심성을 구하고자 할 때 사용하는 척도이다. 매개중심성은 다른 노드 간 연결을 특정 노드가 중개하는 혹은 매개하는 빈도를 이용하여 정의되며, i 번째 노드의 매개중심성 $C_b(i)$ 는 다음과 같다.

$$C_b(i) = \sum_{j \neq i} \sum_{k \neq i, k \neq j} \frac{g_{jik}}{g_{jk}}, \quad i = 1, \dots, N.$$

위세중심성은 중요한 노드에 연결된 노드가 중요하다는 관점에서 창안되었다. 다시 말해 구조적으로 중요하지 않은 여러 노드와 연결되어 있을 때 보다 중요한 몇 개 노드에 연결되어 있을 때 해당 노드의 중심성이 높다는 개념이다. 위세중심성은 인접행렬 $A = (a_{ij})$ 의 N 개 노드에 대하여 각각 c_i ($i = 1, \dots, N$)의 중요도가 부여되었다면 i 번째 노드의 위세중심성 $C_e(i)$ 는 다음과 같다.

$$C_e(i) = \sum_{j=1}^N a_{ij}c_j, \quad i = 1, \dots, N.$$

$c = (c_1, \dots, c_N)'$ 로 표기하면 바로 위 식은 $\lambda c = Ac$ 이다. 단, c 는 행렬 A 의 가장 큰 고유값(eigenvalue) λ 에 대응하는 고유벡터(eigenvector)이다. 따라서 위세중심성을 고유벡터 중심성이라고도 부른다. 본 논문에서는 2-모드 네트워크를 1-모드 네트워크로 변환했기 때문에 중심성 지표를 해석할 때 과목과 연관 지어 해석해야 한다. 따라서 연결정도중심성이 높은 학생은 국지적 입장에서 수강인원이 많은 과목을 주로 수강하는 학생이며, 근접중심성이 높은 학생은 네트워크 전역에서 수강인원이 많은 과목을 여러 개 수강하는 학생이고, 매개중심성이 높은 학생은 자신이 소속된 학과나 학년의 학생들과 달리 다른 교양과목을 많이 수강하는 학생이고, 위세중심성이 높은 학생은 자신 주위의 친구들이 인기과목을 많이 수강하는 학생이라고 해석할 수 있다.

일반적으로 N 이 클 때 전체 네트워크를 동시에 해석하는 것은 용이하지 않다. 따라서 먼저 전체 네트워크를 링크가 끊어지지 않은 노드들의 집합인 컴포넌트(component)로 분해하거나 더 나아가 컴포넌트를 커뮤니티(community)로 세분한 후 그것을 바탕으로 전체 네트워크를 해석한다. 전체 네트워크로부터 커뮤니티를 도출한 결과는 모듈화 지수(modularity index)인 Q 값으로 평가되는데 Q 값은 다음과 같다.

$$Q = \frac{1}{2L} \sum_{i=1}^N \sum_{j=1}^N \left(a_{ij} - \frac{d_i d_j}{2L} \right) \delta(co_i, co_j),$$

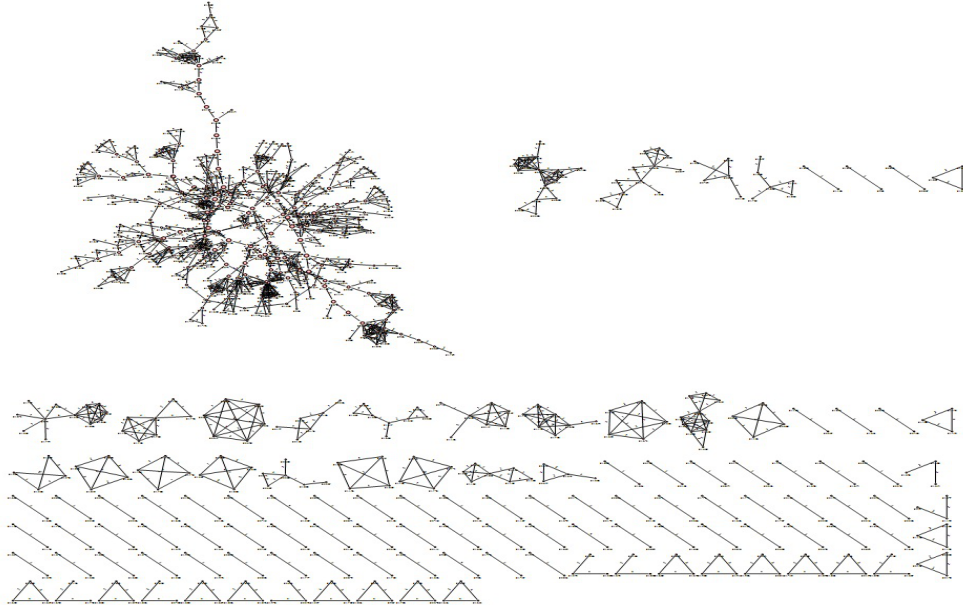
단, co_1, \dots, co_N 은 개별 노드가 속한 커뮤니티를 나타내며, $\delta(co_i, co_j)$ 는 0이나 1의 값을 갖는 지시 함수(indicator function)인데, 만일 $co_i \neq co_j$ 이면 $\delta(co_i, co_j) = 0$ 이고, $co_i = co_j$ 이면 $\delta(co_i, co_j) = 1$ 이다. Q 값이 클수록 커뮤니티들의 모듈화 된 정도가 크다고 할 수 있기 때문에 각 커뮤니티들은 독립적인 기능을 가진 노드들의 집합이라고 할 수 있다.

