

# 백화점 거래 데이터를 이용한 상품 네트워크 연구

김혜경

경희대학교 경영대학 & 경영연구원  
(kimhk@khu.ac.kr)

김재경

경희대학교 경영대학 & 경영연구원  
(jaek@khu.ac.kr)

Chen Qiu Yi

경희대학교 경영대학 & 경영연구원  
(chenqiuji@khu.ac.kr)

본 연구에서는 기업에 이미 전산화되어 표준적인 형태로 존재하는 거래데이터를 이용하여 상품을 노드(node)로 놓고 동일 고객이 구매한 상품을 연결선(edge)으로 이은 상품 네트워크를 구성하였다.

사회 네트워크 분석에 널리 이용되는 중심성(Centrality)과 집중도(Centralization)를 구해서 고객의 구매패턴에서 중심이 되는 상품을 파악하였으며, 다른 상품들과 직간접적으로 연계되어 판매되는 상품관계의 총체적 흐름을 파악하고자 하였다. 또한 성별, 연령별, 판매지역, 그리고 계절별 고객의 구매활동으로부터 도출되는 상품 네트워크에서 어떤 차이가 나타나는지 네트워크의 관점에서 밝히고자 하였다. 본 연구의 결과는 상품간의 구매관계 정보를 이용하여 교차판매, 상향판매, 그리고 추가판매 등을 보다 적극적으로 유도함으로써 기업의 매출증대와 더불어 판매상품의 다양성을 확보하기 위한 전략구축 방법과 평가 방법을 제시한다.

논문접수일 : 2009년 11월 30일    게재확정일 : 2009년 12월 16일    교신저자 : 김재경

## 1. 서론

고객들의 니즈가 복잡하고 다양해지면서 기업들은 고객들의 니즈를 보다 정확하게 이해하기 위한 노력들을 하게 되었고, 이들 기업들에게는 필연적으로 경쟁력에 있어 상대적 우위 확보 수단에 대한 요구가 증가하게 되었다. 고객별로 차별화된 마케팅을 실시하는 일대일 마케팅(one to one marketing)이나 고객을 이해하고 고객과의 관계를 강화시켜 나가는 고객관계관리(CRM : customer relationship management) 등의 중요성이 부각되면서

이를 실현하기 위한 다양한 솔루션들이 개발되어 왔다(Alex et al., 2000). 대표적인 예로서, 상품검색 에이전트(agent)나 다양한 형태의 개인화(personalization) 서비스를 들 수 있다. 이러한 솔루션들은 방대한 양의 상품 정보로부터 고객이 원하는 정보만을 필터링 해주는 정보 필터링(information filtering) 기술에 기반을 둔 것으로서, 불필요한 정보를 고객 스스로가 필터링하는 노력을 줄여주는 효과를 거둘 수는 있으나 고객의 취향에 부합할만한 상품에 관한 정보를 능동적으로 제공할 수는 없다. 이에 대한 대안으로 등장한 상품추천시스템은 통

\* 본 연구는 2009년도 경희대학교 연구박사지원에 의한 결과임(KHU- 20090418). 본 논문은 2009년 한국지능정보시스템학회 추계학술대회에서 최우수논문상을 수상하였으며, 지능정보연구 편집위원회에 의해 게재확정된 논문을 알려드립니다.

계적 기법과 지식탐사기술(knowledge discovery technology)을 이용하여 고객의 취향에 가장 부합하는 상품을 추천해주는 시스템이다(Adomavicius and Tuzhilin, 2005; Sarwar, et al., 2000; Kim et al., 2009). 이와 같은 개인화 서비스는 개인의 선호와 니즈를 이해하여 관련정보를 도출할 수 있는 충분한 데이터 확보가 해당 기업에 전제되어야 한다. 그러나 다양한 채널을 통하여 판매되는 상품과 고객의 수가 많아지면서, 개별 고객을 기준으로 데이터분석 결과를 도출할 때 아주 적은 수의 선호도 데이터를 사용하게 되고, 결국 제공되는 개인화 서비스의 질을 확보하지 못하는 결과를 초래하게 된다(Agrawal and Srikant, 1994; Michael and Gordon, 2004).

본 연구에서는 이러한 문제를 해결하기 위하여 개별고객을 기준으로 적합상품을 선택하여 해당 정보를 제공하는 개인화 서비스의 차원이 아닌, 동일 데이터를 기반으로 판매되는 상품들간의 관계 정보를 도출하고 도출된 정보를 이용하여 홍보 대상이 되는 고객 또는 상품을 선택함으로써 기업에 유용한 마케팅 정보를 제공한다는 차원에서 문제를 다루고자 한다. 다시 말해, 고객들에게 유용한 정보를 제공함으로써 그들의 편의를 도모하는 한편, 기업의 교차판매(cross sell) 및 매출 증대에 초점을 맞춘 상품 네트워크 분석 방법을 제시하고자 한다.

본 연구의 상품 네트워크 분석 목적은, 첫째 고객의 구매패턴에서 중심이 되는 상품을 파악하는 것이다. 구체적으로는 다른 상품과 동시에 구매되는 빈도가 높은 상품을 비롯하여, 구매 활동이 단일 상품의 범주를 넘어 다른 상품들과 연계해서 판매되는 흐름을 파악하게 될 것이다. 둘째는 성별, 연령별, 판매지역, 그리고 시군별로 고객의 구매활동으로부터 도출되는 상품 네트워크의 실체에서 어떤 차이가 나타나는지 네트워크의 관점에

서 밝히고자 한다. 이 과정에서 분석 단위별로 시각적으로 제시되는 네트워크 다이어그램을 도출된다. 상품 네트워크의 특성은 사회 네트워크 분석(Social Network Analysis)에서 사용하는 대표적인 분석 지표인 중심성과 집중도(centralization)를 구해 분석하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 본 연구에서 제시하는 상품 네트워크에 대한 개괄적인 논의와 연구데이터를 이용한 상품 네트워크 생성방법 및 분석 방법을 설명하였다. 제 3장에서는 실제 거래 데이터를 이용하여 고객의 성별, 연령별 네트워크 및 판매점 위치, 그리고 시군별 상품 네트워크에 대한 실증적 분석 내용을 다루었다. 마지막으로, 제 4장에서는 본 연구의 결론 및 향후 연구방향을 기술하였다.

## 2. 상품 네트워크

### 2.1 상품 네트워크 생성 방법

네트워크 연구는 물리학이나 생물학, 수학과 같은 자연과학을 넘어 1990년대 말부터는 대기업의 물류관리시스템, 마케팅전략 등에 적용되고 있다. 최근에는 사회 네트워크 분석 등을 통해 빠르게 변화하는 사회현상을 설명하거나 기업의 조직 및 재무구조와 금융시장 등 그 응용의 폭이 점점 확대되고 있으며, 경제학이나 사회학 등의 다양한 학문분야의 연구자들이 사회 네트워크라는 개념들을 이용하여 광범위한 문제들을 연구하고 있다(Albert and Barabási, 2002; Barabási and Albert, 1999; Shahriar et al., 2007). 특히 네트워크 분야의 선도적인 연구들이 정립한 유용한 개념 및 측정량들이 표준화, 정량화 되고 있어 네트워크를 기반으로 복잡한 사회현상을 이해하려는 시도들은 꾸준히 증

가하는 추세이다(김용학, 2003). 또한, 컴퓨팅 기술의 발달로 대규모의 네트워크를 자동적으로 생성하고 분석하는 기술이 발전하여 기존의 사회 네트워크 연구에도 급속한 발전이 이루어지고 있다.

본 연구에서는 기업에서 CRM을 위하여 이미 전산화되어 표준적인 형태로 존재하는 거래데이터를 이용하여 네트워크를 구성하였다. 거래데이터는 각 상품을 구매한 고객의 목록이 주어져 있어, 전체 네트워크는 고객과 상품으로 노드가 생성되는 이분 네트워크(Bipartite Network)가 된다(Barabási and Albert, 1999; Zhou, 2007). 즉, 고객들의 집합과 상품집합을 고려하여 한 고객이 구매한 상품들을 연결선(edge)으로 이으면, 고객간 그리고 상품간 직접적인 연결은 없는 전형적인 이분 그래프의 형태가 된다(Newman et al., 2001). 이와 같은 이분 그래프는 네트워크 문헌에서 영화배우들과 그들이 출연한 영화, 그리고 생화학적 반응과 그 반응에 관여하는 신진대사물질 등의 형태로 찾아 볼 수 있다(Albert and Barabási, 2002; Newman, 2001).

