

소셜 네트워크 분석(SNA)을 활용한 택배 서비스 수요 분석

Delivery Service Demand Analysis Using Social Network Analysis (SNA)

오 경 은 (Kyungeun Oh)	KAIST 경영대학 석사
김 수 림 (Sulim Kim)	KAIST 경영대학 박사과정
최 한 별 (HanByeol Stella Choi)	명지대학교 경영정보학과 조교수, 교신저자
이 희 석 (Heeseok Lee)	KAIST 경영대학 교수

요 약

코로나19 이후 비대면 소비 사회로의 전환이 일어나며 생활밀착형 라스트마일 물류 서비스인 택배 기반의 세분화된 도시 물류의 중요성이 대두되고 있다. 본 연구는 소셜 네트워크 분석(SNA)을 활용하여 지역 및 상품 특성별로 수요의 관계성을 분석하고자 한다. 전자상거래에서 수요 파악에 활용되었던 장바구니 네트워크와 공동 구매 상품 네트워크를 확장한 분석 결과 택배 주문 건 수 데이터 속에서 상품 카테고리 패턴을 발견하고 지역별 차이를 확인하였다. 이를 바탕으로 본 연구에서 제시하는 SNA 기반 분석이 도시 물류의 재고 분산 배치나 상품(SKU) 선정 전략에 효과적으로 적용될 수 있음이 확인되었다.

키워드 : 배송, 택배, 수요 분석, 커뮤니티 탐지, 소셜 네트워크 분석

I. 서 론

전자상거래의 영향으로 2000년대 이후 국내 택배 산업은 가파른 성장세를 보이며, 택배 시장의 전체 매출액은 2015년 4조 3,440억 원에서 2020년 7조 4,920억 원으로 72.5% 성장했다(노정명, 2021). 택배 서비스 수요의 급증은 공급시스템 전반에 영향을 줄 수 있어, 택배 산업의 변화 동향을 예측하여 선제적 대응방안을 마련하는 것이 중요하다(박

한영 등, 2021).

특히, 코로나19 이전과 이후 택배 산업의 가장 큰 차이점은 고객의 택배 서비스 이용 빈도가 증가했다는 점이다. 국내 택배 이용 횟수는 2020년 경제활동인구 1인당 약 122회로 전년 대비 22.7회 증가하였다.¹⁾ 비대면 소비 사회로의 전환이 빠르

- 1) ① 국가물류통합정보센터a. ‘년도별 생활물류통계 - 택배 물동량 추이’자료. 2021. <https://www.nlic.go.kr/nlic/parcelServiceLogistics.action>.
② 국가물류통합정보센터b. ‘년도별 생활물류통계 - 택배 이용횟수 추이’자료. 2021. <https://www.nlic.go.kr/nlic/parcelServiceUseCount.action>.

† 본 논문은 카이스트 경영대학 디지털 혁신연구센터의 지원과 카이스트 경영대학 공공혁신연구센터 연구 기금에 의하여 연구되었음.

게 일어나며, 택배나 배달 대행 등 생활밀착형 물류 서비스가 가정에 필수 서비스로 자리잡은 것이다. 또한, 택배를 중심으로 도시 물류의 공간적 범위가 넓어지고 동시에 세분화되는 경향이 나타났다. 광역시·도 단위의 도시 물류 정책 외에도 시·군·구 또는 읍·면·동 단위 도시 물류 정책의 필요성이 제기되고 있다(한국교통연구원, 2021). 물류시설 측면에서는 지금까지 대형 물류단지 중심으로 신규 단지 개발이 주로 진행되었다면 이제는 중소규모의 복합물류시설에 대한 정책 강화가 필요하다는 목소리가 나오고 있다(정승주, 2015).

나아가 고객의 택배 이용 품목 측면에서도 변화가 있었다. 식품 카테고리의 경우 코로나19 발생 전후로 이용 건수가 2.19건에서 4.03건으로 83.8% 증가하며 가장 큰 폭의 변화를 보였다. 도서 61.6% 증가, 가전제품 54.7% 증가, 생활용품 49.6% 등 전반적으로 증가했지만, 패션과 같은 품목의 경우 감소하였다(박한영 등, 2021). 코로나 발생 전후로 상품 카테고리별 택배 이용 건수 증가율의 변화가 각기 다르다는 점은, 상품의 민감도에 따라 택배 서비스의 공급 전략에 차별화가 필요함을 의미한다.

택배 산업의 성장에 따라 중요해지고 복잡해진 고객의 니즈를 충족하기 위해, 택배 상품 카테고리 특성 및 지역 세분화에 기반한 분석이 대두되고 있다. 이는 도시 물류나 중소규모 복합물류시설의 상품·재고 공급전략에도 활용될 수 있다. 최근 기업의 물류 운영 방식의 변화가 요구되고 있으며, 고객 수요의 폭발적인 증가에 대비하는 네트워크 유연성 확보가 중요한 이슈로 떠오르고 있다. 특히 마이크로 풀필먼트 센터와 같은 도시 물류 또는 중소형 물류센터는 공간적 제약으로 인해 대형 물류센터에 비해 상품 선정 및 재고 관리가 중요하다.

학술적인 관점에서 택배와 관련된 기존의 국내 연구는 주로 택배 물량 예측 모형이나 물류 비용 최적화에 한정되어 있었다(김태현 등, 2020; 박인선, 이상현, 2018; 이양수 등, 2005). 물량 예측 모형은 택배 서비스 수요의 합계 값에만 초점을 두었

기에 상품 및 지역 관점에서 특성을 발견하는 관점의 연구는 부족한 실정이다. 물류 비용 최적화는 택배 물량 데이터에 네트워크 개념을 적용한 데서 의미가 있으나, 적절한 물류센터나 작업자의 수를 도출하는 수학적 문제에 초점을 두고 있어 수요에 관한 연구라고 보기는 어렵다. 소셜 네트워크 분석(SNA, Social Network Analysis)을 적용하여 택배 물량 데이터 속에서 상품과 지역의 패턴을 발견하고, 택배 서비스 수요 연구는 이런 관점에서 중요하다. 항공·항만 분야에서 SNA를 적용하여 공급망의 중심성 척도 분석을 진행한 사례는 있었지만, 배송 또는 택배 분야에서는 SNA를 적용하여 분석한 연구가 미진하다(김벽진, 안승범, 2017; 임병학, 2011). 또한, 중심성 척도 분석 외에도 이분 네트워크 구조 및 커뮤니티 탐지와 같은 다양한 SNA 기법의 물류 분야로의 확대도 필요하다. 이를 통한 택배 서비스 수요 연구는 코로나19 이후 프리 뉴 노멀(pre-new normal) 시대에 필요한 생활물류 전략을 마련할 수 있을 것으로 보인다.

본 연구에서는 SNA(Social Network Analysis)를 적용하여 상품 카테고리 및 지역 특성 관점에서 택배 서비스 수요를 분석하고 유의미하게 묶을 수 있는 수요 그룹을 확인할 수 있는 방법을 제시하고자 한다. 본 분석을 위하여 물류 현장에서 상품 및 재고를 분산 배치하는 데 유효한 전략을 도출하고, 나아가 택배산업의 제조·운송·창고 네트워크를 재구성하는 데 활용할 수 있는 이론적 근거를 마련하고자 한다.

II. 문헌 연구

2.1 국내 택배 서비스 수요 연구

국내 택배 서비스의 수요에 대한 연구는 정량적 연구가 주를 이루어 왔다. 코로나19 이후 소비자의 택배 이용행태 변화를 분석한 연구는 설문조사를 기반으로 택배 서비스에 관한 양적, 질적 변화

동향을 연구하였다. 특히 택배사업자가 서비스를 공급함에 있어서 상품 카테고리별로 나누어서 전략을 구상할 필요성이 있음을 시사하였고, 향후 택배 서비스 수요 연구의 방향을 설정하는 데 기여하였다(박한영 등 2021). 택배 서비스를 지역 관점에서 분석한 한국교통연구원의 보고서는 지역 간 택배 물량 자료를 기초로 하여 국내에서 처음으로 O-D(Origin-Destination) 연구를 시도하였다(서상범 등, 2015). 전국 단위 지역 간 물량 이동 실태를 분석하여, 택배 거점 네트워크 재설계나 지역특성에 따른 차별적 정책 발굴 등을 위한 접근체계를 마련하였다. 또한 지역에 대한 체계적 분석을 통해 지역특성에 맞는 정책 개발이 필요함을 시사하였다.

택배 물량 또는 수요 예측 모형 연구는 회귀모형, 시계열모형, 비정형 데이터를 접목한 회귀모형 등 다양한 기법을 적용하여 진행되어 왔다(박인선, 이상현, 2018; 이우승, 2004; 전현주, 유명식, 2020). 이러한 연구들은 대부분 택배 물량 합계 값의 예측에 초점을 맞춘 것으로, 산업의 성장성을 예측하고 그 규모를 추정하는 데 유용하다. 또한 택배 서비스 수요에 영향을 미치는 요인이 무엇인

지를 제안하기도 했다. 하지만 수요와 관련하여 양적인 측면에서만 시사점을 제공하고 있어, 서비스의 질적인 개선 및 상세한 전략에 실질적으로 적용하기는 어렵다는 한계가 있었다. <표 1>에 정리된 바와 같이 기존의 연구들은 상품이나 지역 관점에서 택배 서비스 수요를 분석하는 것이 의미있음을 시사하였지만, 실제로는 수요의 합계 값을 예측하는 모형을 개발하고 영향을 주는 요인들을 탐색하는 연구에 그쳤음을 알 수 있었다. 또한 연구결과를 실제 물류 현장에 적용하기 위한 시도까지는 이루어지지 않았다. 따라서 택배 서비스 수요를 상품 및 지역 간의 관계 관점에서 세분화하여 분석하고, 도시 물류의 상품 선정 및 재고 수준을 결정하는 데 실질적으로 활용할 수 있는 연구가 필요함을 확인할 수 있었다.