4. 경상대학의 수강신청 행태 분석

S 대학교 10개 단과대학 중에서 경상대학의 네트워크 특성은 자세히 살펴보고 그 외 단과대학은 단과대학별 비교에서 간략히 살펴보고자 한다.

Table 4.1. Distribution of size of components(SoC) of College of Economics and Business Administration

SoC	2	3	4	5	6	7	8	10	12	18	420	Total
Frequency	68	25	8	2	4	3	2	1	2	1	1	117

**Figure 4.1.** Diagram of 117 components of College of Economics and Business Administration

4.1. 전체 네트워크 특성

경상대학은 786개의 학생 노드와 1,245개의 링크로 이루어져 있는 네트워크이다. 경상대학 네트워크의 특성 값을 살펴보면 밀도가 0.004로 노드들 간의 링크 수가 적었으며, 평균 링크 수는 1.58개로 나타났고, 군집계수는 0.69로 주위 이웃들이 서로 많이 링크되어 있었다.

4.2. 응집구조(cohesion) 분석

응집구조 분석은 긴밀하게 연결되어 있는 노드들을 그룹으로 묶어 네트워크 구조를 쉽게 파악하는 데 목적을 둔다. 경상대학은 전체 노드의 갯수가 786개로 많아 네트워크의 해석이 어렵기 때문에 전체 네트워크를 먼저 컴포넌트로 분해하였다. Table 4.1은 경상대학 네트워크의 컴포넌트 크기에 대한 도수분포이고, Figure 4.1은 117개의 모든 컴포넌트를 그림으로 나타낸 것이다. 117개의 컴포넌트 중에서 가장 작은 컴포넌트와 가장 큰 컴포넌트는 각각 2명, 420명의 학생으로 이루어져 있다. 본 논문에서는 단과대학별로 가장 큰 컴포넌트를 찾아내고 그 컴포넌트를 바탕으로 단과대학 전체 네트워크를 해석하는 방법을 제안하고자 한다. 경상대학의 경우는 가장 큰 컴포넌트가 경상대학 전체 노드의 53.4%를 차지하였다. 그러나 가장 큰 컴포넌트 또한 420명으로 이루어져 있어 네트워크의 특성을 해석하기가 쉽지 않기 때문에 컴포넌트를 커뮤니티로 더 세분하여 전체 네트워크를 해석하고자 한다.

3절에서 언급한 것처럼 컴포넌트를 커뮤니티로 세분하면 각각의 커뮤니티는 비슷한 속성을 가진 노드들로 이루어진다. 본 논문에서는 매개중심성 방법을 써서 커뮤니티를 세분하였는데 이 방법은 매개중