거래 데이터로부터 얻어지는 이분 네트워크는 같은 종류의 노드로만 이루어진 두 개의 일반적인 네트워크로 투영할 수 있다. 다시 말해, 고객을 노드로, 두 고객이 동일한 상품을 구매할 경우 두 고객을 연결선으로 이은 고객 네트워크와 상품을 노드로, 두 상품을 모두 구매한 고객이 있으면 연결선으로 이은 상품 네트워크가 투영될 수 있다.

본 논문에서는 이렇게 투영된 두 네트워크 중에 상품 네트워크의 특성을 분석하고자 한다. 사람들 간의 연결관계라 할 수 있는 고객 네트워크는 사회 네트워크 연구분야에서 이미 그 중요성과 의미가 대두되어 추천시스템과 같은 개인화 시스템을 다루는 연구자들에 의하여 이미 연구가 촉발된 바 있으며, 인터넷 사용자들을 대상으로 사용자간의

연결관계에 관한 연구 또한 같은 맥락에서 이루어지고 있다(Arazy et al., 2007; Huang et al., 2007; Kim and Kim, 2009). 그러나 거래 데이터로부터 투영된 하나의 네트워크인 상품 네트워크는 아직까지 심도 있는 분석과 활용방안에 관한 논의가 부족한 상황이다. 상품 네트워크는 상품들간의 직간접적인 관계 정보를 제공해 줄 수 있는 유용한 틀로써, 교차판매 또는 상향판매 등의 마케팅 전략 구축과 매장에서의 상품 배치 등에 활용할 수 있는 정보를 도출 할 수 있다는 점에서 그 연구의 필요성과 유용성이 있다. 현재 상품관계 정보에 기반한 마케팅 활동의 예는 한 고객이 동시에 구매한 상품 정보를 이용하는 장바구니 분석(Market Basket Analysis)이 있다(Chen et al., 2004; Craig et al., 1998). 본 연구에서 제시하는 상품 네트워크는 장바구니 분석을 통해 얻어지는 상품관계 정보를 고객의 선호정보를 추적하여 시간적으로 확대한다는 의미를 갖는다. 다시 말해, 동일시점에 하나의 장바구니에 담긴 상품의 관계뿐 아니라, 일관된 선호를 지니는 단일 고객이 구매한 상품간 관계가 네트워크에 반영된다.

연구 데이터는 2007년 1월부터 12월까지 L 백화점의 서울시내 모든 지점에서 발생하는 상품판매 데이터를 이용하였다. 데이터 수집은 백화점에서 비교적 구매빈도가 높은 상위 고객등급의 거래 내역을 대상으로 하였으며, 성별, 연령별, 지역별, 시즌별 분석단위로 분류하였다. 거래 데이터에서 고객별로 동일 상품 구매빈도가 2회 이상인 기록은 제외하여 하나의 상품에 대한 한 고객의 거래기록은 단일한 것으로 간주하였다. 분석단위 데이터 집합내에서 상품 x와 상품 y가 동일 고객 A에 의해 구매되었다면, 두 상품이 연계된다. 이 연계들을 L 백화점의 남녀성별, 연령별, 지역별, 시즌별 상품 관계 네트워크를 생성하는 방법으로 설정하고 사

회 네트워크 분석을 실시하였다.

## 2.2 상품 네트워크 분석 방법

본 네트워크 분석에는 밀도분석, 중심성 분석, 집중도 분석을 사용한다(김용학, 2003; 손동원, 2002; Wouter et al., 2005). 밀도는 하나의 네트워크에서 노드간 연결된 정도로, 네트워크내 노드 사이에 존재하는 연결선이 네트워크에 포함된 모든 노드가 가질 수 있는 최대한의 연결선에서 차지하는 비중으로 계산된다. 밀도가 높은 네트워크는 정보의 교류가 활성화되어 정보의 확산이 빠른 네트워크이다. 중심성 분석은 한 네트워크에서 중심 노드가 무엇인지를 도출하는 유용한 방법이다. 연결 정도 중심성은 네트워크에서 한 노드가 다른 노드들과 얼마만큼 직접 관계를 맺고 있는가를 통하여 그 노드가 중심에 위치하는 정도를 계량화하는 지표이다. 즉, 한 노드에 직접적으로 연결되어 있는 점들의 합으로 얻어지는 값이다. 이 값은 한 점에서 직접적으로 연결되어 있는 점의 수에 한정하여 측정되기 때문에 노드간 간접 관계는 측정되지 않는다. 반면, 매개 중심성은 해당 노드가 정보를 전달하고 중개하는 위치에 있는가에 초점을 맞추는 개념이며, 근접 중심성은 정보를 얻는 원천으로서 거리가 얼마나 가까운가에 초점을 맞춘 개념이다.

본 연구에서는 다양한 중심성 개념 중에서 매개 중심성과 근접 중심성에 주목한다. 매개 중심성이란 네트워크 내에서 한 노드가 담당하는 매개자 혹은 중개자 역할의 정도로서 중심성을 측정하는 방법이다. 예를 들어 상품 x와 상품 y가 오직 상품 z를 통해서만 관계를 맺을 수 있는 경우, 상품 z는 높은 매개 중심성을 갖는다고 판단 할 수 있다. 이 상품 z는 잠재적으로 다른 상품들 사이를 통제할

수 있는 브로커의 역할을 하는 것이다. 구체적인 측정방법은 다음과 같다. 어떤 한 쌍의 노드 x와 y 사이에 존재하는 z의 매개 중심성의 측정은 네트워크 상에서 x와 y를 연결하는 가장 짧은 경로들 중에서 z가 포함되어 있는 경로의 비율로 측정한다. 즉, x와 y 사이에 존재하는 짧은 거리경로 중에서 z를 통과해야 하는 경로숫자의 비율이 중요하게 반영된다. 근접 중심성은 네트워크내 간접연결을 모두 고려한 측정 방법으로, 한 노드가 네트워크내 모든 다른 점들에 대한 최단거리의 합으로 정의된다. 따라서 네트워크의 총체적인 연결관계를 파악할 수 있는 지표이고, 그 값이 높을수록 네트워크에서 다른 모든 노드들로부터 짧은 거리에 놓인 가까운 노드라는 의미가 된다.

중심성이 한 노드에 초점을 맞춘 분석지표라면, 집중도는 한 네트워크 전체가 특정 노드에 몰리는 정도를 표현한다. 즉, 소수의 노드에 집중적으로 연결선의 흐름이 몰리게 되면 그 네트워크는 집중도가 높은 네트워크가 된다. 반면에 다양한 노드에 고르게 연결선이 분포되어 있으면 집중도가 낮은 네트워크가 된다. 따라서 밀도는 노드간 연결관계의 수에 추론되는 평균 개념인데 반해, 집중도는 노드간 연결선의 분산 개념을 지니게 된다. 예를 들어, 밀도가 높은 상품 네트워크는 소수의 상품에 집중되는 정도는 떨어지며, 밀도가 낮으면 소수 상품노드에 연결선이 집중된다는 의미를 지니게 되어 집중도가 높은 네트워크가 된다. 따라서 밀도와 집중도는 반비례하는 추세를 갖는다. 집중도도 중심성과 마찬가지로, 근접 집중도와 매개 집중도의 개념으로 분석될 수 있다.

상품 네트워크를 분석하기 위하여 앞서 기술한 바와 같이 성별, 연령별, 지역별, 시즌별로 분석단위를 설정하여 각 단위내에서 대비시켜 분석하였다. 예를 들면, 남성의 거래데이터로부터 생성된

상품 네트워크의 중심성 분석결과와 여성의 상품 네트워크의 결과와의 비교가 이루어진다.

이와 같은 분석 방법을 통하여 상품 노드 및 상품군의 구체적인 역할과 위치를 계량화하고 평가할 수 있으며, 실제로 기업에서는 그 전체 구조를 가시적으로 표현하여 쉽게 이해할 수 있다. 이렇게 도출되는 상품 네트워크는 연구 데이터를 제공한 L 백화점의 판매 상품들의 연계와 상호작용에 대한 현실적인 결과를 제공할 것으로 기대할 수 있다. 본 연구에서 사용한 분석 프로그램은 UCINET 과 Pajek을 사용하였다.