2.2 물류 분야 SNA 활용 연구

택배 서비스 수요를 상품 및 지역 간의 관계 관점에서 분석하는 데에는 SNA 기법을 활용할 수 있다. SNA는 사회 연결망 데이터를 활용하여 사회적 실체 간의 관계 구조를 분석하는 학제적 연

<표 1> 국내 택배 서비스 수요 관련 선행 연구

출처	연구 방법론	내용
박한영 등(2021)	설문조사	<ul style="list-style-type: none"> 코로나19 이후 양적 측면에서 이용건수 56.5% 증가, 질적 측면에서 이용채널, 품목, 수령방법 변화 발생 이용 품목별 민감도에 따라 택배사업자의 서비스자원 공급전략 차별화 필요
서상범 등(2015)	O-D 추정 및 상관관계 분석	<ul style="list-style-type: none"> 수도권 택배 물량이 전체의 약 35%를 차지 대규모 수요지 인근의 물량 집중으로, 도심 물류 효율화를 위한 네트워크 변화(multi-hub), 인프라 및 제도 개선 필요 택배 물량은 지역별로 도착과 발생 특성이 확연하게 다르므로 지역에 대한 체계적 분석 필요
이우승(2004)	회귀분석	<ul style="list-style-type: none"> 사회경제적 요인(전자상거래 법규 및 제도의 정비, 정보인프라에 대한 투자, 컴퓨터 보급 등)의 변화에 따른 택배 물량 인과관계 분석 및 수요 예측
박인선, 이상현 (2018)	시계열분석	<ul style="list-style-type: none"> 연평균성장률에 의존하던 기존 모형에서 나아가 시계열 패턴으로 수요 예측 단일지수평활법, 윈터스의 가법모형, 계절형 ARIMA모형 중 계절성을 고려한 ARIMA 모형이 가장 적합
전현주, 유명식 (2020)	회귀분석	<ul style="list-style-type: none"> 비정형 데이터(상관계수를 활용한 키워드)를 접목하여 택배 프로세스의 비용을 최소화하는 다중회귀모형으로 수요를 예측하여 예측률 1.2% 개선

구로, 사회학, 심리학, 생물학, 물리학 등 다양한 분야에서 사용되고 있는 접근법이다(Carter, 2008; 강성경 등, 2016). SNA는 분석 단위(unit of analysis)로써 행위자에 초점을 맞추는 것이 아니라, 네트워크를 구성하고 행위자 간의 관계에 초점을 맞추기 때문에 기존의 연구 방법론으로는 달성되지 않는 결과를 얻을 수 있다(Craig et al., 2007). 과거 물류 관련 문헌연구들의 연구 방법론은 사례연구(Case study)가 주요했다. 하지만, 정량적 엄격성이 요구되면서 점차 SNA가 대안적인 연구 방법으로 각광받게 되었고, 물류 분야에서도 SNA를 적용하는 다양한 시도들이 생겨났다(Craig et al., 2007). 특히, 공급망 관련 연구를 위해 주로 SNA를 적용하였는데, 중심성 척도를 가지고 자동차 공급망을 분석하거나, 항공 공급망의 네트워크 밀도가 공급망 통합에 미치는 영향을 연구한 사례 등이 있다(Kim et al., 2011; Lee, 2005). 항공 분야에서는 비행 지연이 발생하는 메커니즘을 분석하고 위험 관리 전략을 제시하거나, 항공화물 글로벌 네트워크 구조와 지역별 물동량 패턴을 분석했으며, 항공우편물 네트워크의 중심성 변화를 추적하여 항공운송 물류 네트워크 역량을 연구하기도 했다(Wang and Li, 2020; 김벽진, 안승범, 2017; 손재철 등, 2021). 항만 분야에서는 컨테이너항만 네트워크의 중심성과 구조적 공백 분석을 통해 항만 생산성에 영향을 주는 변수를 분석하는 시도가 있었다(임병학, 2011). 운송 분야에서는 운송 계획을 개선하기 위한 도구로써 SNA의 효과를 입증한 연구가

진행되었다(Islam et al., 2017).

선행 연구를 살펴본 결과, <표 2>와 같이 국내외 다양한 물류 연구 중 택배 분야에 SNA를 적용한 연구는 본 연구가 최초의 시도이다. 또한 그동안 물류 분야에 적용된 SNA는 노드의 중심성 척도를 분석한 경우가 대다수였기에, 본 연구에서는 전자상거래에서 주로 진행되었던 상품 네트워크 연구에 착안하여 커뮤니티 탐지 등을 새롭게 시도해 보았다.

2.3 상품 네트워크 연구

전자상거래의 급격한 성장과 함께 고객들은 상품 과부하에 직면하였다. 고객들이 서비스에 노출되는 상품을 쉽게 고르지 못하는 문제를 겪게 되면서, 일대일 마케팅, 개인화, CRM 등의 마케팅 전략 필요성이 강조되었다(Sawar et al., 2000). 장바구니 분석은 보다 적절한 상품을 고객에게 제안하기 위해 등장한 연관 규칙 데이터 마이닝 기법으로, 한 번에 거래되는 결제 데이터 또는 단일 쇼핑 상황에서 고객이 구입한 제품의 바스켓(묶음) 구성을 연구하는 방법론이다(Gary and Ann, 2000). 본 방법론은 수백 또는 수천 개의 중복되는 연관 규칙 중에서 흥미로운 규칙을 강화하는 데 집중하는데, 다소 주관적일 수 있는 ‘흥미로운 규칙 강화’의 명확성을 개선하기 위한 시도가 진행되어 왔다. 그 시도 중 하나가 네트워크를 모델링하여 상품 커뮤니티를 탐지하는 방식인데, 이를

<표 2> 물류 분야 SNA 활용 선행 연구

출처	물류 관련 연구 분야	활용한 SNA 개념
Kim et al.(2011)	자동차 공급사슬	중심성 척도(연결중심성, 매개중심성, 근접중심성)
Lee(2005)	항만 공급사슬	연결중심성 및 밀도
Wang et al.(2020)	항공 연착	중심성 척도(연결중심성, 매개중심성, 근접중심성)
김벽진, 안승범(2017)	국제 항공화물	지니지수, 군집화 계수, 밀도, 가중밀도
손재철 등(2021)	항공우편물	중심성 척도(연결중심성, 매개중심성, 근접중심성)
임병학(2011)	컨테이너 항만	중심성 척도(연결중심성, 근접중심성, 고유벡터)
Islam et al.(2017)	교통 계획	중심성 척도(매개중심성, 고유벡터)

장바구니 네트워크(Market Basket Network, MBN)라고 한다(Raeder and Nitesh,2011). 한편, 언제 어디서든 인터넷만 연결되면 원하는 상품을 구매할 수 있게 되면서 한 번에 여러 상품을 주문할 필요가 줄어들었다. MBN은 단일 거래 내에서 동시에 구매한 상품 간의 연관성에 초점을 맞추고 있었기에, 동시성에 대한 의존도를 낮추는 새로운 연구 방법이 필요했다. 이러한 맥락에서 등장한 상품 네트워크 분석(Product Network Analysis)은 일정 기간 내에 함께 구매된 모든 상품 간의 관계에 대한 네트워크로 확장한 연구다(Kim et al. 2012). 거래 데이터는 고객-상품 이분 네트워크로 표현할 수 있는데, 여기에 투영 기법을 적용하여 고객이 공동으로 구매한 상품끼리 연결되는 네트워크를 추출한다. 이를 공동 구매 상품 네트워크(Co-purchased Product Network, CPN)라고 한다(Huang et al. 2007).

MBN과 CPN의 위상학적 특징과 성능을 분석한 연구(Kim et al., 2012) 에 따르면, 동시 구매 여부와 무관하게 상품 간의 관계를 파악할 수 있는 CPN이 MBN보다 더 많은 잠재적 상품 정보를 제공할 수 있으며, 일상적으로 필요하지만 자주 구매되지 않

는 상품들에 대한 정보까지 담을 수 있었다. 장기적으로 고객의 구매 선호도를 분석할 때에는 CPN이 보다 적절한 방법론이라고 할 수 있다.

상품 네트워크는 고객 수요 관점에서도 활용 가능하다. 고객의 수요는 고객이 구매했거나 관심이 있는 상품에 반영되는데, 동일한 고객이 단기간에 함께 구매했거나 구매할 예정인 둘 이상의 상품을 공동 수요 상품(Co-demanded products)으로 볼 수 있다. 공동 수요 상품 네트워크(DACCN)를 기반으로 고객 수요를 발견하고 커뮤니티를 탐지하는 모델에 대한 연구가 진행되기도 했다(Wu et al., 2021).

<표 3>과 같이 상품 네트워크 관련 선행 연구를 살펴보면, 택배 서비스 수요와 전자상거래의 고객 수요가 유사한 점이 있음을 알 수 있었다. 택배 서비스 수요 역시 일정 기간 동안 복수의 상품을 함께 구매하고자 하는 CPN의 특성을 가지고 있다. 택배 물량 데이터는 전자상거래에서 발생한 주문 데이터를 기반으로 물류센터에서 출하되어 각 지역으로 배송되기 때문에, 고객-상품의 관계를 지역-상품의 관계로 치환하여 SNA를 적용 가능하다고 보았다.

<표 3> 상품 네트워크 관련 선행 연구

출처	연구 방법	주요 연구 내용
Russell et al.(2000)	장바구니 분석	<ul style="list-style-type: none"> 동일한 쇼핑 여정 내에서 소비자가 상품을 선택하는 의사결정 과정 장바구니 분석 모델 접근법
Troy Raeder et al. (2011)	장바구니 네트워크(MBN)	<ul style="list-style-type: none"> 장바구니 내 연관성 규칙 정량화를 위해 거래 데이터를 상품 네트워크로 모델링하고 커뮤니티 발견 상품 커뮤니티를 위한 척도 개발
Kim et al.(2012)	상품 네트워크	<ul style="list-style-type: none"> 모든 상품 간 관계에 대한 확장된 네트워크 수준에서 그 관계를 분석하기 위해 상품 네트워크 제안 MBN과 CPN의 위상학적 특징과 성능 분석
Huang et al.(2007)	공동 구매 상품 네트워크(CPN)	<ul style="list-style-type: none"> 고객 구매 행동을 이해하기 위해 거래 데이터를 활용하여 고객-상품 이분 네트워크 분석 과거 거래를 기반으로 상품을 추천하는 대규모 협업 필터링 기반 추천 알고리즘에 대한 정당성 확보
Wu et al.(2021)	공동 수요 상품 네트워크(DACCN)	<ul style="list-style-type: none"> 공동 수요 제품 네트워크를 기반으로 관련 소비자 수요를 발견하기 위한 모델 제안 중복 커뮤니티 탐지 알고리즘 사용 장바구니 분석에 비해 모듈화 되고 다양하고 실용적인 커뮤니티 탐지

Ⅲ. 연구 방법

본 연구는 택배 서비스 수요를 상품과 지역 관점에서 분석하기 위해 다음 연구 질문에 답하는 것을 목표로 한다.