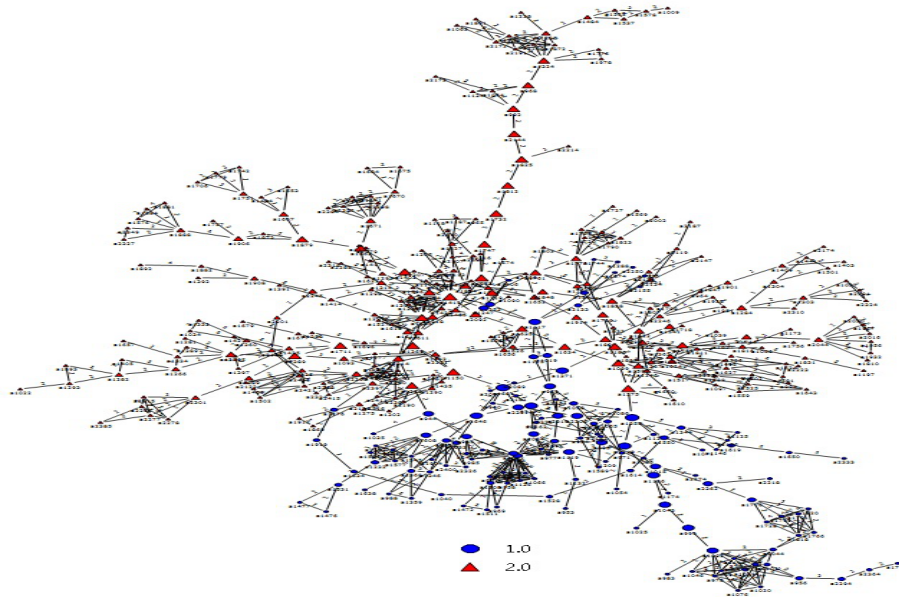


Figure 4.2. Diagram of communities corresponding to the largest component of College of Economics and Business Administration (Number of communities = 2)

심성이 큰 노드부터 차례로 제거시켜 가면서 세분하는 방법이다. Figure 4.2와 Figure 4.3은 가장 큰 컴포넌트를 각각 2개, 6개의 커뮤니티로 세분한 그림이다. 비록 가장 큰 모듈화 지수 값은 23.6이고 (Figure 4.2), 그 다음으로 큰 모듈화 지수 값은 16.7이지만 (Figure 4.3) 두 값의 차이가 크지 않고 2개 보다 6개의 커뮤니티를 해석하기가 용의하다고 판단하여 커뮤니티의 갯수를 6개로 결정하였다. 커뮤니티의 해석에 앞서 커뮤니티에 따라 중심성 지표들이 서로 다른지를 검정하기 위해 다변량 분산분석(MANOVA)을 수행하였다. Table 4.2는 커뮤니티별 노드 수와 평균 평점, 4가지 중심성 지표의 평균 값을 나타낸 것이다. 단, 괄호 안에 있는 값은 측도별 커뮤니티의 순위를 나타낸다. MANOVA의 유의 확률 값이 0.0001 미만이기 때문에 6개 커뮤니티들 간에 중심성 지표가 서로 다르다고 할 수 있다. 6개 커뮤니티별 특징은 다음과 같이 요약할 수 있다.

- 커뮤니티 1: 네트워크 전역에서 수강인원이 많은 과목을 주로 수강하는 학생들의 집합.
- 커뮤니티 2: 국지적 의미에서 수강인원이 많은 과목 위주로 수강하고, 친구들 역시 똑같은 형태를 보이는 학생들의 집합.
- 커뮤니티 3: 매개중심성이 높은 것으로 보아 자신의 무리와 다른 교양과목을 많이 수강하는 학생들의 집합.
- 커뮤니티 4: 자신의 무리와 다른 과목을 수강하는 학생들의 집합.
- 커뮤니티 5: 학점이 가장 높고 네트워크 전체에서 수강인원이 많은 과목을 여러 개 수강하는 학생들의 집합.
- 커뮤니티 6: 모든 관점에서 중심성이 낮은 학생들의 집합.

이제 커뮤니티 분석을 통해 경상대학 학생들의 수강신청 행태를 살펴보고자 한다. 특별히 다른 커뮤니티에 비해 GPA가 높고 중심성 지표 값이 큰 노드로 구성되어 있는 ‘커뮤니티 5’를 사례로 제시하고자