### 3. 상품 네트워크 분석

거래 데이터의 상품 네트워크 분석은 기존의 방법으로 알려지지 않은 상품간 관계를 도출하여 마케팅 담당자에게 새로운 지식을 제공하기 위한 작업으로 본 연구에서의 분석은 구매패턴에 차이가 발생하는 고객, 지점위치, 시즌 단위로 실시되었다. 거래 데이터는 고객의 성별과 연령을 기준으로 분류하였으며, 또한 거래가 발생한 시즌(계절)과 매장이 위치한 지역에 따라 분류하여 각각의 상품 네트워크를 구축하였다. 각각 구축된 네트워크는 지정된 네트워크 특성 지표를 이용하여 정량적으로 측정하여 비교 분석하였다.

#### 3.1 성별 제품 네트워크 분석

먼저 전체 거래 데이터를 성별을 기준으로 2개

<표 1> 남성과 여성의 데이터 분포

	남성	여성
고객 수	1167	2621
상품 수	1950	2352
거래 건수	14396	39177

<표 2> 상품 네트워크 분석 : 성별(남녀)

	남성	여성
밀도	0.0879	0.1181
연결집중도	0.3496	0.4467
근접중심성	0.5012	0.5213
근접집중도	-	-
매개중심성	0.000517	0.000397
매개집중도	0.00842	0.00832
고립 노드수	2	1

집합으로 분류하고, 각각의 구매데이터로부터 상품 네트워크를 생성하였다. 남성과 여성의 거래 데이터의 분포는 <표 1>과 같다. 분석 데이터의 남녀 고객의 수는 1167명과 2621명으로 여성이 약 70%의 비중을 차지하지만, 남성 1인의 평균 구매건수는 약 12이며 여성은 약 14건으로 여성의 구매빈도가 다소 높은 것으로 나타났다. 여성의 구매상품은 2,352가지로 2,621×2,352의 매트릭스를 만들게 되고, 남성의 경우에는 구매상품이 1,950종으로 1,167×1,950의 매트릭스로 표현된다.

<표 2>는 여성과 남성의 거래 데이터로부터 생성한 상품 네트워크의 분석 결과 값을 정리한 표이다. 일반적으로 희박행렬을 형성하게 되는 거래 데이터를 기반으로 생성된 상품 네트워크는 10% 내외의 밀도를 나타내었다. 또한 다른 상품과 전혀 연결선을 형성하지 않는 상품 즉, 단일 고객이 단일 상품을 선택하여 고립된 노드는 남성의 경우는 2개, 여성은 하나만 발생하였다.

거래건수가 여성 고객이 남성 고객보다 많기 때문에 여성 고객의 상품네트워크가 남성보다 높은 밀도의 네트워크가 생성될 가능성은 잠재적으로 높았지만, 여성 고객이 구매한 상품 네트워크의 연결집중도가 남성고객보다 높게 나타난 것은 흥미로운 결과이다. 일반적으로 밀도가 높은 네트워크가 집중도가 낮은 네트워크가 생성되는 경향이 있

<표 3> 중심성간 상관관계

(a) : 남성의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.855**	0.864**
근접중심성	0.855**	1	0.650**
매개중심성	0.864**	0.650**	1

(b) : 여성의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.935**	0.838**
근접중심성	0.935**	1	0.747**
매개중심성	0.838**	0.747**	1

\*\* : Significant at  $p < 0.01$  level.

음을 전술한바 있다. 본 결과는 여성고객의 상품네트워크에서 연결선의 분포가 특정 노드에 몰려있다는 것을 의미하는 것으로 여성고객이 남성고객보다 특정 제품을 집중적으로 구매하는 패턴이 있음을 의미한다.

또한, 여성의 상품네트워크가 남성의 상품 네트워크 보다 밀도가 높다는 것으로부터 특정 상품을 원천으로 다른 모든 상품까지의 연결흐름이 여성의 상품네트워크가 보다 빠르다는 것을 간접적으로 예측할 수 있다. 이를 보다 직접적으로 나타내주는 값이 근접중심성 결과값으로 모든 상품에 도달하는 연결선의 평균거리가 여성의 상품 네트워크에서 남성보다 다소 높게 나타났다. 근접중심성 분석결과는 여성과 남성의 차이가 크게 나지 않는 편이다. 이는 근접 중심성의 핵심 개념인 노드간 네트워크 거리 개념으로는 상품의 중요성을 명확하게 포착할 수 없다는 것을 의미하기도 한다. 그러나 매개중심성 분석결과 남성의 상품네트워크에서 평균 매개중심성이 여성의 상품네트워크보다 1.3배정도 높게 나타나 분명한 차이가 나타났다.

다. 그러나 여성과 남성의 상품네트워크의 매개집중도는 비슷한 수준으로 측정되어, 두 네트워크 모두 전체 네트워크의 관점에서 중요한 매개자 역할을 하는 상품의 집중 정도가 비슷한 것을 알 수 있다.

한편, 각 중심성간 상관관계를 통하여 각 중심성값이 제공하는 정보의 차이를 확인하기 위한 상관분석을 실시하였다. 결과는 <표 3>에 나타난 바와 같이 통계적으로 유의한 상관관계를 지니지만, 각 중심성간에 차이가 존재한다. 연결중심성과 매개중심성간에는 상관관계수 0.864의 높은 상관관계를 보이지만, 매개중심성과 근접중심성간에는 비교적 낮은 상관관계를 나타냈다. 즉, 상품간에 관계를 통제하는 브로커의 역할을 하는 상품은 전체 상품간의 거리계산을 통해 얻어지는 근접중심성이 높은 상품과는 차이가 있음을 알 수 있다. 특히, 남성과 여성을 비교해보면 남성의 상품 네트워크에서 그 차이가 더 크다는 것을 알 수 있다. 따라서 남성의 거래데이터로부터 도출된 상품네트워크에서 매개중심성이 높은 상품은 추가판매나 교

&lt;표 4&gt; 상품군별 중심성 비교

(a) : 남성의 상품 네트워크

	연결중심성(밀도)		매개중심성		근접중심성	
1	남성 캐주얼	0.1957	남성 캐주얼	0.0023	남성 캐주얼	0.5415
2	남성 정장	0.1693	남성 정장	0.0015	침구/수예	0.5354
3	침구/수예	0.1573	섬유잡화	0.0014	남성 정장	0.5351
4	섬유잡화	0.1560	주방용품	0.0013	섬유잡화	0.5326
5	주방용품	0.1478	침구/수예	0.0013	주방용품	0.5255
6	셔츠	0.1418	셔츠	0.0011	셔츠	0.5232
7	란제리/내의	0.1239	란제리/내의	0.0010	아동	0.5190
8	아동	0.1229	화장품	0.0009	란제리/내의	0.5158
9	화장품	0.1092	아동	0.0007	화장품	0.5144
10	시티웨어	0.1047	주방가전	0.0007	시티웨어	0.5126

(b) : 여성의 상품 네트워크

	연결중심성(밀도)		매개중심성		근접중심성	
1	섬유잡화	0.2422	섬유잡화	0.0016	섬유잡화	0.5699
2	주방용품	0.2254	주방용품	0.0014	침구/수예	0.5591
3	침구/수예	0.2117	침구/수예	0.0010	주방용품	0.5568
4	시티웨어	0.1946	남성 캐주얼	0.0009	시티웨어	0.5524
5	셔츠	0.1909	시티웨어	0.0008	남성 정장	0.5501
6	남성 정장	0.1898	남성 정장	0.0007	셔츠	0.5469
7	남성 캐주얼	0.1845	셔츠	0.0007	란제리/내의	0.5445
8	란제리/내의	0.1741	란제리/내의	0.0007	남성 캐주얼	0.5445
9	화장품	0.1492	화장품	0.0006	아동	0.5337
10	아동	0.1475	주방가전	0.0005	기타의류	0.5335

차판매 등을 위한 마케팅 활동에 기존정보와 다른 참조 정보를 제공할 수 있을 것으로 기대된다.