- (1) 택배 서비스 수요 연구에 적합한 네트워크는 상품과 지역 네트워크 중 어떤 것인가?
- (2) 네트워크 내에서 그룹(상품 그룹 또는 지역 그룹)은 어떻게 나타나는가?
- (3) 그룹 내에서 중심이 되는 노드(상품 또는 지역)는 무엇인가?

이를 위해 <그림 1>과 같은 연구 절차에 따라 분석을 진행하였다. 우선 택배 물량 데이터를 이분 네트워크로 변환한 후 2개의 투영 네트워크를 생성한다. 이후 노드 간 유사도 분석이라는 기준에 따라 연구에 적합한 네트워크를 선정하고 커뮤니티를 탐지한 뒤, 해당 커뮤니티의 중심성 척도 및 수요 특성을 분석하는 프로파일링을 진행한다.

SNA는 사회적인 관계를 네트워크로 구성하는 데서 출발하는데, 행위자(개인 또는 조직)를 노드(node)로, 행위자 간의 관계를 링크(link)로 연결하며, 그 연결의 정도에 따라 가중치(weight)를 부여하여 표현한다. 택배 물량 데이터는 상품 카테고리, 배송 지역, 송장건수를 포함하고 있다. 여기서 상품 카테고리나 지역을 노드로 하고, 해당 상품 카테고리나 지역을 주문한 지역이 있을 경우 두 개의 노드를 연결하는 링크를 생성하여 네트워크 구조를 만들 수 있다. 이러한 형태의 네트워크는 상품과 지역이라는 두 가지 다른 성격의 노드 집합을 가

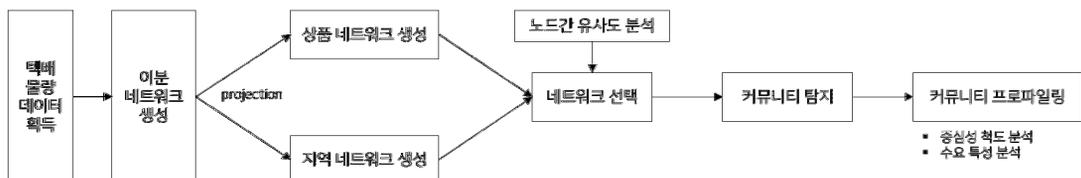
지고 있다는 점에서 이분 네트워크라고 부르는데, 투영 처리를 통해 상품 카테고리 노드 집합 또는 지역 노드 집합만으로 구성된 네트워크를 얻을 수 있다. 두 네트워크 중 택배 서비스 수요 분석에 보다 적절한 네트워크를 선택하기 위한 검증은 시도하였다. 노드 간의 성격 차이가 두드러지는 네트워크가 보다 유의미한 커뮤니티 탐지가 가능하다는 점에 착안하여 노드 간 유사도 분석을 수행하였다.

이 과정을 통해 선택된 네트워크에 몇 가지 커뮤니티 탐지 알고리즘을 적용하여 비슷한 성격을 가지는 그룹을 도출한다. 탐지된 커뮤니티가 어떤 특성으로 인해 같은 그룹으로 묶였는지 분석하는 프로파일링 과정도 함께 진행한다. 물류 분야에서 물동량이나 물류거점을 분석할 때 주로 활용하는 중심성 척도와 지역별로 세분화하여 살펴보면서, 물류 현장에 실질적으로 적용될 수 있는 의미를 찾아보고자 한다.

Ⅳ. 네트워크 분석

4.1 데이터

본 연구는 국내 택배 시장점유율이 50%에 가까운 A 택배사의 2021년 4월 한 달간의 택배 물량 데이터를 연구에 사용하였다. 원본 데이터는 배송일자, 배송품목명, 수취인주소 등으로 구성되어 있는데, 이를 상품과 지역 특성을 분석하기에 적합한 단위로 가공하였다. 연구에 활용한 데이터는 <표 4>와 같이 배송품목이 포함된 상품 카테고리나 배송된 수취인주소를 포함하고 있는 지역 구분



<그림 1> 연구방법 도식화

〈표 4〉 택배 물량 데이터 가공 예시

연월	대분류	중분류	소분류	시도	시군구	읍면동	송장건수
202104	디지털/가전	휴대폰	휴대폰액세서리	서울특별시	은평구	갈현동	761

으로 그룹화 하여 송장건수를 집계하였다.

연구 데이터는 지역 기준으로 서울특별시로 한정하였다. 서울특별시는 국내 전체 택배 물량의 21.6%를 차지하는 큰 규모이자 주민등록 인구 수 기준으로 전국의 18.7%를 차지하고 있어 대표성을 가지고 있다. 네트워크의 지역 노드는 도심 물류 센터의 배송 권역 기준과 유사한 수준인 시·군·구 단위로 지정하였다. 따라서 서울시 행정구 개수인 25개로 지역 노드가 생성된다.

상품 카테고리 기준 연구 대상은 A사가 관리하고 있는 상품 카테고리 대분류 10개를 모두 포함하였으며, 식품, 패션의류, 생활/건강, 화장품/미용, 패션잡화, 디지털/가전, 도서/음반, 가구/인테리어, 스포츠/레저, 출산/육아로 분류되고 있다. 네트워크의 노드로 대분류 10개를 사용할 경우 1개 대분류에 속해 있는 날개 상품이 최대 180개에 이르러 상품 속성의 정의가 명확하지 않다고 보았다. 반면 소분류를 노드로 할 경우 노드 개수가 900개를 넘게 되면서 속성을 부여하기에는 너무 작은 단위라고 판단하였다. 따라서 네트워크의 상품 카테고리 노드는 중분류를 기준으로 하여 128개를 생성하였다.

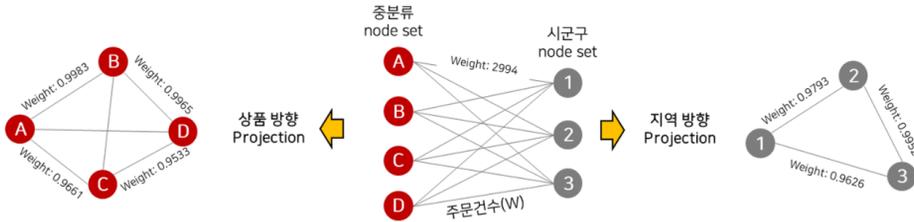
택배 물량 데이터를 기반으로 서울특별시 행정구와 상품 카테고리 노드로 구성된 네트워크를 구축하기 위해 Python의 Network X 모듈을 활용하였다. 만들어진 네트워크는 행정구 25개와 중분류 128개를 합하여 총 153개의 노드로 구성되었다. 해당 기간 동안 모든 행정구에서 모든 중분류에 대해 최소한 1건 이상의 주문이 발생했기에 링크는 3,200개가 생성되었다. 이 때 노드 간의 연결강도를 나타내는 가중치는 행정구별로 해당 상품 카테고리 중분류를 주문한 주문건수를 나타낸다.

4.2 이분 네트워크 생성

택배 물량 데이터를 활용하여 구축한 네트워크는 상품과 지역이라는 두 가지 성격의 노드 집합을 동시에 포함하고 있는데, 이를 SNA에서는 이분 네트워크(Bipartite Network)라고 한다. 이분 네트워크는 노드를 두 개의 뚜렷한 집합으로 나눌 수 있고 링크는 두 집합 사이에서만 연결되는 네트워크를 뜻한다(Huang *et al.*, 2007). 이분 네트워크 개념을 활용한 연구는 앞서 언급된 상품 네트워크 분석 외에도, 할리우드 배우와 영화 네트워크, 과학 협업 네트워크 등 다양한 분야에서 진행되어 왔다(Barabasi *et al.*, 2002). 이분 네트워크는 단일 노드 집합으로 구성된 네트워크에 비해 광범위한 정보를 담고 있기 때문에 이를 압축하기 위한 방식으로 대수학(algebra, 代數學)의 연산 중 하나인 투영(projection)을 사용한다. 투영을 통해 두 노드 집합 간의 관계가 아닌 특정 노드 집합 내의 관계 구조를 볼 수 있게 된다(Zhou *et al.*, 2007). 예를 들어 두 노드 집합을 각각 X와 Y라고 할 때, X 방향 또는 Y 방향으로 투영하여 두 개의 새로운 네트워크를 만들 수 있다. X 방향으로 투영한 네트워크에서 적어도 하나의 공통의 인접 Y 노드를 가지고 있는 경우에 두 개의 X 노드가 연결되는 방식이다(Zhou *et al.*, 2007).

4.3 투영 네트워크 생성 및 선택

앞서 택배 물량 데이터에서 서울특별시 행정구와 상품 카테고리 중분류를 두 개의 노드 집합으로 바라보고 지역-상품 이분 네트워크를 생성하였다. 다음으로는 투영을 통해 지역 네트워크와 상품 네트워크를 각각 추출하였는데 이 과정은



<그림 2> 지역-상품 이분 네트워크와 투영 처리 도식화

<그림 2>와 같이 나타낼 수 있다.

위 그림에서 중분류 node set과 시군구 node set을 연결하는 링크의 weight는, 예를 들어 1번 시군구에서 A 카테고리 상품을 주문한 건수인 2994건으로 표현할 수 있다. 상품 방향으로 투영한 네트워크의 weight는 A 카테고리 상품과 B 카테고리 상품이 각 시군구에서 얼마나 유사한 주문 패턴을 보여주는지를 나타내는 코사인 유사도 값인 0.9983으로 표현할 수 있다. 지역 방향으로 투영한 네트워크의 weight는 1번 시군구와 2번 시군구가 카테고리별로 얼마나 유사한 주문 패턴을 보여주는지를 나타내는 코사인 유사도 값인 0.9793으로 나타낼 수 있다.

