

The Usage Characteristics of Twitter, and Their Relationship with Gender, Age, and Brand Preferences

Hyung Jun Ahn*

Abstract

With the increasing popularity of social network services (SNSs), there have been many attempts to analyze the users of SNSs. By doing so, the characteristics and preferences of the users can be understood, which can help companies provide personalized information and services that they need or are relevant for them. This study aimed to analyze the usage behavior of Korean Twitter users from various perspectives to deepen the understanding of it. For this research goal, an online survey was conducted for the users of Twitter and the data about their actual usage were collected using the open API of Twitter. Factor analysis of the data revealed five factors that explain about 69.3% of the usage variables. It was also investigated how the factors are related to gender, age, and brand preferences. The results showed that the usage behavior of Twitter is largely affected by age ($p < 0.001$), and also by gender through an interaction effect ($p < 0.05$). Also, the factors showed significant statistical correlations with the brand preferences of the users.

▶ Keyword : SNS, Twitter, brand preference, user characteristics

I. Introduction

인터넷 환경은 2000년대 중반의 Web 2.0 시대를 지나 이제 소셜 및 스마트 모바일 시대에 이르렀다. 특히 많은 사용자들이 다양한 기능과 특성을 가진 소셜 네트워크 서비스(Social Network Services; SNS)에 접속한 채 매일 상당한 시간을 보내고 있으며, 이는 사회적 교류 활동뿐 아니라 뉴스 및 정보 공유, 쇼핑 등 매우 다양한 활동들을 수행하는 데 활용된다[1]. SNS는 현재도 점점 성장해 가는 추세이며 사회 및 경제의 다방면에서 그 중요성은 아무리 강조해도 지나치지 않는다고 할 수 있다.

개인들에게 SNS가 다양한 편익과 가치를 제공해 주는 반면, 기업들에게는 SNS를 통해 고객들을 만나고 소통해야 하는 과

제가 있다. SNS는 그 특성상 네트워크 관계를 통해 매우 다양한 상호작용을 가능케 하며, 풍부한 기능이 있고, 사용자들이 직접 원하는 종류의 정보를 취사선택할 수 있도록 해 주기 때문에 SNS에서 전통적 미디어에서와 같이 일방향적이며 무차별적인 고객관리를 수행하는 것은 매우 비효율적이라고 할 수 있다. 기업들로서는 SNS 상에서 고객들을 잘 이해하는 것이 무엇보다 중요하며 이를 바탕으로 개인화 및 차별화된 고객관리 활동을 수행해야 할 필요가 있다. 이렇게 함으로써 주로 사용자들의 구매 및 평점 정보에 의존해 왔던 협업필터링(collaborative filtering)과 같은 개인화 기법을 확장시킬 수도 있다[1, 2].

이와 관련하여 SNS 상에서 사용자의 특성을 이해하고자 하는 연구들이 최근에 많이 수행되고 있다. 페이스북이나 트위터

• First Author: Hyung Jun Ahn, Corresponding Author: Hyung Jun Ahn
*Hyung Jun Ahn(hjahn@hongik.ac.kr), School of Business, Hongik University
• Received: 2016. 01. 14, Revised: 2016. 02. 03, Accepted: 2016. 02. 22.

의 사용 특성을 바탕으로 개인의 성향을 추정하는 연구들이 그 좋은 예라고 할 수 있다[3, 4]. 그러나 이런 연구들은 주로 영어권 사용자들을 대상으로 수행되어 왔으며, 특히 그러한 SNS 사용 특성을 구성하는 기본 요인들이 무엇인지, 그러한 요인이 사용자들의 인구통계학적 특성이나 제품 선호도와 관련되어 있는지 등에 대한 연구들은 아직 미흡한 실정이다.

따라서 본 연구에서는 사용자들의 SNS 활용 특성이 사용자 개개인의 차별화된 특성과 선호를 반영한다고 가정하고 사용자들의 트위터 이용 특성에 대한 다양한 분석을 통해 그에 대한 이해의 폭을 넓히고자 하였다. 이를 위해 국내 트위터 이용자에 대한 설문을 실시하고 응답자들의 트위터 사용 정보를 공개 API(Application Programming Interface)를 통해 수집하였다. 수집한 데이터의 요인분석을 통해 트위터 사용 특성의 주요 요인들을 추출하고, 응답자들의 나이 및 성별과 이러한 요인과의 관계를 분석하였다. 더불어 트위터 사용 특성과 사용자들의 브랜드 선호도 간의 상관관계를 분석하였다.

본 논문의 구성은 다음과 같다. 2절에서는 관련 연구를 간략히 소개한다. 3절은 연구 목적과 연구 방법을 설명한다. 4절은 분석 결과를 제시한다. 5절은 토론 및 결론을 제시한다.