또한 보다 일반화된 분석결과를 얻기 위하여 상품노드가 포함된 상품군별 네트워크 특성을 알아보았다. 각 상품군별로 포함된 상품노드의 밀도와 중심성 값의 평균값을 조사하여 상위 10개 카테고리를 선별한 후, 내림차순으로 정렬하여 <표 4>에 나타내었다. 남성의 경우는 남성 캐주얼의 매개 중심성이 상대적으로 높게 나타났고, 그 뒤를 이어 남성 정장, 섬유잡화, 주방용품, 침구/수예, 셔츠, 란

제리/내의, 화장품, 아동, 주방가전 순으로 매개 중심성이 높은 것으로 나타났다. 반면, 여성은 섬유잡화와 주방용품의 매개 중심성이 높았고, 침구/수예, 남성 캐주얼, 시티웨어, 남성 정장, 셔츠, 란제리/내의, 화장품, 주방가전의 순으로 매개 중심성이 높게 나타났다. 이외에 연결중심성과 근접중심성 등을 고려하여 판단할 때도 남성의 경우는 보통 캐주얼과 정장 등 남성옷을 중심으로 쇼핑이 이루어 진다는 것을 유추할 수 있다. 반면, 여성의 상품네트워크에서 가장 높은 매개중심성이 측정

<표 5> 연령별 데이터 분포

	under 30	31 to 33	34 to 37	38 to 40	over 40
고객수	1612	857	782	361	176
상품수	1725	1706	1887	1849	1356
거래건수	11752	11156	14847	10755	5063

<표 6> 상품 네트워크 분석 : 연령별

	under 30	31 to 33	34 to 37	38 to 40	over 40
밀도	0.05	0.0704	0.0931	0.1079	0.0873
연결집중도	0.2641	0.3346	0.3813	0.406	0.3323
근접중심성	0.4585	0.489	0.5077	0.516	0.5084
근접집중도	-	-	-	0.3063	-
매개중심성	0.0006827	0.00062	0.00052	0.000506	0.000718
매개집중도	0.01318	0.01483	0.0091	0.0109	0.01265
고립 노드수	6	2	1	0	1

된 섬유잡화와 그 후순위 상품카테고리간 값의 차이가 남성과 비교하여 크지 않다는 점에서 주방용품 및 침구 등 다양한 상품들을 대상으로 쇼핑이 이루어진다고 판단된다.

### 3.2 연령별 제품 네트워크 분석

마케팅 기획에 중요한 지표가 되는 연령을 기준으로 상품 네트워크를 분석하기 위하여 연령대를 기준으로 연구 데이터를 분류하였다. 연구 데이터에서 거래량의 분포가 집중된 30대를 세분하여 분석을 실시하였다. 각 세부분류의 상품 노드의 수가 비교적 균일하게 분포하도록 30세 이하, 31세~33세, 34세~37세, 38세~40세, 40세 초과로 데이터를 다섯개의 연령집합으로 분류하였다. 각 데이터 집합의 거래건수는 각 집합에 속한 고객 구매력의 차이로 인하여 30세이하의 평균거래건수는 약7건으로 가장 낮으며, 38~40세의 구매건수는 약 30건

으로 가장 많아 차이를 나타낸다. 연령별 데이터 분포는 <표 5>과 같다.

<표 6>은 연령집합별 상품 네트워크의 분석결과를 정리한 표이다. 38세~40세 연령집합의 고객수, 상품수, 그리고 거래건수를 고려할 때, 다른 연령집합의 상품 네트워크 보다 상대적으로 높은 밀도를 나타냈다. 밀도와 더불어 근접중심성 또한 타 네트워크와 비교하여 다소 높은 것으로 나타났다. 즉, 한 고객이 다양한 상품에 대한 구매활동이 이루어지고 있으며, 이로 인하여 고립된 상품 노드가 없는 네트워크인 동시에 전체 네트워크의 관점에서 상품 노드간 연계거리가 짧은 네트워크를 생성하게 된다. 뿐만 아니라 연결집중도 또한 높은 네트워크를 생성하여 타 연령집합보다 특정 상품에 대한 선호도의 집중력이 높은 것으로 나타났다. 반면에 30세 이하의 연령에서는 밀도와 연결집중도 모두 다른 연령집합의 상품 네트워크보다 낮게 나타나 구매 상품의 범위가 넓지 않을 뿐 만 아니라,



&lt;표 7&gt; 중심성간 상관관계: 연령별

(a) : 30세 이하 고객의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.745**	0.882**
근접중심성	0.745**	1	0.528**
매개중심성	0.882**	0.528**	1

(b) : 31~33세 고객의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.837**	0.867**
근접중심성	0.837**	1	0.611**
매개중심성	0.867**	0.611**	1

(c) : 34~37세 고객의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.852**	0.872**
근접중심성	0.852**	1	0.670**
매개중심성	0.872**	0.670**	1

(d) : 38~40세 고객의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.952**	0.844**
근접중심성	0.952**	1	0.757**
매개중심성	0.844**	0.757**	1

(e) : 40세초과 고객의 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.878**	0.872**
근접중심성	0.878**	1	0.688**
매개중심성	0.872**	0.688**	1

\*\* : Significant at  $p < 0.01$  level.

특정상품에 대한 집중도 또한 높지 않다는 것을 확인할 수 있다.

한편, 40세 이상의 연령집합에서는 가장 높은 매개중심성을 갖는 상품 네트워크를 생성한다. 즉,

두 가지의 구매상품을 매개해 주는 특징을 갖는 상품노드의 발견이 비교적 쉬운 네트워크로 구매상품의 확대를 위한 마케팅 활동에 유용한 정보를 제공할 수 있다. 또한 31~33세 연령집합의 상품네트워크는 매개중심성은 비교적 높지 않지만, 매개집중도가 높은 네트워크이기 때문에 전체 상품노드 중에서 소수의 상품노드에 매개성이 집중되어 있는 네트워크로 판단 된다. 따라서 구매상품 범위 확대를 유도하는 정보 서비스 제공에 높은 효과를 기대할 수 있는 상품노드 탐색에 유용한 네트워크라 할 수 있다.

그리고 성별 상품네트워크 분석과 마찬가지로 각 중심성간 상관관계를 통하여 각 중심성값이 제공하는 정보의 차이를 확인하기 위한 상관분석을 실시하였다. 결과는 <표 7>에 나타난 바와 같이 통계적으로 높은 상관관계를 보이지만, 각 네트워크의 중심성 사이에는 분명한 차이가 존재한다. 매개중심성과 근접중심성간의 상관관계는 밀도가 높은 38세~40세 연령집합에서 가장 높게 나타난 반면, 밀도가 낮은 30세 이하 연령에서 상관관계가 비교적 낮게 나타난다. 다양한 거래데이터를 통하여 보다 심도 있는 검증이 필요하겠지만, 본 연구 데이터에서는 네트워크 밀도가 중심성 측정지표들간의 상관관계에 영향을 주는 것으로 나타났다. 즉, 밀도가 높은 상품네트워크에서 그렇지 않은 네트워크보다 매개중심성과 근접중심성간의 상관관계가 높은 경향이 있다.

연령별 분석에서도 성별 분석에서와 동일한 방법으로 상품노드가 포함된 상품군별 특성을 조사하였다. 각 상품군별 밀도와 중심성 값의 평균값을 조사한 결과는 <표 8>과 같다. 밀도가 높은 상위 10개 상품군은 각 연령대에서 다른 결과가 도출되었다. 즉, 30세 이하 고객의 상품 네트워크에서 밀

도가 가장 높은 상품군은 섬유잡화이고, 30세~37세 고객의 상품 네트워크에서는 남성 캐주얼 상품군이며, 38세에서 40세까지 고객들의 상품 네트워크에서는 주방용품이 다른 상품군에 비하여 많은 연결선을 갖는 상품군으로 나타났다. 저연령층에서는 스타킹, 스카프 그리고 넥타이와 같은 비교적 저가의 상품이 포함된 섬유잡화 상품군에서 많은 판매가 이루어지는 반면, 경제적으로 구매력이 높은 고연령층에서 비교적 고가의 상품이 다수 포함되는 상품군에 대한 구매가 활발히 이루어지고 있음을 보여주는 결과이다. 한편, 33세 이하 연령대에서는 화장품이 밀도가 높은 상위 10개 상품군에 포함되지만, 나머지 연령대에서 그렇지 않았으며, 38세 이상 연령층에서는 그 이하 연령층에서 나타나지 않은 골프 상품군이 포함되었다. 또한 40세 이상의 고객의 상품네트워크에서는 다른 연령대에서 모두 포함하는 시티웨어 상품군이 포함되지 않았으며, 아동류 상품군도 연령대별 차이가 나타났다. 이와 같이 연령층마다 상이한 상품 네트워크가 생성되는 것을 확인할 수 있으며, 연령별로 생성된 상품 네트워크의 밀도는 고객의 라이프스타일과 구매력이 명확하게 반영되는 지표임을 알 수 있다.