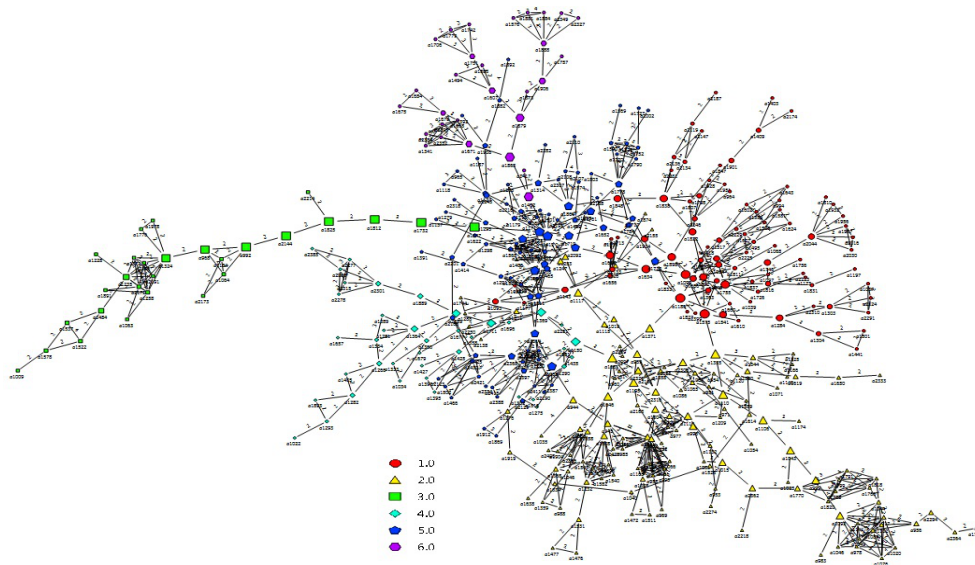


Figure 4.3. Diagram of communities corresponding to the largest component of College of Economics and Business Administration (Number of communities = 6)

Table 4.2. Number of nodes(NoN), averageGPA, and mean of four centrality indices by community

Community	NoN	GPA	Centrality indices			
			Degree	Closeness	Betweenness	Eigenvector
1	91	3.38(2)	0.0094(4)	0.0660(2)	0.00480(5)	1.31×10^{-6} (4)
2	137	2.08(6)	0.0127(1)	0.0628(4)	0.00500(4)	3.06×10^{-2} (1)
3	28	3.01(5)	0.0098(3)	0.0407(6)	0.00896(1)	7.90×10^{-9} (6)
4	40	3.05(4)	0.0081(5)	0.0632(3)	0.00565(2)	5.05×10^{-6} (3)
5	95	3.42(1)	0.0125(2)	0.0715(1)	0.00533(3)	2.72×10^{-5} (2)
6	29	3.31(3)	0.0098(3)	0.0540(5)	0.00456(6)	4.64×10^{-8} (5)

한다. Figure 4.4에서 ‘1’로 표시한 큰 원을 살펴보면, 특정 학과나 학년으로만 이루어져 있지 않고 다른 학과와 학년도 포함되어 있는 것을 알 수 있다. 이러한 행태는 경상대학의 다른 커뮤니티에서도 나타나고 있다. 노드별로 행태를 살펴보면, ‘2’로 표시된 노드는 회계학과 4학년 학생으로서 커뮤니티의 중심에 위치한 노드이다. 이 학생은 자신이 소속된 학과와 무리지어 수강하기 보다 다른 학과의 학생들과 수강하는 경향을 보였다. ‘3’으로 표시된 노드는 글로벌비즈니스학과 2학년 학생으로서 1학년과 2학년을 연결하는 역할을 하는데, 이와 유사한 역할을 수행하는 노드가 ‘4’로 표시된 노드이다. 이 노드는 회계학과 3학년 학생으로서 두 그룹으로 나뉜 회계학과 3학년 학생들을 연결하는 노드이다.

5. 그 외 단과대학별 특성

본 절에서는 공과대학(EN), 법정대학(L&SS), 인문대학(HU), 자연과학대학(NS)의 가장 큰 컴포넌트를 세분한 커뮤니티의 특징을 살펴보고자 한다. 그 외 IT대학(IT), 생활과학대학(HE), 음악대학(MU), 체육대학(PE), 미술대학(FA)들은 가장 큰 컴포넌트의 크기가 각각 12, 16, 11, 85, 26으로 작아 수강신청 행태를 쉽게 알아볼 수 있기 때문에 컴포넌트의 특징을 살펴보고자 한다. 가장 큰 컴포넌트에서 세분한