택배 서비스 수요 분석은 지역 또는 상품 카테고리 노드 내에서 유사한 특성으로 묶이는 커뮤니티를 발견하는 것에 중점을 두고 있기 때문에, 투영 처리된 두 개의 네트워크 중 노드 간에 특성 차이가 보다 뚜렷하게 나타나는 네트워크를 연구에 활용하는 것이 적합하다고 판단하였다. 따라서 지역 네트워크와 상품 네트워크에 대해 각각 노드 간 유사도 분석을 추가적으로 수행하였다. 지역 네트워크의 25개 노드 간 코사인 유사도는 0.95 ~ 0.99 사이에 분포하는 것으로 나타나 서울특별시 행정구 간의 주문 패턴에는 유의미한 차이가 있다고 보기 어려웠다. 반면 상품 네트워크 128개 노드 간 코사인 유사도는 0.45 ~ 0.99 사이에 넓게 분포하였으며, 상품 카테고리 중분류를 기준으로 한다면 주문 패턴에 차이가 있는 커뮤니티들을 구분하기에 보다 적절하다고 보았다. 투영 네트워크는 원본 네트워크보다 적은 정보를 담고 있기 때문에 정보 손실을 최소화하기 위해 적합한 링크

가중치를 도출하는 것이 연구에 필수적인 과정이다(Zhou et al., 2007). 본 연구는 두 상품 카테고리 중분류 간 주문 패턴이 유사할수록 큰 가중치를 부여하여 네트워크를 구성하였고, 노드 간 코사인 유사도를 가중치로 활용하였다. 코사인 유사도 (Cosine Similarity)는 식 (1)과 같이 표현할 수 있는데, 두 벡터 간의 코사인 각도가 완전히 동일하면 1, 90도 각도를 이루면 0, 180도 반대 방향을 이루면 -1의 값을 가지는 벡터 유사도를 의미한다.

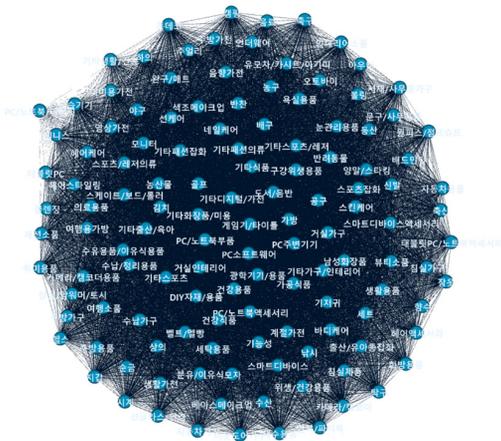
$$\begin{aligned}
 \text{similarity} &= \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1) \\
 &= \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}
 \end{aligned}$$

식 (1)에서 A, B는 상품 네트워크에서의 개별 상품 카테고리를 의미하는 벡터값이다. 상품 네트워크에서 카테고리 중분류를 하나의 벡터로 간주하고 서울특별시 행정구별 주문건수를 벡터의 원소로 본다면, 두 카테고리 중분류의 주문 패턴이 비슷하다는 것은 벡터의 코사인 유사도가 높다는 것을 의미한다. 본 연구에서는 25개 차원(dimension)을 가진 상품 카테고리 중분류 128개 벡터에 대해 <표 5>와 같이 코사인 유사도를 계산하여 네트워크 링크의 가중치로 적용하였다.

<그림 3>은 최종적으로 생성된 상품 네트워크의 시각화 결과를 보여준다. 128개의 상품 카테고리 노드 간 8,128개의 링크가 밀도 있게 서로를 연결하며, 링크의 가중치로 코사인 유사도가 부여되는 완전 네트워크의 모습을 하고 있다.

〈표 5〉 상품 카테고리 중분류 간 코사인 유사도

벡터(중분류)	벡터(중분류)	코사인 유사도
바디케어	클렌징	0.999627
구강위생용품	생활용품	0.999616
바디케어	스킨케어	0.999586
	(생략)	
PC/노트북/태블릿	기타패션잡화	0.475711
PC/노트북/태블릿	기능성	0.455672
PC/노트북/태블릿	기타생활/건강	0.453837



〈그림 3〉 이분 네트워크에서 추출한 상품 네트워크

4.4 커뮤니티 탐지

본 연구에서는 합리적인 품질과 결과를 제공하는 동시에 간편하게 사용 가능한 비중첩 커뮤니티 탐지법을 적용하였다. 이에 상품 네트워크에서 커뮤니티를 발견하는 기존 연구에서 주로 사용된 비중첩 커뮤니티 탐지법인 Girvan-Newman 알고리즘, Louvain 알고리즘, Label Propagation 알고리즘을 시행하였다(Newman, 2006; Raeder and Chawla, 2011).

Girvan-Newman 알고리즘은 가장 널리 사용되는 커뮤니티 탐지법으로, 각 링크에 대해 매개 중심성을 계산한 후 매개 중심성이 가장 높은 링크를 제거하며 커뮤니티를 탐지한다(Girvan and Newman,

2002). 남아 있는 링크에 대한 매개 중심성을 다시 계산하기를 반복하면, 링크가 제거되면서 네트워크 내의 커뮤니티가 분리된다. 이 알고리즘은 식 (2)와 같이 표현되는 모듈성(modularity) Q 가 최대화되는 지점에서 커뮤니티가 적절하게 탐지된 것으로 판단한다(Raeder and Chawla, 2011). 매개 중심성 재계산만 하면 된다는 점에서 간편하지만, 수천 개가 넘는 노드를 가진 대형 네트워크에서는 비효율적이라는 단점이 있다(Newman, 2004). 식 (2)에서 e_{ii} 는 커뮤니티 i 의 노드를 커뮤니티 i 의 다른 노드에 연결시키는 링크의 비율을 뜻하며, a_i 는 커뮤니티 i 의 노드에 연결되는 링크의 비율을 나타낸다. 즉 주어진 네트워크에서 커뮤니티 내 링크 수, 동일한 도 분포(degree distribution)를 가진 랜덤 네트워크에서 커뮤니티 내 링크 수 두 값의 차이를 측정한다.

$$Q = \sum_i (e_{ii} - a_i^2) \quad (2)$$

Louvain 알고리즘은 네트워크를 커뮤니티로 분할하는 강도를 측정하는 모듈성을 최적화하는 접근 방식이다(Blondel et al., 2008; Girvan and Newman, 2002). 모듈성은 -0.5에서 1 사이의 척도 값으로 커뮤니티 외부 대비 커뮤니티 내부 링크의 상대적인 밀도를 측정한다. 가중치가 있는 네트워크에서 모듈성 Q 는 식 (3)에 의해 산출된다.

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(C_i, C_j) \quad (3)$$

식 (3)에서 A_{ij} 는 i 와 j 노드를 연결하는 링크 사이의 가중치, $k_i = \sum_j A_{ij}$ 는 i 노드에 인접한 링크의 가중치 합계를 말한다. C_i 는 i 노드가 할당된 커뮤니티를 뜻하고, δ 함수는 i 와 j 노드가 같으면 1, 다르면 0의 값을 가지며, $m = \frac{1}{2} \sum_{ij} A_{ij}$ 를 나타낸다(Blondel *et al.*, 2008). 모듈성을 최대화하기 위해 알고리즘은 두 가지 단계를 반복한다. 첫 번째 단계는 노드를 선정하여 인접한 커뮤니티에 재배치하여 모듈성을 최적화하는 단계이다. 모듈성이 커지면 해당 노드를 그 커뮤니티에 배정하고 어떠한 변화도 일어나지 않을 때까지 이 작업을 수행한다. 두 번째 단계는 모듈성의 글로벌 최대값을 찾는 단계이다. 커뮤니티를 하나의 노드로 간주하여 셸프 루프 형태의 링크를 만들고, 커뮤니티 내부 링크의 가중치 합계 값을 이 링크의 가중치로 부여한다. 커뮤니티를 간을 연결하는 링크는 가중치를 합쳐서 하나의 링크로 하여 새로운 네트워크를 생성한다. 새롭게 만든 네트워크에 대해 첫 번째 단계부터 다시 수행하면서 모듈성이 변하지 않을 때까지 이 작업을 반복한다(Blondel *et al.*, 2008).

마지막으로 Label Propagation 알고리즘은 준지도(semi-supervised) 기계 학습 알고리즘의 하나이다. 이는 라벨(커뮤니티)이 지정되지 않은 데이터에 라벨을 할당하면서 커뮤니티를 찾아가는 방식으로, 특정 노드의 커뮤니티는 주변 노드들이 가장 많이 속해 있는 커뮤니티라는 직관적인 개념에서 출발한다. 처음 알고리즘을 시작할 때 모든 노드가 각자의 라벨을 지니고 있다. 노드의 순서를 랜덤하게 배치하여 노드의 라벨을 이웃한 노드에서 가장 빈도가 높은 라벨로 업데이트 한다. 모든 노드의 라벨이 이웃한 노드에서 가장 빈도가 높은 라벨이 될 때 알고리즘을 종료한다. 앞서 소개된 알고리즘들과 비교했을 때 모듈성을 사용하지 않는다는 차이점이 있으며, 동일한 조건에서 좀 더 다양한 커뮤니티를 찾아내는 경향이 있다(Zhu and Gharamani, 2002).

커뮤니티 탐지에 어떤 알고리즘이 적절한지를 평가하는 것은 정량적 평가와 정성적 평가가 모두 수행되어야 한다(Keziban *et al.*, 2011). 본 연구는 물류 분야에 적합한 알고리즘이 무엇인지를 밝히는 것보다도 커뮤니티 탐지를 적용하여 물류 분야 SNA 연구를 확장하는 것에 초점을 맞추고 있다. 따라서 Girvan-Newman, Louvain, Label Propagation 알고리즘의 성능에 대한 정량적 평가가 아닌 정성적 평가만을 진행하였다. 각각의 알고리즘으로 상품 커뮤니티를 탐지한 후 해석하기에 가장 용이한 알고리즘을 채택하였다.

4.5 커뮤니티 프로파일링

커뮤니티의 특성을 파악하는 프로파일링 단계에서는 먼저 커뮤니티별로 중심성 척도를 분석하였다. 중심성 척도는 물류 분야 SNA 연구에서도 자주 사용되어 왔던 개념으로, 본 연구에서는 상품 네트워크에서 특정 상품 카테고리가 다른 상품 카테고리와의 함께 주문되는 경향이 얼마나 강한지를 파악하는 기준으로 사용하였다. 중심성 척도(Centrality Measures)는 네트워크 내에서 노드의 상대적 중요성을 나타내는 척도이다. 대표적으로 연결 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 고유벡터 중심성(eigenvector centrality)이 있다(Wasserman and Faust, 1994). 이러한 중심성 척도들은 비가중(unweighted) 네트워크에 대해 정의되었기 때문에, 모든 노드와 링크는 동일한 가중치를 가진다고 가정한다(Singh *et al.*, 2020).