마찬가지로, 근접중심성지표를 통해 확인한 상품군별 특성을 살펴보면, 전연령대에서 남성 캐주얼, 섬유잡화, 셔츠, 침구/수예, 트레이셔널 상품군이 상위 10위내에 포함되어 있다. 즉, 해당상품군은 상품 네트워크에서 타 상품에 도달하는 거리가 비교적 짧은 상품이 많이 포함되는 상품군으로, 다른 상품과 동시에 구매될 확률이 높은 상품군이다. 연령대별로는 30세 이하에서는 모피/피혁 상품군이 특징적으로 나타났는데, 이는 행사매장에서 높은 판매량에 기인한 것으로 분석되었다. 이외에

&lt;표 8&gt; 상품군별 중심성 비교 : 연령별

(a) 30세 이하 고객의 상품 네트워크

	밀도		근접중심성		매개중심성	
1	섬유잡화	0.1156	섬유잡화	0.5053	섬유잡화	0.0028
2	남성 캐주얼	0.1006	셔츠	0.4918	남성 캐주얼	0.0025
3	주방용품	0.0944	침구/수예	0.4915	주방용품	0.002
4	침구/수예	0.0865	남성 캐주얼	0.4877	침구/수예	0.0017
5	시티웨어	0.0835	시티웨어	0.4853	시티웨어	0.0015
6	셔츠	0.0811	트래디셔널	0.484	트래디셔널	0.0013
7	트래디셔널	0.0756	생활잡화	0.4736	주방가전	0.0013
8	란제리/내의	0.0736	모피/피혁	0.4716	셔츠	0.0012
9	화장품	0.0663	화장품	0.4706	란제리/내의	0.0011
10	생활잡화	0.0618	커리어	0.4696	화장품	0.0011

(b) 30세~33세 고객의 상품 네트워크

	밀도		근접중심성		매개중심성	
1	남성 캐주얼	0.1524	섬유잡화	0.5275	남성 캐주얼	0.0023
2	섬유잡화	0.144	주방용품	0.5236	섬유잡화	0.0022
3	주방용품	0.1402	침구/수예	0.5201	주방용품	0.0022
4	침구/수예	0.1255	남성 캐주얼	0.5196	침구/수예	0.0017
5	셔츠	0.1156	셔츠	0.5179	트래디셔널	0.0015
6	시티웨어	0.1124	시티웨어	0.5142	셔츠	0.0012
7	트래디셔널	0.1054	트래디셔널	0.5078	시티웨어	0.0011
8	란제리/내의	0.098	화장품	0.5016	화장품	0.0011
9	화장품	0.0914	란제리/내의	0.5006	주방가전	0.0011
10	생활잡화	0.0818	생활잡화	0.4976	란제리/내의	0.001

(c) 34세~37세 고객의 상품 네트워크

	밀도		근접중심성		매개중심성	
1	남성 캐주얼	0.2044	남성 캐주얼	0.5537	주방용품	0.0018
2	섬유잡화	0.1789	섬유잡화	0.5447	남성 캐주얼	0.0017
3	주방용품	0.1706	침구/수예	0.5406	섬유잡화	0.0017
4	침구/수예	0.1631	주방용품	0.538	침구/수예	0.0013
5	셔츠	0.1574	트래디셔널	0.5366	트래디셔널	0.0011
6	트래디셔널	0.1564	셔츠	0.5342	셔츠	0.001
7	시티웨어	0.1366	란제리/내의	0.5286	주방가전	0.0008
8	란제리/내의	0.1352	시티웨어	0.5265	란제리/내의	0.0008
9	아동	0.1238	기타의류	0.5262	시티웨어	0.0008
10	기타의류	0.1237	아동	0.5237	화장품	0.0008

(d) 38세~40세 고객의 상품 네트워크

	밀도		근접중심성		매개중심성	
1	주방용품	0.1934	주방용품	0.5507	섬유잡화	0.0019
2	섬유잡화	0.1892	섬유잡화	0.5494	주방용품	0.0014
3	남성 캐주얼	0.18	남성 캐주얼	0.5458	남성 캐주얼	0.0011
4	셔츠	0.1711	셔츠	0.5443	침구/수예	0.001
5	트래디셔널	0.1671	트래디셔널	0.543	란제리/내의	0.001
6	침구/수예	0.165	침구/수예	0.5423	트래디셔널	0.001
7	란제리/내의	0.1502	란제리/내의	0.5364	셔츠	0.0009
8	아동	0.1448	아동	0.5347	골프	0.0008
9	골프	0.1351	생활잡화	0.5318	대형가전	0.0008
10	시티웨어	0.1319	골프	0.5278	아동	0.0008

(e) 40세 초과 고객의 상품 네트워크

	밀도		근접중심성		매개중심성	
1	남성 캐주얼	0.1565	기타의류	0.5386	남성 캐주얼	0.0021
2	기타의류	0.1481	남성 캐주얼	0.5355	섬유잡화	0.0016
3	섬유잡화	0.1443	섬유잡화	0.5318	아동	0.0016
4	주방용품	0.135	셔츠	0.5285	주방용품	0.0014
5	셔츠	0.1337	주방용품	0.5282	셔츠	0.0014
6	침구/수예	0.1273	침구/수예	0.5258	트래디셔널	0.0013
7	트래디셔널	0.1243	트래디셔널	0.5231	침구/수예	0.0013
8	아동	0.115	아동	0.5189	골프	0.0012
9	골프	0.1064	대형가전	0.5176	주방가전	0.0012
10	란제리/내의	0.1037	골프	0.5175	기타의류	0.0012

34세 이상 고객의 상품네트워크에서는 아동 상품군이 특징적으로 나타났으며, 38세 이상 연령층에서는 골프 상품군이, 40세 이상 연령대에서는 대형가전이 근접중심성이 높은 상품군으로 나타났다.

상품군별 매개 중심성 비교에서는 37세 이하 고객의 모든 상품 네트워크에서 동일한 상품군이 상위 10개에 포함된다. 즉, 매개중심성은 37세 이하 연령층에서 두드러진 차이를 나타내지 않음을 알 수 있다. 이외에 38세에서 40세 고객의 상품 네트워크에서 매개역할을 하는 상품 노드가 많이 포함

되는 상품군은 대형가전이며, 40세 이상에서는 주방가전으로 나타났다.

### 3.3 시즌별 상품 네트워크 분석

본 절에서는 연구데이터를 계절별 네 개 집합으로 분류하여 상품 네트워크를 분석하였다. 각 시즌은 봄(3월~5월), 여름(6월~8월), 가을(9월~11월), 겨울(12월~2월)로 데이터가 분류되었으며, 각 데이터의 분포는 <표 9>과 같다.