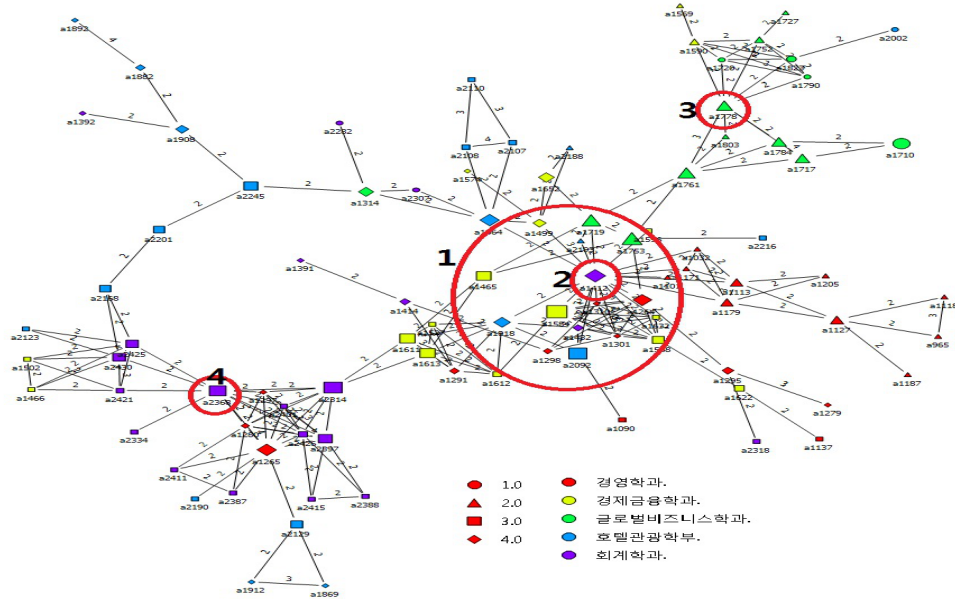


Figure 4.4. Diagram of the fifth community corresponding to the largest component of College of Economics and Business Administration (Number of communities = 6)

커뮤니티별로 중심성 지표들이 서로 다른지를 검정하기 위해 MANOVA 분석을 실시했다. Table 5.1은 4개 단과대학(EN, L&SS, HU, NS)에 대해 가장 큰 컴포넌트를 세분한 커뮤니티별로 중심성 지표의 평균값과 MANOVA분석의 유의확률 값을 나타낸 것이다. 분석결과 4개의 단과대학 모두 가장 큰 컴포넌트를 세분한 커뮤니티별로 중심성 지표가 서로 다른 것으로 나타났다.

공과대학은 가장 큰 컴포넌트가 3개의 커뮤니티로 세분되었고, 노드 수는 각각 180명, 146명, 38명이었다. 특히 첫 번째 커뮤니티는 위세중심성을 제외한 나머지 중심성 지표가 다른 커뮤니티에 비해 높은 것으로 보아 수강인원이 많은 과목을 여러 과목 수강하며, 자신의 속한 학과나 학년의 학생들과 달리 다른 과목을 많이 수강하는 학생들의 모임이라고 할 수 있다. 법정대학도 3개의 커뮤니티로 세분되었고, 노드 수는 각각 207명, 50명, 59명이었다. 특히 첫 번째 커뮤니티는 모든 중심성 지표가 다른 커뮤니티에 비해 월등히 높은 것으로 보아 주위 친구들과 함께 수강인원이 많은 과목을 여러 과목 수강하고, 주위 친구들과 공통으로 수강하는 과목외에도 다른 교양과목을 많이 수강하는 학생들의 모임이라고 할 수 있다. 인문대학은 5개 커뮤니티로 세분되었고, 노드 수는 각각 79명, 72명, 61명, 67명, 47명이었다. 특히 두 번째 커뮤니티는 매개중심성을 제외한 나머지 중심성 지표 값이 다른 커뮤니티에 비해 높은 것으로 보아 수강인원이 많은 과목을 여러 과목 수강하며, 주위 친구들이 인기과목을 많이 수강하고 있는 학생들의 모임이라고 할 수 있다. 자연과학대학은 3개의 커뮤니티로 세분되었고, 노드 수는 각각 105명, 50명, 44명이었다. 특히 첫 번째 커뮤니티는 모든 중심성 지표가 다른 커뮤니티에 비해 월등히 높은 것으로 보아 법정대학의 첫 번째 커뮤니티와 유사하다고 할 수 있다.

IT대학에서 가장 큰 컴포넌트는 정보통신학과와 컴퓨터학과가 각각 그룹을 형성하였다. 두 번째로 큰 컴포넌트도(컴포넌트 크기 = 10) 대체로 학과 위주로 이루어져 있다. 생활과학대학에서 가장 큰 컴포넌트는 식품영양학과만으로 이루어져 있고, 두 번째로 큰 컴포넌트는(컴포넌트 크기 = 15) 주로 의류학과로 이루어져 있다. 음악대학은 가장 큰 컴포넌트와 두 번째로 큰 컴포넌트가(컴포넌트 크기 = 5) 모

Table 5.1. Number of nodes(NoN) and mean of four centrality indices by college and community

College	Community	NoN	Centrality indices				p-value
			Degree	Closeness	Betweenness	Eigenvector	
EN	1	180	1.36	9.40	0.40	2.37	<0.0001
	2	146	1.33	8.03	0.20	4.51	
	3	38	0.23	1.48	0.11	0.01	
L&SS	1	207	5.17	25.61	1.63	4.84	<0.0001
	2	50	0.69	5.24	0.34	0.00	
	3	59	1.02	5.96	0.47	0.07	
HU	1	79	0.85	2.81	0.58	0.006	<0.0001
	2	72	1.17	3.27	0.53	4.147	
	3	61	0.67	2.38	0.35	0.008	
	4	67	0.75	2.83	0.36	0.001	
	5	47	0.50	2.01	0.29	0.002	
NS	1	105	3.06	5.87	0.62	0.005	<0.0001
	2	50	1.03	2.74	0.31	0.000	
	3	44	2.18	2.83	0.19	0.005	