본 연구에 활용되는 투영 처리된 두 개의 네트워크는 모든 노드가 서로 연결된 완전 네트워크이자, 링크의 가중치에서만 차이를 보이는 가중(weighted) 네트워크이다. 따라서 단순히 노드에 연결된 링크 수(degree)를 비교하거나, 해당 노드를 경유하는 최단 경로의 수(betweenness)를 비교하거나, 모든 노드에 도달하는 속도(closeness)를

비교하거나, 중요 노드와의 연결 정도(eigenvector)를 비교하는 중심성 척도들은 본 연구에서 노드의 중심성을 판단하기에 적절하지 않다. 이에 본 연구에서는 가중 네트워크의 중심성 척도를 계산하기 위하여 중심성에 가중치를 반영한 척도를 활용하였다(Barrat *et al.*, 2004; Newman, 2004). 가중 연결 중심성 척도는 식 (4)와 같이 표현할 수 있다.

$$S_i = \sum_j^N a_{ij} w_{ij} \quad (4)$$

i 는 중심성 척도 계산 대상 노드, j 는 그 외의 모든 노드들, N 은 전체 노드의 개수, a_{ij} 는 이항 변수로 i 와 j 가 연결되어 있으면 1, 연결되어 있지 않으면 0의 값을 가지며, w_{ij} 는 연결의 가중치를 나타낸다(Barrat *et al.*, 2004).

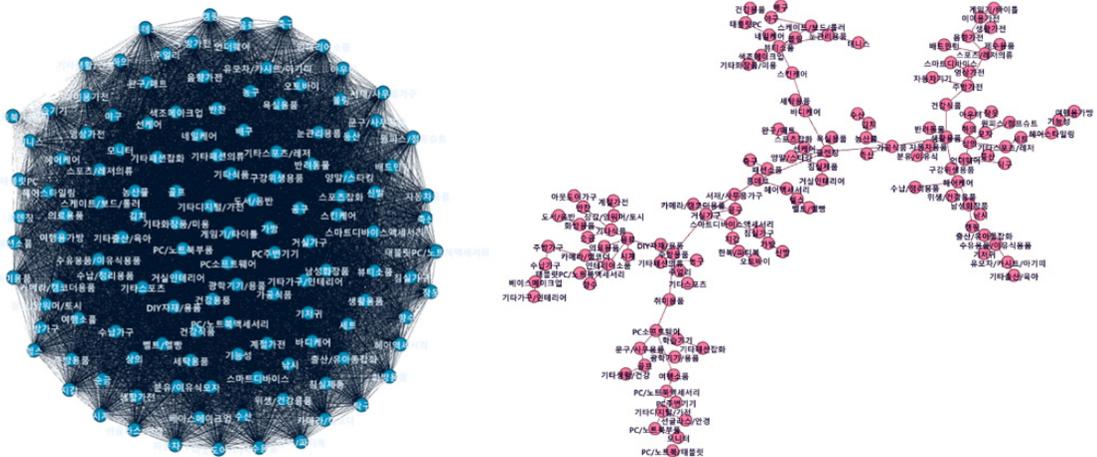
중심성 척도 계산 외에도 커뮤니티의 특성을 파악하기 위해 지역별로 상품 커뮤니티에 대한 수요가 어떻게 달라지는지에 주목하였다. 히트맵을 활용하여 상품 커뮤니티의 수요가 지역별로 어떻게 달라지는지 시각화 하였으며, 지역별 주문건수의 규모에 편차가 있음을 고려하여 데이터 스케일링을 진행하였다.

V. 연구 결과

5.1 커뮤니티 탐지 결과

상품 카테고리 중분류 네트워크(N1)와, 여기서 핵심적인 관계만을 추출한 네트워크(N2) 두 가지 네트워크에 대해 커뮤니티 탐지 결과를 비교하였다. N1은 128개의 노드 간 8,128개의 링크가 코사인 유사도를 가중치로 가지며 밀도 있게 연결되고 있다. 반면 N2는 Prim 알고리즘으로 MST(최대 신장 트리, Maximum Spanning Tree) 기법을 적용하여 핵심적인 관계만을 추출한 것으로 128개의 노드 간 127개의 링크로 연결된다. 이 때 주문 패턴의 유사도가 높은 관계만 추출하는 것이 연구의도에 적합하기 때문에 최소 신장 트리가 아닌 최대 신장 트리를 사용하였다. Gephi 프로그램을 활용하여 N1과 N2를 <그림 4>와 같이 표현하였다.

MST는 모든 노드를 순환 구조 없이 가중치 합계가 최대가 되도록 연결한 네트워크를 의미한다. MST를 적용한 네트워크로 커뮤니티를 탐지하면 정확도와 효율성 측면에서 우수하다는 연구결과가 있으며, 모듈성을 활용한 커뮤니티 탐지 알고리즘은 네트워크의 전체 링크 수에 영향을 받는다는 연구 결과도 있다(Basuchowdhuri *et al.*, 2015;



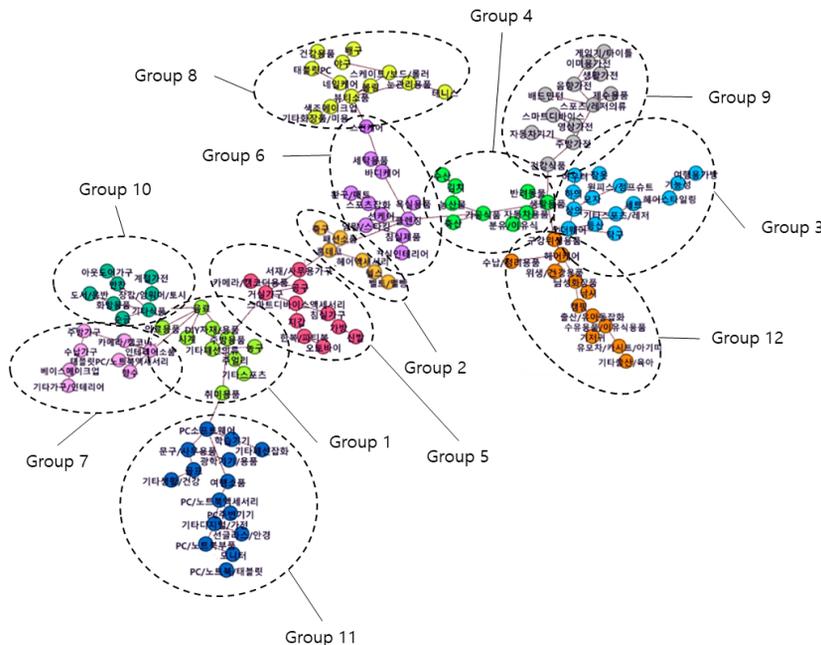
<그림 4> N1(상품 카테고리 중분류 전체 네트워크)와 N2(MST를 적용한 네트워크)

Basuchowdhuri *et al.*, 2014; Fortunato and Barthélemy, 2007). 즉 밀도가 높은 N1에 그대로 커뮤니티 탐지 알고리즘을 적용하는 것은 적절하지 않은 것이다. 상품 커뮤니티 관련 기존 연구에서도 MST 적용과 유사하게 가중치가 낮은 링크를 제거하여 소수의 고객이 주문한 상품 간의 관계를 제거하고 대다수의 고객이 선호하는 상품 간의 관계만을 남기는 방식을 사용하였다(Kim *et al.*, 2021). 실제로 N1 네트워크에 Girvan-Newman 알고리즘을 적용하여 커뮤니티를 탐지하면 서로 규모가 크게 차이 나는 2개의 커뮤니티가 발견되며, Louvain 알고리즘과 Label Propagation 알고리즘 적용 시 커뮤니티를 탐지하지 못하는 것으로 나타났다. 반면 N2 네트워크에 Girvan-Newman 알고리즘을 적용하면 2개의 커뮤니티를, Louvain 알고리즘을 적용하면 12개의 커뮤니티를, Label Propagation 알고리즘을 적용하면 20개의 커뮤니티를 도출할 수 있었다.

상품 네트워크에서 커뮤니티를 탐지할 때 적절한 커뮤니티의 수를 결정하는 기준으로 해석의 용

이성을 사용할 수 있다. 유용한 커뮤니티란 고객 행동에 대한 실질적인 통찰력을 제공할 수 있을 만큼 충분히 큰 규모이면서, 사람이 해석할 수 있을 만큼 충분히 작아야 한다(Raeder and Chawla, 2011). Girvan-Newman 알고리즘이 찾은 2개의 커뮤니티는 그 규모가 크고 커뮤니티별로 많은 상품을 포함하고 있어 상품 관계를 쉽게 해석할 수 없었다. Label Propagation 알고리즘은 커뮤니티당 포함된 상품 수가 너무 적고 균일하지 않았다. 따라서 <그림 5>와 같이 Louvain 알고리즘으로 도출된 12개의 커뮤니티 기준을 활용하였다.

탐지된 12개의 커뮤니티는 최소 6개에서 최대 16개의 카테고리 중분류를 포함하고 있다. Group 3과 같이 ‘농산물’, ‘수산’, ‘축산’ 카테고리가 동일한 커뮤니티에 묶인다는 점, Group 8과 같이 가전 관련 카테고리가 동일한 커뮤니티로 분류된다는 점, Group 10처럼 PC와 관련된 카테고리들이 하나의 커뮤니티로 묶인다는 점 등을 보았을 때, 비슷한 주문 패턴을 보이는 상품 카테고리들을 직관적



<그림 5> N2에 Louvain 알고리즘을 적용하여 커뮤니티 탐지한 결과

으로 잘 그룹화 하였음을 알 수 있다.

5.2 커뮤니티 프로파일링 결과

앞서 탐지된 12개의 커뮤니티가 어떤 특징을 가지고 있는지 중심성 및 수요 특성을 분석함으로써 택배 서비스와 도시 물류의 전략을 세우는 데 유효한 이론적 근거를 마련하고자 하였다. 먼저

중심성 파악을 위해 <표 6>과 같이 가중 네트워크의 연결 중심성 척도인 강도를 분석하였는데, 주문 패턴이 유사한 상품 카테고리 그룹 내에서도 카테고리별로 강도에 편차가 발생하였다.