&lt;표 9&gt; 시즌별 데이터 분포

	봄	여름	가을	겨울
고객수	2926	2984	3082	2943
상품수	1785	1 802	1782	1847
거래건수	15594	16781	16507	14454

&lt;표 10&gt; 상품 네트워크 분석 : 계절별

	봄	여름	가을	겨울
밀 도(연결중심성)	2926	2984	3082	2943
연결집중도	0.19746	0.28695	0.26986	0.2137
근접중심성	0.444391	0.448142	0.442027	0.425741
근접중심도	-	-	-	-
매개중심성	0.0007	0.000669	0.000696	0.000701
매개집중도	0.01209	0.0182	0.02099	0.02355
고립 노드수	8	13	14	22

시즌별 상품 네트워크의 특징은 <표 10>과 같다. 여름의 상품 네트워크는 밀도와 연결집중도 모두 높게 나타나, 고객들이 여름에 다양한 상품을 구매하지만, 동시에 일부 상품에 구매가 집중되는 경향이 두드러지는 시즌임을 알 수 있다. 반면에

봄의 상품 네트워크에서는 연결집중도가 가장 낮게 나타나 특정 상품에 연결이 집중되는 경향이 상대적으로 약한 시즌으로 분석되었다. 여름의 상품네트워크 밀도가 비교적 높다는 것으로부터 근

&lt;표 11&gt; 중심성간 상관관계 : 시즌별

(a) 봄시즌 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.774**	0.895**
근접중심성	0.774**	1	0.569**
매개중심성	0.895**	0.569**	1

(b) 여름시즌 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.718**	0.873**
근접중심성	0.718**	1	0.502**
매개중심성	0.873**	0.502**	1



(c) 가을시즌 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.722**	0.868**
근접중심성	0.722**	1	0.493**
매개중심성	0.868**	0.493**	1

(d) 겨울시즌 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.687**	0.876**
근접중심성	0.687**	1	0.471**
매개중심성	0.876**	0.471**	1

\*\* : Significant at  $p < 0.01$  level.

접중심성이 다소 높을 것으로 예측되었으며, 실제 분석 결과도 높은 값이 측정되었다. 그러나 시즌별 상품네트워크의 매개중심성은 모든 시즌에서 비슷한 수준으로 측정되어, 시즌별차이가 나타나지 않았다. 다만 가을과 겨울에서 봄과 여름보다 높은 매개집중도가 나타나는 것으로 보아 매개자 역할을 하는 상품의 집중 정도가 가을과 겨울에서 다소 높을 것으로 판단된다.

밀도와 중심성성간 상관분석 결과는 <표 11>에 요약하였다. 모든 시즌의 상품 네트워크에서 밀도와 중심성 지표간의 상관관계는 비교적 높게 나타났다으나, 중심성 지표간 상관관계는 비교적 낮게 나타나는 경향이 있다.

네트워크 밀도와 중심성값이 높게 나온 상품군 분포를 통하여 시즌별 판매상품의 차이를 살펴본 결과는 <표 12>와 같다.

<표 12> 상품군별 중심성 비교 : 시즌별

(a) 봄시즌 상품 네트워크

	밀 도		근접중심성		매개중심성	
1	남성 캐주얼	0.094	남성 캐주얼	0.4876	섬유잡화	0.0023
2	섬유잡화	0.089	섬유잡화	0.4876	남성 캐주얼	0.0022
3	주방용품	0.0734	셔츠	0.4792	주방용품	0.0017
4	셔츠	0.0732	시티웨어	0.4718	란제리/내의	0.0014
5	시티웨어	0.067	트래디셔널	0.4716	셔츠	0.0014
6	트래디셔널	0.066	주방용품	0.4709	화장품	0.0014
7	란제리/내의	0.0624	란제리/내의	0.4697	시티웨어	0.0014
8	아동	0.0564	침구/수예	0.4674	트래디셔널	0.0013
9	골프	0.0559	아동	0.4668	골프	0.0011
10	화장품	0.0534	커리어	0.4617	아동	0.001





(b) 여름시즌 상품 네트워크

	밀도		근접중심성		매개중심성	
1	트래디셔널	0.0942	트래디셔널	0.4898	트래디셔널	0.0021
2	남성 캐주얼	0.0939	침구/수예	0.4895	남성 캐주얼	0.0021
3	침구/수예	0.0884	주방용품	0.4818	침구/수예	0.0019
4	주방용품	0.0879	남성 캐주얼	0.4817	주방용품	0.0018
5	시티웨어	0.0748	시티웨어	0.4771	섬유잡화	0.0017
6	란제리/내의	0.072	섬유잡화	0.4721	란제리/내의	0.0014
7	섬유잡화	0.0714	화장품	0.4711	시티웨어	0.0013
8	셔츠	0.0695	란제리/내의	0.4673	주방가전	0.0011
9	화장품	0.0625	골프	0.4641	화장품	0.0011
10	기타의류	0.0593	커리어	0.462	기타의류	0.0011

(c) 가을시즌 상품 네트워크

	밀도		근접중심성		매개중심성	
1	섬유잡화	0.0992	주방용품	0.4918	섬유잡화	0.0032
2	주방용품	0.097	섬유잡화	0.4857	주방용품	0.0025
3	남성 캐주얼	0.0868	남성 캐주얼	0.4801	남성 캐주얼	0.002
4	침구/수예	0.0777	셔츠	0.4775	침구/수예	0.0019
5	셔츠	0.076	침구/수예	0.474	셔츠	0.0014
6	란제리/내의	0.0681	아동	0.4702	화장품	0.0013
7	아동	0.0626	란제리/내의	0.468	란제리/내의	0.0013
8	트래디셔널	0.0593	생활잡화	0.4609	아동	0.0011
9	화장품	0.0553	트래디셔널	0.4609	생활잡화	0.001
10	시티웨어	0.0552	시티웨어	0.4558	골프	0.001

(d) 겨울시즌 상품 네트워크

	밀도		근접중심성		매개중심성	
1	섬유잡화	0.086	섬유잡화	0.483	섬유잡화	0.0034
2	주방용품	0.0678	침구/수예	0.4641	주방용품	0.002
3	남성 캐주얼	0.0607	남성 캐주얼	0.4572	남성 캐주얼	0.0017
4	침구/수예	0.0572	주방용품	0.4563	트래디셔널	0.0016
5	트래디셔널	0.0546	트래디셔널	0.4549	화장품	0.0015
6	아동	0.0519	아동	0.4517	란제리/내의	0.0014
7	화장품	0.0497	셔츠	0.4468	침구/수예	0.0012
8	란제리/내의	0.047	시티웨어	0.4436	아동	0.0012
9	골프	0.0432	란제리/내의	0.4435	주방가전	0.0012
10	셔츠	0.0424	화장품	0.4415	골프	0.0009

<표 13> 지역별 데이터 분포

	강남	강북
고객수	1372	1889
상품수	1411	1658
거래건수	12662	21633

골프 상품군은 봄과 겨울의 상품 네트워크에서 밀도가 높은 상품군에 속한다. 이를 통하여 봄과 겨울에 홍보행사 등을 통하여 높은 판매량이 발생할 수 있지만, 타상품들과의 관계가 보다 폭넓게 확장되는 계절은 중심성이 높게 나타나는 다른 계절일 가능성이 있음을 알 수 있다. 또한, 여름에는 계절적 특징으로 인하여 일반의류 매장에서 판매되는 상품 외에 홈웨어나 바캉스의류가 포함되는 기타의류 상품군에 대한 판매가 두드러지는 특징을 나타낸다. 한편, 매개중심성이 높은 상품군은 모든 계절에서 남성 캐주얼, 란제리/내의, 섬유잡화, 주방용품, 화장품 등을 포함하는 것으로 나타났다.

### 3.4 지역별 상품 네트워크 분석

본 절에서는 연구데이터를 강남과 강북지역으

<표 14> 상품 네트워크 분석 : 지역별

	강남	강북
밀도	0.107276	0.150714
연결집중도	0.37695	0.43823
근접중심성	0.51037	0.531869
근접집중도	-	-
매개중심성	0.000677	0.000536
매개집중도	0.01157	0.00837
고립 노드수	5	3

로 나누어 상품 네트워크를 생성하여 분석하였다. 분류된 데이터의 분포는 <표 13>과 같다. 강북에 위치한 백화점에서 고객 1인당 연간 평균 구매건수는 약 11로 강남보다 다소 높으며, 구매되는 상품 또한 강남지역보다 다양한 것으로 나타났다.

<표 14>는 지역별 상품 네트워크의 분석 결과를 요약한 표이다. 결과에서 보여주듯, 강북지역의 상품 네트워크는 강남지역의 상품 네트워크보다 밀도와 연결집중도가 높다. 즉, 강북지역에서 강남 지역보다 다양한 상품이 팔리는 동시에 특정 제품에 구매가 집중되는 경향을 알 수 있다. 근접중심성은 강북지역이 다소 높지만, 큰 차이가 나지 않

<표 15> 중심성간 상관관계 : 지역별

(a) 강남지역 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.790**	0.883**
근접중심성	0.790**	1	0.624**
매개중심성	0.883**	0.624**	1

(b) 강북지역 상품 네트워크

	연결중심성	근접중심성	매개중심성
연결중심성	1	0.896**	0.847**
근접중심성	0.896**	1	0.733**
매개중심성	0.847**	0.733**	1

\*\* : Significant at p < 0.01 level.