Table 5.2. The total number of links(tNoL), the number of links between students within the same department(NoLb), and the ratio(= NoLb/tNoL) in the largest component by college

Department	IT	E&BA	EN	FA	L&SS	HE	MU	HU	NS	PE
tNoL	19	816	343	86	528	25	30	657	326	1239
NoLb	17	501	123	32	359	25	30	396	264	1239
ratio	0.90	0.59	0.36	0.37	0.68	1	1	0.60	0.81	1

두 관현악과로만 이루어져 있다. 체육대학에서 가장 큰 컴포넌트는 2개 커뮤니티로 세분되는데($Q = 15.0$), 그 중에서 한 커뮤니티는 사회체육학과 학생만으로 이루어져 있고, 나머지 커뮤니티 또한 사회체육학과 1학년으로만 이루어져 있다. 미술대학의 가장 큰 컴포넌트는 26명으로 이루어져 있다. Figure 5.1에서 '1'로 표시한 부분은 미술대학의 가장 큰 컴포넌트 내에 속해 있는 하위 네트워크(sub-network)로 밀도가 1인 클릭(clique) 형태의 네트워크이며, 모든 링크의 가중값은 2로 동일한 네트워크이다. '1'과 같은 집단이 형성된 이유는 '미술과 문화의 이해', '취업과 진로' 과목을 동일하게 수강했기 때문이다. 주로 서양화학과 3학년(파랑색 정사각형 모양)으로 구성되어 있지만 여러 학과의 학생들과 같이 속해 있는 행태를 보였다. 노드별로 보면 '2'와 '3'으로 표시한 노드는 서로 구조적으로 유사하며 그룹과 그룹을 연결하는 역할을 하고 있다 (Gould와 Fernandez, 1989).

한편, 1-모드 네트워크에서 링크된 학생들을 학과 중심으로 변환하여, 즉 같은 학과 학생들끼리의 링크이면 1로 정의하고, 서로 다른 학과 학생들끼리의 링크이면 0으로 정의하여 전체 링크 중에서 같은 학과 학생들끼리의 링크의 비율을 계산한 결과는 Table 5.2와 같다. 공과대학, 미술대학, 경상대학, 인문대학, 법정대학은 비교적 단과대학 내에서 학과별 교류가 잘 이루어지고 있는 반면에 체육대학, 음악대학, 생활과학대학, IT대학, 자연과학대학은 같은 학과, 같은 학년 중심으로 수강신청이 이루어지고 있었다.

6. 중심성 지표와 학점과의 관계 모형

본 절에서는 네트워크 분석을 통해 얻은 4가지 중심성 지표가 학점에 미치는 영향을 살펴보고자 한다. 이를 위해 단과대학별로 가장 큰 컴포넌트에 대해 먼저 Box-Cox 변환을 써서 중심성 지표를 개별적으로

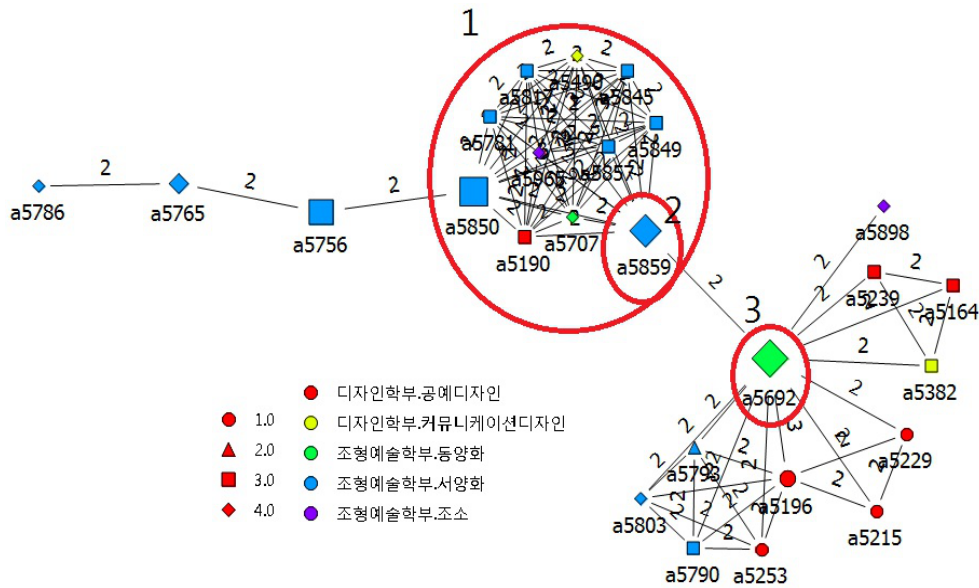


Figure 5.1. Diagram of the largest component of College of Fine Arts

Table 6.1. Number of nodes(NoN), regression parameter estimates, and adjusted $R^2 (R_a^2)$ by college