편차가 비교적 큰 Group 2 커뮤니티의 경우, ‘상의’ 카테고리의 중심성 척도가 5.977로 가장 높았으며 ‘여행용가방’ 카테고리가 0.969로 가장 낮았다. ‘상의’와 ‘여행용가방’은 지역별 주문 패턴이

<표 6> 커뮤니티 내 카테고리별 중심성 척도

1		2		3		4	
category	centrality	category	centrality	category	centrality	category	centrality
음료	4.992	홈데코	3.996	상의	5.977	생활용품	5.989
주방용품	4.99	헤어엑세서리	1.997	언더웨어	2.97	가공식품	3.997
기타패션의류	2.884	헬스	1.996	모자	1.996	농산물	1.998
주얼리	1.981	패션소품	1.989	기타스포츠/레주	1.989	김치	1.998
DIY자재/용품	0.998	벨트/멜빵	0.997	아우터	1.987	자동차용품	1.992
의료용품	0.998	축구	0.99	세트	1.982	축산	0.999
시계	0.998			헤어스타일링	1.96	수산	0.999
기타스포츠	0.99			기능성	1.94	반려동물	0.997
농구	0.897			하의	0.997	분유/이유식	0.996
				원피스/점프슈트	0.997		
				잠옷	0.994		
				등산	0.993		
				탁구	0.981		
				여행용가방	0.969		
5		6		7		8	
category	centrality	category	centrality	category	centrality	category	centrality
거실가구	3.994	클렌징	5.996	인테리어소품	4.987	뷰티소품	5.981
스마트디바이스엑세서리	2.993	바디케어	2.998	주방가구	1.997	네일케어	1.994
공구	2.974	선케어	1.998	수납가구	1.995	색조메이크업	1.993
침실가구	1.992	양말/스타킹	1.998	베이스메이크업	1.985	스케이트/보드/롤러	1.987
지갑	1.991	스킨케어	1.998	향수	0.998	태블릿PC	1.986
가방	1.99	침실제품	1.998	태블릿PC/노트북 액세서리	0.997	눈관리용품	1.986
한복/파티복	1.98	스포츠잡화	1.994	카메라/캠코더	0.993	야구	1.958
카메라/캠코더용품	0.998	욕실용품	0.999	기타가구/인테리어	0.989	볼링	0.995
신발	0.996	세탁용품	0.999			기타화장품/미용	0.995
오토바이	0.986	거실인테리어	0.998			건강용품	0.991
서재/사무용가구	0.979	완구/매트	0.995			테니스	0.99
						배구	0.967

〈표 6〉 커뮤니티 내 카테고리별 중심성 척도(계속)

9		10		11		12	
category	centrality	category	centrality	category	centrality	category	centrality
스포츠/레저의류	3.974	기타식품	4.966	PC소프트웨어	3.968	구강위생용품	2.995
주방가전	3.973	장갑/압위머/토시	1.996	PC/노트북액세서리	2.983	수유용품/이유식용품	2.993
음향가전	2.985	화방용품	1.994	문구/사무용품	2.947	헤어케어	2.99
건강식품	1.997	반찬	1.976	기타디지털/가전	2.928	남성화장품	2.986
영상가전	1.989	계절가전	0.997	PC/노트북부품	2.683	출산/유아동잡화	1.994
이미용가전	1.986	도서/음반	0.997	취미용품	1.995	유모차/카시트/아기띠	1.994
생활가전	0.998	아웃도어가구	0.982	여행소품	1.989	캠핑	1.992
게임기/타이틀	0.995	순금	0.976	학습기기	1.938	수납/정리용품	0.998
스카트디바이스	0.994			골프	1.922	기저귀	0.997
배드민턴	0.994			PC주변기기	0.996	기타출산/육아	0.997
계수용품	0.988			광학기기/용품	0.991	위생/건강용품	0.994
자동차기기	0.983			모니터	0.978	납시	0.992
				기타생활/건강	0.955		
				선글라스/안경	0.954		
				기타패션잡화	0.951		
				PC/노트북/태블릿	0.723		

유사하여 같은 커뮤니티로 분류되지만, 그 중에서도 특히 ‘상의’는 다른 카테고리들과 함께 주문되는 경향이 강한 카테고리이며, ‘여행용가방’은 단독으로 주문되는 경향이 강한 카테고리라고 해석할 수 있다.

반면 카테고리별 강도의 편차가 크지 않은 Group 11 커뮤니티의 경우, ‘구강위생용품’ 카테고리의 중심성 척도가 2.995로 가장 높았고, ‘납시’ 카테고리는 0.992로 가장 낮게 나타났다. Group 11의 경우 Group 2에 비해 다른 카테고리들과 함께 주문되는 경향이 강한 카테고리가 존재하지는 않는다고 볼 수 있다.

다음으로 상품 수요 특성을 파악하기 위해, 지역별로 나누어서 상품 커뮤니티별 주문건수를 확인하였다. 그 결과 모든 지역에서 대량의 주문이 발생하는 상품 커뮤니티가 존재하였다. ‘음료’, ‘주방용품’, ‘기타패션의류’ 등이 포함된 Group 0, ‘상의’, ‘언더웨어’, ‘모자’ 등이 포함된 Group 2, ‘생활용품’, ‘가공식품’, ‘농산물’ 등이 포함된 Group 3가 대표적이었다. 식품, 의류, 생활용품은

지역별 특성에 관계없이 모든 지역에서 수요가 높다는 것을 의미한다.

VI. 결 론

6.1 요약

본 연구는 택배 물량 데이터를 활용하여 상품과 지역 관점에서 택배 서비스 수요를 분석하였다. SNA 를 활용하여 상품 카테고리 노드와 지역 노드로 구분되는 이분 네트워크를 구성하였으며, 투영 처리를 통해 상품 네트워크와 지역 네트워크를 추출하여 연구를 진행하였다. 상품 네트워크는 노드 간 코사인 유사도가 넓게 분포하여 커뮤니티 탐지에 용이하다는 점에서 택배 서비스 수요 연구에 보다 적합한 것으로 나타났다. 상품 네트워크의 가중치는 코사인 유사도를 활용하였고, MST로 간소화된 네트워크에 Louvain 커뮤니티 탐지 알고리즘을 적용한 후 프로파일링을 진행하였다. 그 결과 네트워크 내에서 지역별 주문 패턴이 유사한

상품 커뮤니티 12개를 발견했으며, 지역별로 특색 있는 상품 수요를 찾을 수 있었다. 각 상품 커뮤니티 내에서 중심성이 높아 다른 상품 카테고리와의 함께 주문될 가능성이 높은 상품 카테고리 또한 찾을 수 있었다.

6.2 이론적 함의

본 연구의 이론적 함의는 다음과 같다. 본 연구는 소셜 네트워크 분석(SNA) 방법을 국내 택배 서비스 수요를 상품 및 지역 관계 관점으로 세분화하여 실증하였다. 택배와 관련된 많은 선행 연구들은 주로 택배 서비스의 전체 수요 합계의 값을 예측하는 모형 연구에 집중되어 있었다(박인선, 이상현, 2018; 이우승, 2004; 전현주, 유명식, 2020; 최형림 등, 2008). 본 연구는 전체 택배 서비스 수요를 지역과 상품으로 세분화하여, 지역별로 어떤 상품 수요가 많이 발생하는지를 실증하기 위해 항공, 항만 등 기존 물류 분야에서도 제한적으로 사용되던 SNA를 기반으로 지역-상품 간의 관계를 분석하였다.

특히, 김용희(2020)의 연구에 따르면, 네트워크 분석에서 사용되는 다양한 분석 알고리즘에 따라서 분석 결과와 해석이 달라질 수 있다고 하였다. 본 연구에서도 지역-상품 네트워크 간의 관계를 분석하기 위해 적합한 SNA 분석 알고리즘을 정의하기 위해, Louvain, Girvan-Newman, Label Propagation, Girvan-Newman, Label Propagation 총 다섯 가지 알고리즘을 분석하였다. 첫째, Louvain 알고리즘은 노드의 수가 비교적 균일한 12개의 커뮤니티를 탐지하여 해석의 용이성 측면에서 적절했다. 둘째, 큰 단위로 커뮤니티를 찾고 싶을 때는 Girvan-Newman 알고리즘을 사용하는 것이 적절했다. 마지막으로, 세분화된 커뮤니티를 찾고 싶을 때는 Label Propagation 알고리즘을 사용할 수도 있을 것으로 보인다. 그 외, Girvan-Newman 알고리즘은 2개의 커뮤니티를 찾아내어 커뮤니티 1개 당 많은 상품이 포함되면서 해석이 쉽지 않았고,

Label Propagation 알고리즘은 20개의 커뮤니티를 찾아내어 커뮤니티 1개당 적은 수의 상품이 포함되면서 통찰력을 얻기에 적절하지 않았다. 즉, 본 연구에서는 SNA를 기반으로 택배 서비스 수요를 분석하여 지역(커뮤니티)에 특성화 된 상품 수요가 있음을 도출하고, SNA 방법론과 상세 분석 알고리즘 중 Louvain, Girvan-Newman, Label Propagation들이 상품 네트워크 분석에 활용도가 높다는 점을 확인하였다. 이에 본 연구에서는 Girvan-Newman, Louvain, Label Propagation 알고리즘에 대한 정성적 평가를 통하여 상품 커뮤니티 해석에 용이한 알고리즘을 찾아냈다는 함의를 가진다.

6.3 실무적 함의

본 연구는 택배사 관점에서 택배 물량 데이터의 활용 방향성과 비즈니스적 가치를 제시하였다. 전자상거래 서비스를 직접 제공하는 판매채널에 의존하지 않고서도, 택배사 자체적으로 보유하고 있는 데이터를 활용하여 장바구니 분석이나 교차판매 분석 등 고객의 수요를 면밀히 파악할 수 있음을 보여주었다. 그 동안 택배사는 규모의 경제 중심의 운송 네트워크를 담당하는 공급자의 역할이 주를 이루었다. 향후 택배 물량 데이터의 비즈니스적 가치에 주목하여 포장, 보관, 수요예측 전략을 마련하고 주문관리부터 배송까지 전반을 담당하게 된다면, 택배사들이 전자상거래 시장에서 하나의 플랫폼으로서 주도권을 획득할 수 있을 것이다.