아 거리개념으로는 네트워크 간 차별성을 분명히 할 수 없었다. 그러나 매개 집중도는 강남지역이 더 높게 나타나, 네트워크에서 상품간 브로커역할을 하는 상품노드를 보다 쉽게 찾을 수 있는 상품 네트워크를 생성하고 있다.

지역별 밀도와 중심성성간 상관분석 결과는 <표 15>와 같다. 강남지역의 상품네트워크는 강북지역의 상품 네트워크에 비하여 지표간 상관계

수가 다소 낮은 것으로 나타났다.

지역별 분석에서도 동일한 방법으로 상품노드가 포함된 상품군별 특성을 조사하였다. 지역별로 각 상품군에 포함된 상품 노드의 밀도와 중심성값의 평균값을 조사한 결과는 <표 16>과 같다.

강남과 강북지역의 상품 네트워크는 <표 16>이 보여주는 바와 같이 등위에는 차이가 존재하지만, 분석 지표별로 포함되는 상위 10개 상품군은 유사

<표 16> 상품군별 중심성 비교: 계절별  
(a) 강남지역 상품 네트워크

	밀도		근접중심성		매개중심성	
1	주방용품	0.2101	기타의류	0.5529	주방용품	0.0022
2	기타의류	0.2007	주방용품	0.55	남성 캐주얼	0.0022
3	남성 캐주얼	0.1867	시티웨어	0.5456	섬유잡화	0.0016
4	시티웨어	0.1854	남성 캐주얼	0.5418	시티웨어	0.0014
5	섬유잡화	0.1755	섬유잡화	0.5406	침구/수예	0.0014
6	침구/수예	0.1665	침구/수예	0.5388	기타의류	0.0014
7	셔츠	0.1578	셔츠	0.5317	셔츠	0.0012
8	트래디셔널	0.1447	란제리/내의	0.5295	트래디셔널	0.0011
9	란제리/내의	0.1432	트래디셔널	0.527	골프	0.001
10	골프	0.1355	골프	0.5245	란제리/내의	0.001

(b) 강북지역 상품 네트워크

	밀도		근접중심성		매개중심성	
1	주방용품	0.3077	주방용품	0.5907	주방용품	0.0022
2	침구/수예	0.2694	침구/수예	0.5774	섬유잡화	0.0015
3	트래디셔널	0.2516	트래디셔널	0.5721	침구/수예	0.0014
4	섬유잡화	0.2516	섬유잡화	0.5717	남성 캐주얼	0.0013
5	시티웨어	0.2483	시티웨어	0.5701	트래디셔널	0.0013
6	생활잡화	0.2419	생활잡화	0.5656	시티웨어	0.0012
7	남성 캐주얼	0.2311	남성 캐주얼	0.5627	생활잡화	0.0012
8	란제리/내의	0.2166	란제리/내의	0.5596	주방가전	0.0012
9	주방가전	0.2165	주방가전	0.5595	란제리/내의	0.001
10	셔츠	0.2144	셔츠	0.5562	화장품	0.0009

하게 조사되었다. 그러나 강남과 강북지역의 밀도가 높은 상품군에 다소간의 차이가 나타났다. 기타 의류와 골프 상품군은 강남지역의 상품 네트워크에서만 밀도가 높은 상품군에 포함되었고, 강북지역에서는 생활잡화, 주방 가전 상품군이 포함되어 주로 구매하는 상품군에 차이를 드러냈다.

상품 판매를 통해 매출을 증대시켜야 하는 기업들은 상당한 마케팅 비용을 투자해야 한다. 이때, 기업이 비용을 절감하고 마케팅 효과를 높이기 위해서는 적합한 대상상품과 고객을 선별하는 것이 중요하다. 이때, 이와 같은 네트워크 분석 결과가 활용될 수 있다. 예를 들어, 특정 상품군의 상품 홍보를 위하여 DM 발송 대상을 선정할 때, 세분화된 고객의 상품 네트워크에서 중심성과 집중도 분석결과가 의미 있는 차별성을 지닌 대상을 우선적으로 고려할 수 있다. 마찬가지로 특정 고객을 대상으로 적합한 홍보상품을 선택해야 하는 상황에서도 참조할 정보가 제공된다. 교차판매 또는 상향판매 등의 목적으로 상품 정보를 제공하는 경우, 또는 기업 전체 상품 판매에 영향을 주는 상품을 선정하여 집중 홍보 활동을 기획하는 경우가 가지적인 예가 될 것이다.

#### 4. 결론

본 논문에서는 대형 백화점의 거래 자료에 근거한 이분 네트워크로부터 상품 네트워크를 구성하여 그 특성들을 성별, 연령별, 계절별, 그리고 지역별로 그 특성들을 연구하였다. 구성된 상품 네트워크는 밀도, 중심성, 그리고 집중도 지표를 이용하여 분석되었다. 전반적으로 분석단위내에서 매개 중심성보다 근접중심성 결과의 차이가 크지 않은 것으로 나타났다. 이는 근접 중심성의 핵심 개념인 노드간 네트워크 거리 개념으로는 상품의 중요성

을 명확하게 포착할 수 없다는 것을 의미하기도 한다. 한편, 각 중심성값이 제공하는 정보의 차이를 확인하기 위한 중심성간 상관분석 실시 결과를 보면 통계적으로 유의한 상관관계를 지니지만, 각 중심성간에 차이가 존재함을 알 수 있었다. 특히, 밀도와 매개중심성, 또는 근접중심성간에는 높은 상관계수가 관측되었지만, 매개중심성과 근접중심성간에는 비교적 낮은 상관관계를 나타냈다. 즉, 상품간에 관계를 통제하는 브로커의 역할을 하는 상품은 전체상품간의 거리계산을 통해 얻어지는 근접중심성이 높은 상품과는 차이가 있음을 알 수 있다. 또한 보다 일반화된 분석결과를 얻기 위하여 상품노드가 포함된 상품군의 네트워크 특성을 조사한 결과에서도 분석단위별로 측정지표간 차이가 발견되었다. 이는 상이한 상품군에 대하여 각 지표의 활용방안을 달리해야 함을 시사한다. 거래 데이터를 이용한 상품 네트워크는 구매기록이 있는 상품을 대상으로 네트워크가 구성되기 때문에 고립된 상품 노드가 거의 없고, 상품 노드간 연계 거리가 짧은 네트워크를 생성하게 된다. 이와 같은 네트워크 구조는 근접 중심성과 밀도 등의 결과를 토대로 도출되기도 한다.

한편, 하나의 상품 네트워크에서 밀도와 연결 집중도가 동시에 높은 경우가 발견되었는데 이는 해당 분석단위내 고객들의 구매활동이 다양한 상품을 대상으로 이루어지고 있지만, 실제로 특정 상품을 집중적으로 구매하는 패턴이 있음이 의미한다. 즉, 판매 빈도가 매우 낮은 다수의 상품들과 판매빈도가 매우 높은 소수의 상품으로 구성된 네트워크 구조를 예상할 수 있다. 연구데이터에서는 여성고객의 상품 네트워크에서 이러한 특성이 발견되었다. 연령별로 생성된 상품 네트워크의 밀도는 고객의 라이프스타일과 구매력이 명확하게 반영되는 지표임이 발견되었다. 이외의 연령별 특징으

로는, 40세 이상의 연령집합에서는 가장 높은 매개 중심성을 갖는 상품 네트워크가 생성되는 반면, 31~33세 연령집합에서는 매개집중도가 높은 네트워크를 생성하는 것으로 나타났다. 지역별 상품네트워크에서 밀도와 중심성성간 상관분석 결과 강남지역의 상품네트워크는 강북지역의 상품 네트워크에 비하여 지표간 상관계수가 다소 낮은 것으로 나타났다.