College	NoN	Intercept	Gender	Grade	Centrality indices				R_a^2
					Degree	Closeness	Betweenness	Eigenvector	
IT	12	4.71***	-	-0.21	-6.58***	-	0.006*	-	0.56
E&BA	420	2.48**	0.27**	0.24**	-0.67	25.74**	0.04	0.11	0.17
EN	364	3.12**	0.10	0.24***	0.10	-7.77	-0.18	0.52	0.08
FA	26	3.26**	-0.06	-0.05	3.16	-4.63	-3.52	0.02	0.15
L&SS	316	3.90**	0.35**	0.22**	0.11	-5.36	-0.10	-0.36	0.14
HE	16	1.28	-	-	16.9	0.07*	39.8**	0.11	0.10
MU	11	3.15	0.43	-0.11	-	-	-	-	0.03
HU	326	3.12**	0.37**	0.14**	-0.05	-8.14	-0.18	-0.10**	0.07
NS	199	0.49	0.26**	0.29**	1.99	11.71	-0.09	-0.02	0.18
PE	85	2.24**	0.58**	0.15	4.61	-3.14	-0.06	-0.03	0.04

로 정규화 하였다. 설명변수로 중심성 지표 이외에 학년과 성별을 함께 고려하였다. Table 6.1은 10개 단과대학별 회귀분석 결과이다. IT대학은 연결정도중심성, 근접중심성, 위세중심성 간에 다중공선성이 존재하여 이 지표들 중에서 통계적으로 가장 유의한 연결정도중심성만 모형에 포함하였고, 모든 개체의 성별이 동일하여 성별효과는 모형에서 제외하였다. 생활과학대학은 모든 개체의 성별과 학년이 동일하여 이 두 효과를 모형에서 제외하였다. 음악대학은 모든 개체의 네 가지 중심성 지표 값이 거의 유사하여 모두 제외하였다. 모형에서 제외된 설명변수들의 추정값은 Table 6.1에서 ‘-’으로 표시하였고, 회귀 계수 추정량의 p -값 < 0.001 이면 ***, p -값 < 0.05 이면 **, p -값 < 0.01 이면 *로 각각 표시하였다.

단과대학별 회귀분석 결과를 살펴보면, 경상대학은 성별, 학년, 근접중심성이 학점에 유의하였으며, 인문대학은 성별, 학년, 위세중심성이 학점에 유의하였다. 즉, 경상대학은 근접중심성이 높은 고학년 여학생, 인문대학은 위세중심성이 낮은 고학년 여학생의 학점이 높았다. 공과대학, 법정대학, 자연과학대학,

체육대학은 학점에 유의하게 영향을 미치는 중심성 지표는 없었다. 법정대학과 자연과학대학은 성별과 학년만 유의하였고 고학년 여학생의 학점이 높았다. 공과대학은 학년 변수만 유의하였고 고학년의 학점이 높았으며, 체육대학은 성별만 유의하였고 여학생의 학점이 높았다. IT대학은 연결정도중심성이 낮고 매개중심성이 높은 학생, 생활과학대학은 근접중심성과 매개중심성이 높은 학생의 학점이 높은 것으로 나타났다. 미술대학과 음악대학은 학점에 유의하게 영향을 미치는 변수가 없는 것으로 나타났다. 중심성 지표 간에 다중공선성이 있는 IT대학과 음악대학을 제외하고 수정된 결정계수(R_a^2)의 크기 순서로 나열하면 ‘자연과학대학 ≈ 경상대학 ≈ 미술대학 ≈ 법정대학 > 생활과학대학 ≈ 공과대학 ≈ 인문대학 ≈ 체육대학’과 같았다. 특히 IT대학의 경우는 타 단과대학과 비교하여 R_a^2 값이 매우 큰데 이는 연결정도중심성이 학점에 미치는 영향이 매우 유의하고(p -값 = 0.0036), 또한 두 변수간의 표본상관계수가 -0.678 (p -값 = 0.015)로 매우 강한 선형관계에서 기인한 것으로 생각된다.