물류센터 관점에서는 여러 물류센터에 상품 및 재고를 분산 배치하는 전략에 시사하는 바가 있었다. 본 연구에서는 지역별 주문 패턴이 비슷한 상품 카테고리 그룹, 즉 상품 커뮤니티를 도출하였다. 주문 패턴이 비슷한 상품 카테고리를 같은 물류센터에 보관한다면 묶음 배송의 가능성이 증가하여 물류비용을 절감할 수 있을 것이다. 또한 모든 지역에서 높은 수요를 보이는 상품 커뮤니티도 있었던 반면, 특정 지역에서 주문이 두드러지는 커뮤니티도 나타났다. 이는 다른 지역과 구분되는

해당 지역만의 특별한 수요 패턴으로 해석된다. 도시 물류 및 중소형 물류센터는 공간의 한계로 인해 상품 선정이나 재고 수준을 결정하는 것이 중요하다. 따라서 지역별로 수요가 두드러지는 상품 커뮤니티를 찾아서 해당 지역을 커버하는 물류센터에 집중 배치하는 전략을 고려해볼 만하다.

제한된 SKU 별 차별화된 관리 전략도 응용성이 높다. 상품 네트워크에 대한 중심성 척도 분석 결과, 동일한 커뮤니티 내에서도 상품 카테고리마다 다른 상품 카테고리들과 함께 주문되는 경향이 각각 다르다는 것을 발견하였다. 즉, 중심성 척도가 높은 상품 카테고리는 재고를 충분히 확보하고, 중심성 척도가 낮은 카테고리는 SKU 품목 수를 확대하여 다양성을 확보하는 전략 등을 생각해볼 수 있다. 중심성 척도를 물류센터의 재고 보관비용을 산정할 때 활용하는 등 비즈니스적으로도 충분히 연결 가능하다. 이러한 상품 카테고리 간의 관계성은 단순한 상품별 수요예측이나 클러스터링 기법으로는 얻을 수 없는 의미 있는 결과다.

마지막으로 전자상거래 서비스 및 플랫폼 관점에서는 그들의 기술 역량을 물리적인 공간과 오프라인으로 확장할 수 있는 가능성을 제시하였다. 전자상거래 분야에서 지속적으로 발전해왔던 장바구니 네트워크, 상품 네트워크 분석의 연구 결과는 협업 필터링 기반 상품 추천이나 고객 성향 분석 등 온라인 상에 적용되는 데 머물렀다. 향후에는 전자상거래 주문 데이터를 상품-고객 관점에서 분석하는 것에서 한 단계 나아가, 상품-고객-지역(배송지)으로 확장함으로써 기술의 적용 범위를 오프라인으로 넓힐 수 있을 것으로 본다. 최근 많은 전자상거래 서비스 및 플랫폼들이 풀필먼트에 역량을 집중하고 있다는 점에서, 오프라인으로의 기술력 확장은 중요한 경쟁력이 될 것이다.

6.4 향후 연구 제언

본 연구에 기반한 하기의 향후 연구가 기대된다. 우선 연구에 사용한 택배 물량 데이터는 전국

단위가 아닌 서울특별시에 한정되어 있어 국내 택배 서비스 수요 분석으로 일반화하기 어려운 측면이 있다. 도시가 아닌 도서산간, 농촌 지역 등으로 데이터의 지역적 범위가 확대된다면, 상품 카테고리 간의 주문 패턴에 보다 뚜렷한 차이점이 드러날 수 있을 것으로 보인다. 또한 지역 네트워크의 노드 간 유사성 차이가 좀 더 커지는 결과를 얻을 가능성도 있다.

사용된 택배 물량 데이터가 특정 한 해의 1개월치 데이터라는 점에서 물량의 추세, 계절성, 주기성과 같은 시계열 패턴을 반영하지 못했다는 한계가 있다. 기존 연구에 따르면 택배는 계절적 특성에 따른 월별 물량 변화가 심하기 때문에(서상범 등, 2015), 향후 택배 서비스 수요와 관련된 SNA 연구에서는 시계열분석 방법을 접목해볼 필요가 있다.

또한 택배 물량 데이터에 포함된 상품 카테고리 및 지역 분류만 가지고 연구를 진행했기 때문에, 사회적인 이슈나 계절적인 요인 등 택배 물량 및 물류 인프라에 영향을 주는 다양한 변수들을 고려하지 못하였다. 따라서 연구 결과를 물류 현장의 재고 분산 배치 등에 즉시 적용하기에는 한계가 있다. 향후 다른 변수들을 가중치로 하여 SNA 모델에 반영한다면 정교한 택배 서비스 수요 연구가 가능할 것으로 보인다.

마지막으로 분석 결과를 기반으로 지역별 수요 제품 커뮤니티가 수요 지역의 특성을 반영한다는 사실을 확인할 수 있었다. 예를 들어 대형 디지털 상품 상업지구가 있는 '용산구'의 경우 'PC소프트웨어', 'PC/노트북액세서리', '문구/사무용품' 등이 포함된 상품 커뮤니티인 Group 10의 주문건수가 다른 상품 카테고리에 비해 높게 나타났다. 반면 '강남구'나 '서초구'와 같은 주거지구는 '클렌징', '바디케어', '선케어' 등이 포함된 상품 커뮤니티인 Group 5의 주문건수가 다른 상품 커뮤니티 대비 높은 결과가 나왔다 (<Appendix 1>). 이로써 지역별 특성 변수를 함께 고려한 SNA 분석을 통하여 제품 커뮤니티를 나누는 새로운 요인을 발견

하고 이를 물류 시스템 개선 전략 연구가 가능할 것으로 보인다.

참 고 문 헌

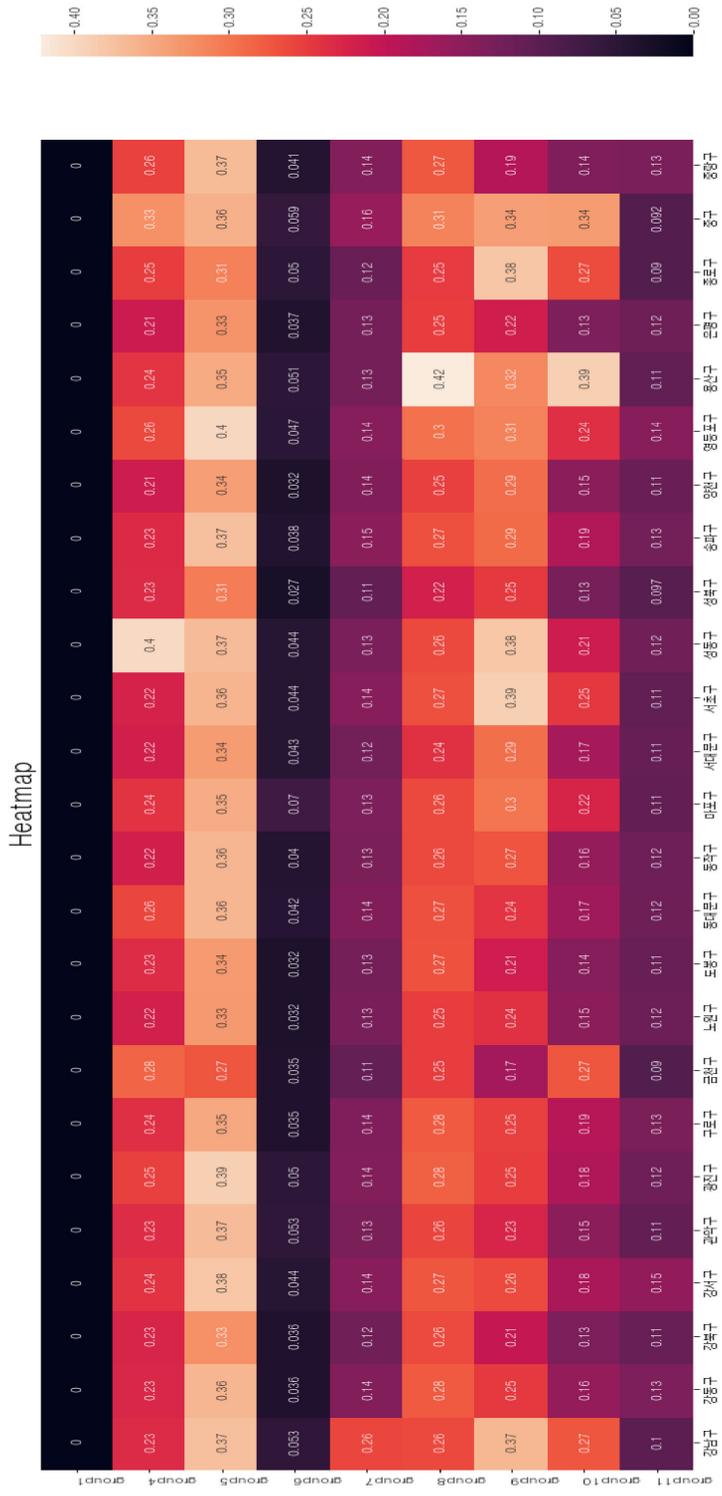
- [1] 강성경, 유환, 이영재, “텍스트 마이닝과 소셜 네트워크 분석을 이용한 재난대응 용어분석”, *Information Systems Review*, 제18권, 제1호, 2016, pp. 141-155.
- [2] 김벽진, 안승범, “SNA를 이용한 국제 항공화물의 네트워크 구조 분석”, *로지스틱스 연구*, 제25권, 제3호, 2017, pp. 59-82.
- [3] 김용희, “소셜 네트워크 분석(Social Network Analysis)기법의 이해와 적용: 네트워크 구조와 클러스터링 그리고 QAP”, *KIPA 조사포럼*, 제34권, 2020, pp. 58-68
- [4] 김태현, 윤승진, 송상화, “라스트마일 배송을 위한 도심 물류 네트워크 설계: 서울시 사례 연구”, *한국SCM학회지*, 제20권, 제1호, 2020, pp. 110-120.
- [5] 노정명, “택배물량 5년만에 85.7% 급증…상위 5사 92% 장악 ‘과점’ 심화”, *소비자경제*, 2021.
- [6] 박인선, 이상헌, “우리나라 택배 물동량 예측 모형에 대한 연구”, *경영교육연구*, 제33권, 제1호, 2018, pp. 47-65.
- [7] 박한영, 김소형, 정승주, 서상범, “코로나19 확산에 따른 국내 택배서비스 이용행태 변화”, *교통연구*, 제28권, 제2호, 2021, pp. 55-64.
- [8] 서상범, 이지선, 김용진, “택배산업의 지역간 물동량 분석을 통한 생활지원형 물류서비스 개선방안 연구”, *한국교통연구원 기본연구보고서*, 2015, pp. 1-195.
- [9] 손재철, 김벽진, 안승범, “항공우편물 네트워크의 중심성 분석에 관한 연구”, *로지스틱스 연구*, 제29권, 제1호, 2021, pp. 29-38.
- [10] 이양수, 김미이, 김신태, 이영훈, “택배 산업의 효율적 물류 네트워크 개선 방향에 관한 연구”, *한국SCM학회지*, 제5권, 제2호, 2005, pp. 43-52.
- [11] 이우승, “전자상거래 촉진요인이 택배서비스 시장에 미치는 영향 분석”, *경제연구*, 제22권, 제2호, 2004, pp. 167-193.
- [12] 임병학, “컨테이너항만 네트워크가 항만 생산성에 미치는 영향에 대한 연구: 사회 네트워크 분석을 중심으로”, *로지스틱스 연구*, 제19권, 제3호, 2011, pp. 19-35.
- [13] 전현주, 유명식, “물동량 예측을 위한 비정형 데이터 활용에 관한 연구”, *한국통신학회 학술대회논문집*, 2020, pp. 373-375.
- [14] 정승주, “2015년 물류정책 연구방향과 과제”, *월간교통*, 제203권, 2015, pp. 43.
- [15] 최형림, 김현수, 박병주, 강무홍, “운송스케줄 정보수집 에이전트 기반 반복운송 경로계획 시스템”, *Information Systems Review*, 제10권, 제1호, 2008, pp. 115-133.
- [16] 한국교통연구원, “2021년 교통정책 연구방향과 과제”, *월간교통*, 제275권, 2021, pp. 36-41.
- [17] Barabási, A. L., H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations”, *Physica A: Statistical Mechanics and its Applications*, Vol.311, No.3-4, 2002, pp. 590-614.
- [18] Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks”, *Proceedings of the National Academy of Sciences*, Vol.101, No.11, 2004, pp. 3747-3752.
- [19] Basuchowdhuri, P., R. Roy, S. Anand, D. R. Srivastava, S. Majumder, and S. K. Saha, “Spanning tree-based fast community detection methods in social networks”, *Innovations in Systems and Software Engineering*, Vol.11, No.3, 2015, pp. 177-186.
- [20] Basuchowdhuri, P., S. Anand, D. R. Srivastava, K. Mishra1, and S. K. Saha, “Detection of communities in social networks using spanning tree”, *Advanced Computing, Networking and Infor-*