본 연구에서는 기존의 사회 네트워크 연구자들이 주로 관심을 갖는 특성들에 초점을 두었지만, 네트워크의 기존 연구방법을 확장하여 적용한다면 온·오프라인 상품 판매 데이터를 이용한 여러 현실적이고 유용한 정보를 얻을 수 있다고 믿는다. 예를 들어, 마케팅부서에서 특정상품에 대한 중심성 지표를 마케팅 활동 시행 이전과 이후에 그려서 비교해 본다면, 마케팅 활동이 상품판매 패턴에 미치는 영향과 같은 것도 정량적이고 객관적으로 살펴 볼 수 있을 것이다. 이외에도, 상품 네트워크 분석의 핵심 가운데 하나는 각 상품노드의 역할과 위치를 구체적으로 도출된다는데 있다. 따라서 상품 네트워크상에서 소외된 변방에 위치한 상품 노드를 발견하고 네트워크상에서 연계된 다른 상품과의 관계를 고려하여 판매전략을 제고할 수 있는 기초자료를 제공할 수 있을 것이다.

본 연구에서 제시한 상품 전략은 소수 상품에 대한 판매증가의 관점이라기 보다는, 상품간의 구매관계 정보를 이용하여 교차판매, 상향판매, 그리고 추가판매 등을 보다 적극적으로 유도함으로써 판매상품의 다양성을 확보하기 위한 전략을 구축하는 방법이다. 장기적으로는 기업의 상품 매출 구조를 제고하기 위한 방법으로 심화 적용될 수 있을 것이다. 이와 같은 방법은 전략시행 전후의 네트워크 구조와 개별노드의 지수변화 비교를 통해 실행에 대한 평가를 정량적으로 할 수 있는 틀이

동시에 구축될 수 있다.

본 연구에서는 단일 데이터를 이용하여 상품 네트워크를 분석하였기 때문에 상품 네트워크의 특성을 일반화하기에는 한계가 있다. 향후 보다 다양한 데이터를 대상으로 네트워크 측정방법 및 적용방안에 관한 논의를 심화시킨 후속 연구가 이루어진다면 의미 있는 연구가 될 것이다.

## 참고문헌

- 김용학, 사회연결망 분석, 박영사, 2003.  
 김용학, 사회연결망 이론, 박영사, 2003.  
 손동원, 사회 네트워크 분석, 경문사, 2002.  
 Adomavicius, G. and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No. 6(2005), 734~746.  
 Agrawal, R., and Srikant, R. "Fast Algorithm for Mining Association Rules in Large Database", *In Proceedings of the 20th International Conference on Very Large Data Base*, (1994), 478~499.  
 Albert R. and A-L. Barabási, "Statistical mechanics of complex networks", *Reviews of Modern Physics*, Vol.74(2002), 47~97.  
 Alex, B., Stephen, S., and Kurt, T. *Building data mining applications for CRM*, New York, McGraw-Hill, 2000.  
 Arazy, I., O. Sana, B. Shapira, and N. Kumar. "Social relationships in recommender systems", *in Proceedings of the Seventeenth Annual Workshop on Information Technologies and Systems*, Montreal, Canada, (2007),

- 146~151.
- Barabási, A-L. and R. Albert “Emergence of scaling in random networks”, *Science*, Vol.286 (1999), 509~512.
- Huang, Z., D. Zeng, and H. Chen. “Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommendation Systems”, *Management Science*, Vol.53, No.7 (2007), 1146~1164.
- Chen, Y. L., K. Tang, R. J. Shen, and Y. H. Hu “Market basket analysis in a multiple store environment”, *Decision Support Systems*, Vol.40(2004), 339~354.
- Craig, S., B. Sergey, and M. Rajeev, “Beyond Market Baskets: Generalizing Association Rules to Dependence Rules”, *Data Mining and Knowledge Discovery*, Vol.2(1998), 39~68.
- Kim, H. K., J. K. Kim, and Y. U. Ryu, “Personalized Recommendation over a Customer Network for Ubiquitous Shopping,” *IEEE Transactions on Services Computing*, Vol.2, No.2 (2009), 140~151.
- Kim, H. K. and J. K. Kim, “Local Scoring Model for Recommender Network”, *In proceeding of 2009 KMIS Spring Conference*, (2009), 145~151.
- Michael, J. A. B. and S. L. Gordon, *Data Mining Techniques : For Marketing, Sales, and Customer Relationship Management*, Wiley, New York, Wiley, 2004.
- Newman, M. E. J. “Scientific collaboration networks: I. Network construction and fundamental results,” *Physical Review*, Vol.64 (2001), 016131.
- Newman, M. E. J., S. H. Strogatz, and D. H. Watts, “Random graphs with arbitrary degree distributions and their applications”, *Physical Review*, Vol.64(2001), 026118.
- Sarwar, B., G. Karypis, and et al. “Analysis of recommendation algorithms for e-commerce,” *In Proceedings of ACM E-commerce 2000 conference* (2000), 158~167.
- Shahriar, T. H. M., G. D. Joseph, and H. Liauqat, “Social Network Analysis and Organizational Disintegration: the Case of Enron Corporation”, *In Proceeding of 2007 International Conference on Information Systems*, 2007.
- Wouter, D. N., M. Andrej, and B. Vladimir, “Exploratory Social Network Analysis with Pajek”, *Cambridge University Press, CAMBRIDGE*, 2005.
- Zhou, T., J. Ren, M. Medo, and Y. C. Zhang, “Bipartite network projection and personal recommendation”, *Physical Review*, Vol.E76 (2007), 046115.

Abstract

## A Network Approach to Derive Product Relations and Analyze Topological Characteristics

Hyea-Kyeong Kim\* · Jae-Kyeong Kim\* · Qiu-Yi Chen\*

We construct product networks from the retail transaction dataset of an off-line department store. In the product networks, nodes are products, and an edge connecting two products represents the existence of co-purchases by a customer. We measure the quantities frequently used for characterizing network structures, such as the degree centrality, the closeness centrality, the betweenness centrality and the centralization. Using the quantities, gender, age, seasonal, and regional differences of the product networks were analyzed and network characteristics of each product category containing each product node were derived. Lastly, we analyze the correlations among the three centrality quantities and draw a marketing strategy for the cross-selling.

**Key Words** : *Product Network, Social Network Analysis, Centrality, Centralization*

---

\* School of Business Administration & Management Research Institute, Kyunghee University

## 저자 소개



김혜경

현재 경희대학교 경영대학에서 연구교수로 재직하고 있다. 경희대학교 물리학과에서 학사, 일반대학원 경영학과에서 MIS 전공으로 석사학위와 박사학위를 취득하였으며, 박사학위논문은 경희대학교 최우수학위논문으로 선정되었다. 주요 관심분야는 고객 관계관리, 상품추천시스템, 사회연결망분석, 복잡계 시스템 등이며, 경영정보학회 학술발표대회 및 지능정보시스템학회 학술발표대회에서 최우수논문상 및 우수논문상을 총5회 수상하였다. IEEE Transactions on Services Computing, Expert Systems, Expert Systems With Applications, LNCS, LNAI 등 에 논문을 게재하였으며, WITS(Workshop on Information Technologies and Systems) 2007, WITS 2008, WEB(Workshop on eBusiness)2009에서 논문을 발표하였다.



김재경

서울대학교에서 산업공학 학사, 한국과학기술원(KAIST)에서 경영정보시스템 전공으로 석사 및 박사학위를 취득하였다. 현재 경희대학교 경영대학 교수로 재직하고 있으며, 한국경영정보학회 부회장, BK21 네트워크 과학에 기반한 경영학 전문연구인력 양성사업팀 단장, 저탄소녹색성장국민포럼산하 그린IT분과위원을 역임중이다. 주요 연구분야로는 비즈니스 인텔리전스, 네트워크 경영, 그린 비즈니스/IT 등이다. Artificial Intelligence Review, Electronic Commerce Research and Applications, European Journal of Operational Research, Expert Systems with Applications, Group Decision and Negotiations, IEEE Transactions on services computing, International Journal of Human-Computer Studies, Technological Forecasting and Social Change, 등 다수의 학술지에 논문을 게재하였으며, 또한 학술지 지능정보연구 편집위원장, Information Technology and Management(SSCI) AE(Associate Editor)를 역임중이다.



Chen Qiu Yi

중국 Nanjing University of Technology 경영학과에서 학사를 취득하였다. 현재 경희대학교 대학원에서 MIS 전공 석사과정에 재학 중이다. 관심분야로는 CRM, 추천시스템, 데이터마이닝 등이다.