7. 결론

본 논문에서는 S 대학교 재학생 11,198명의 2012년 1학기 수강신청 데이터를 이용하여 수강신청 행태를 살펴보기 위해 사회 네트워크 분석을 실시하였다. 이를 위해 2-모드 네트워크 데이터를 1-모드 네트워크로 변환하여 네트워크를 구성하였는데, 변환 시 학과 속성을 배제하기 위해 교양선택 과목만을 대상으로 두 과목이상 공통과목을 수강한 경우에만 두 학생의 링크가 존재한다고 가정하였다. 이렇게 구성된 네트워크로부터 단과대학별로 컴포넌트 분석, 커뮤니티 분석 및 비교, 개별 노드 해석의 순으로 분석하여 수강신청의 행태를 살펴보았으며, 또한 중심성 지표와 성별, 학년이 학점에 미치는 영향을 살펴보았다.

단과대학별 수강신청의 행태가 크게 두 가지로 구분되었는데, 공과대학, 미술대학, 경상대학, 인문대학, 법정대학은 비교적 학과별 교류가 잘 이루어지고 있는 반면 자연과학대학, IT대학, 체육대학, 생활과학대학, 음악대학은 같은 학과, 같은 학년별로 수강신청이 이루어지고 있음을 알 수 있었다. 실제로 자연과학대학 경우 가장 큰 컴포넌트에서 수강신청의 행태가 뚜렷하게 나타난 커뮤니티를 보면 학과별 교류가 있기는 하지만 대체로 학과별로 수강신청이 이루어지고 있었으며, IT대학의 경우 가장 큰 컴포넌트는 정보통신학과와 컴퓨터학과가 각각 그룹을 형성하였고, 체육대학의 경우는 가장 큰 컴포넌트가 사회체육학과 학생만으로, 생활과학대학 경우는 식품영양학과 학생만으로, 음악대학은 관현악과 학생만으로 이루어져 있었다. 한편, 대체적으로 남학생 보다 여학생이, 저학년 보다 고학년의 학점이 높았으며, IT대학 학생 중에서는 수강인원이 많지 않은 과목을 수강할수록, 경상대학 학생 중에서는 수강인원이 많은 과목을 여러 개 수강할수록, 생활과학대학 학생 중에서는 자신이 소속된 학과나 학년의 학생들과 달리 다른 교양과목을 많이 수강할수록, 인문대학 학생 중에서는 자신 주위의 친구들이 인기과목을 많이 수강할수록 높은 학점을 받는 것으로 나타났다.

References

- 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.
- Freeman, L. C. (1979). Centrality in social networks conceptual clarification, *Social Networks*, **79**, 215-239.
- Gould, J. and Fernandez, J. (1989). Structures of mediation: a formal approach to brokerage in transaction networks, *Sociological Methodology*, **19**, 89-126.
- Huh, M.-H. (2010). *Introduction to Social Network Analysis using R*, Freedom Academy, Seoul.
- Huh, M.-H. and Lee, Y. (2011). Monte-Carlo methods for social network analysis, *The Korean Journal of Applied Statistics*, **24**, 401-409.

Lee, Y. H. (2011). *Social Network Analysis*, Pakyoungsa, Seoul.

Sohn, D. W. (2002). *Social Network Analysis*, Kyungmoonsa, Seoul.

Travers, J. and Milgram, S. (1969). An experimental study of the small world problem, *Sociometry*, **32**, 425–443.

사회 네트워크 분석으로 본 수강신청 행태

임성민^a · 김진흠^{b,1}

^a한국임업진흥원, ^b수원대학교 통계정보학과

(2013년 5월 6일 접수, 2013년 7월 13일 수정, 2013년 7월 25일 채택)

요약

본 논문에서는 S 대학교 2012년 1학기 수강신청 자료를 이용하여 학생들의 수강신청 행태를 분석하였다. 학생과 과목의 링크로 이루어진 2-모드 네트워크 자료를 학생과 학생의 링크로 이루어진 1-모드 네트워크 자료로 변환하였다. 학생들의 수강신청 행태를 자세히 살펴보기 위해 전체 네트워크를 단과대학별로 컴포넌트나 커뮤니티 같은 하위 네트워크로 세분한 후 중심성 지표 값을 통해 하위 네트워크의 특징을 설명하였다. 또한 다중 회귀분석모형을 써서 성별, 학년, 중심성 지표가 학점에 미치는 영향을 살펴보았다.

주요용어: 1-모드 네트워크, 중심성, 수강신청, 커뮤니티, 컴포넌트, 사회 네트워크 분석.

¹교신저자: (445-743) 경기도 화성시 봉담읍 와우안길 17, 수원대학교 통계정보학과, 교수.
E-mail: jkimdt65@gmail.com