- matics*, Vol.2, 2014, pp. 589-597.
- [21] Blondel, V. D., J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No.10, 2008, P10008
- [22] Butts, C. T., "Social Network Analysis: A methodological introduction", *Asian Journal of Social Psychology*, Vol.11, No.1, 2008, pp. 13-41.
- [23] Carter, C. R., L. M. Ellram, and W. Tate, "The use of social network analysis in logistics research", *Journal of Business Logistics*, Vol.28, No.1, 2007, pp. 137-168.
- [24] El-adaway, I. H., I. S. Abotaleb, and E. Vechan, "Social network analysis approach for improved transportation planning", *Journal of Infrastructure Systems*, Vol.23, No.2, 2017, 05016004.
- [25] Fortunato, S. and M. Barthélemy, "Resolution limit in community detection", *Proceedings of the National Academy of Sciences*, Vol.104, No.1, 2007, pp. 36-41.
- [26] Girvan, M. and M. E. J. Newman, "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences*, Vol.99, No.12, 2002, pp. 7821-7826.
- [27] Huang, Z., D. D. Zeng, and H. Chen, "Analyzing Consumer-Product Graphs: Empirical Findings and Applications in Recommender Systems", *Management Science*, Vol.53, No.7, 2007, pp. 1146-1164.
- [28] Kim, H. K., J. K. Kim, and Q. Y. Chen, "A product network analysis for extending the market basket analysis", *Expert Systems with Applications*, Vol.39, No.8, 2012, pp. 7403-7410.
- [29] Kim, Y., T. Y. Choi, T. Yan, and K. Dooley, "Structural investigation of supply networks: A social network analysis approach", *Journal of Operations Management*, Vol.29, No.3, 2011, pp. 194-211.
- [30] Lee, P. D., "Measuring Supply Chain Integration: A Social Network Approach", *Supply Chain Forum: An International Journal*, Vol.6, No.2, 2005, pp. 58-67.
- [31] Newman, M. E. J., "Analysis of weighted networks", *Physical Review E*, Vol.70, No.5, 2004, 056131.
- [32] Newman, M. E. J., "Fast algorithm for detecting community structure in networks", *Physical Review E*, Vol.69, No.6, 2004, 066133.
- [33] Newman, M. E. J., "Finding community structure in networks using the eigenvectors of matrices", *Physical Review E*, Vol.74, No.3, 2006, 036104.
- [34] OrmanVincent, G. K., V. Lanayut, and L. Cherifi, "Qualitative comparison of community detection algorithms", *Digital Information and Communication Technology and Its Applications*, 2011, pp. 265-279.
- [35] Raeder, T. and N. V. Chawla, "Market basket analysis with networks", *Social Network Analysis and Mining*, Vol.1, No.2, 2011, pp. 97-113.
- [36] Russell, G. J. and A. Petersen, "Analysis of cross category dependence in market basket selection", *Journal of Retailing*, Vol.76, No.3, 2000, pp. 367-392.
- [37] Sarwar, B., G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce", In *Proceedings of ACM E-commerce 2000 Conference*, Minneapolis, MN, USA.
- [38] Singh, A., R. R. Singh, and S. R. S. Iyengar, "Node-weighted centrality: A new way of centrality hybridization", *Computational Social Networks*, Vol.7, No.6, 2020, pp. 1-33.
- [39] Wang, Y. and Y. Li, "Complexity analysis on the influence factors of the flight delay risk based on SNA", *Open Journal of Social Sciences*, Vol.8, No.5, 2020, pp. 54-71.

- [40] Wasserman, S. and K. Faust, *Social network analysis: Methods and applications*, Cambridge University Press, 1994.
- [41] Watts, D. J. and S. H. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, Vol.393, No.6684, 1998, pp. 440-442.
- [42] Wu, J., Y. Wang, S. Shafie, and D. Zhang, "Discovery of associated consumer demands: Construction of a co-demanded product network with community detection", *Expert Systems with Applications*, Vol.178, 2021, 115038.
- [43] Zhou, T., J. Ren, M. Medo, and Y. Zhang, "Bipartite network projection and personal recommendation", *Physical Review E*, Vol.76, No.4, 2007, 046115
- [44] Zhu, X. and Z. Ghahramani, "Learning from Labeled and Unlabeled Data with Label Propagation", *Technical Report*, Vol.2, No.107, 2002.

〈Appendix〉

〈Appendix 1〉 지역별(서울특별시 25개 행정구) 상품 수요 특성
 (12개의 커뮤니티 중 모든 지역에서 수요가 높은 3개의 커뮤니티 제외)



Information Systems Review

Volume 24 Number 4

November 2022

Delivery Service Demand Analysis Using Social Network Analysis (SNA)

Kyungeun Oh^{*} · Sulim Kim^{**} · HanByeol Stella Choi^{***} · Heeseok Lee^{****}

Abstract

The transition to a non-face-to-face consumer society has rapidly occurred since Covid-19. The need for a subdivided urban logistics policy centered on courier delivery, a life-friendly last-mile logistics service, has been raised. This study proposes a SNS-based method that can analyze the demand relationship by region and product, respectively. We extend the market basket network (MBN) and co-purchased product network (CPN), find product category patterns, and confirm regional differences by using delivery order data. Our results imply that SNA analysis can be effectively applied to inventory distribution or product (SKU) selection strategies in urban logistics.

Keywords: *Delivery, Courier, Parcel, Demand Analysis, Community Detection, Social Network Analysis (SNA)*

* MS, Master of Information Management, KAIST

** Ph.D Candidate, College of Business, KAIST

*** Corresponding Author, Assistant Professor, Department of Management Information Systems, Myongji University

**** Professor, College of Business, KAIST

◎ 저 자 소 개 ◎



오 경 은 (okeee0000@kaist.ac.kr)

KAIST 경영대학에서 석사학위를 취득하였으며, SSG닷컴을 거쳐 현재 네이버 쇼핑에 재직 중이다. 이커머스, 물류, 풀필먼트, 라스트마일과 관련된 플랫폼 프로덕트 매니지먼트 업무를 담당하고 있다.



김 수 림 (sulimkim@kaist.ac.kr)

KAIST 경영대학 박사과정에 재학 중이다. 미국 Washington State University에서 Finance로 학사학위를 취득하였으며, 서울대학교 경영대학에서 경영정보학과 석사학위를 취득하였다. KPMG에서 IT컨설턴트로 재직한 경력이 있다. 주요 관심분야는 인플루언서 마케팅, IT 애플리케이션의 효과성 및 사용자 행태, 전자정부 등이다.



최 한 별 (hbschoi@mju.ac.kr)

명지대학교 경영대학 경영정보학과 조교수로 재직 중이다. KAIST 경영대학 박사학위 취득 후 KAIST 경영대학 디지털혁신연구센터 연구조교수로 재직하는 경력이 있다. 주요 관심분야는 프라이버시, 정보 보안, 공유 경제, 정보 시스템의 사회적 효과 및 비즈니스 애널리틱스 등이다. 지금까지 Journal of Management Information Systems, Decision Support System, Security Journal 등 주요 학술지에 논문을 발표하였다



이 희 석 (hsl@kaist.ac.kr)

KAIST 경영대학 교수로 재직 중이다. 아리조나대학 경영학 박사 취득 후 네브라스카 대학 교수를 역임하였다. 주요 관심분야는 IT Strategy와 Digital Innovation이다. MIS Quarterly, Journal of Management Information Systems, Information and Management 등 주요 학술지에 논문을 발표하였다.

논문접수일 : 2022년 04월 04일

게재확정일 : 2022년 05월 30일

1차 수정일 : 2022년 05월 23일