II. Review of Related Studies

SNS는 단순한 인터넷 커뮤니티로서의 역할에만 머무르지 않고 인터넷 상의 다양한 정보와 서비스를 총 망라하는 방향으로 진화하고 있다. 사용자들에게는 SNS가 필요한 정보, 기회 및 사회적 교류의 공간을 제공한다는 점에서, 기업들에게는 소비자들을 가장 밀접하게 만나고 다양한 상호작용을 수행할 수 있다는 점에서 SNS의 중요성은 갈수록 커지고 있다.

SNS와 관련된 최근 연구들은 사용자들의 SNS의 활용 방식이나 SNS에 공유하는 내용이 사용자들의 다양한 특성을 반영함을 확인하였다. 즉, 사용자의 고유한 특징, 취향 등이 SNS 이용 방식에 드러난다는 것이다. 그러한 사용자의 특징은 사용자가 SNS의 다양한 기능을 활용하는 구체적인 방식, 친구를 맺는 방식이나 혹은 친구의 수, SNS에 공유하는 내용의 종류, SNS 상의 텍스트에 포함된 어휘 등에 드러날 수 있다. 예를 들어 [5]의 연구에서는 SNS 사용과 관련된 다양한 특징, 특정 주제나 감정에 관련된 어휘, 네트워크 연결 구조 등을 분석함으로써 사용자들의 정치적 성향과 인종 등을 추정할 수 있음을 보였다. 또 다른 연구에서는 주로 SNS 상의 텍스트에 포함된 감정어를 활용하여 사용자들의 특정 브랜드에 대한 태도를 분석하였다 [6]. 이는 텍스트 마이닝 분야의 감정 분석과 유사한 접근 방식이라고 할 수 있다[7]. Kandias 등의 연구에서는 YouTube에 올라온 다양한 사용자 데이터를 활용하여 사용자가 얼마나 신뢰성 있는지를 분석하기도 하였다[8].

이와 같은 응용 연구들과 더불어 더 근본적인 사용자들의 특

성, 즉 사용자의 성격 유형을 SNS를 통해 파악하고자 하는 연구들도 존재한다. 예를 들어 [3]의 연구에서는 페이스북의 사용 특징을 분석하여 사회적 어휘, 즉 모임이나 애정 등을 표현하는 어휘의 활용이 외향성과 관련이 있고, 반면 컴퓨터나 인터넷과 관련된 어휘들의 사용이 내향성과 관련이 있음을 보였다. 또 다른 연구에서는 부정적인 성격 특징과 일련의 감정 단어, 문장 부호 등의 상관관계를 보였다[4]. 이 외에도 최근에 SNS와 성격 유형의 관계를 분석한 연구들이 적지 않게 등장하고 있다[9, 10].

이와 같이 SNS를 통해 사용자를 이해하게 되면 기업들은 이를 활용하여 더 나은 정보나 서비스를 사용자들에게 제공할 수 있다. SNS를 비롯한 온라인상에는 항상 너무 많은 서비스, 제품 등이 존재할 수 있는데, 이는 사용자들에게 정보 과잉의 문제를 일으키고 기업들에게는 효과적인 마케팅 활동을 수행하기 어렵게 만든다. 이때 사용자에 대한 이해가 정교할수록 기업들은 효과적으로 개인화된 상호작용을 수행할 수 있다. 예를 들어 [11]에서는 SNS 데이터를 활용하여 사용자들에게 영화, 책, 음악 등을 효과적으로 추천할 수 있음을 보였다[11]. 또 다른 연구에서는 사용자의 SNS 정보를 기타 상황 정보, 즉 날씨, 교통 정보 등과 결합하여 유용한 위치 기반 정보를 제공할 수 있음을 보였다[12]. Ma의 연구에서는 SNS에서 도출할 수 있는 암묵적 연결 정보, 즉 추정된 친구 관계 정보를 활용하여 영화 추천의 정확도를 높일 수 있음을 보였다[13].

이와 같이 SNS는 사용자에 대한 이해를 높여주어 기업들로 하여금 효과적인 마케팅 활동을 수행할 수 있게 해 준다. 이는 기존의 협업필터링과 같은 개인화 연구의 확장이라고 할 수 있다[14]. 그러나 이와 같은 연구들은 미국 등 영어권 국가의 SNS 사용자를 대상으로 수행된 경우가 많으며, 국내에서는 아직 관련연구가 부족한 편이다. 본 연구에서는 이와 같은 기존 연구들을 참고하여 국내 트위터 사용자들의 사용 특성을 분석하며, 사용 특성이 나이, 성별, 브랜드 선호도 등과 유의한 관계를 갖는지에 대해 탐색적인 연구를 수행하였다.

III. Research Objectives and Methodology

1. Objectives and methodology

본 연구의 목적은 국내의 트위터 사용자들의 이용 특성에 대한 다양한 분석을 통해 이에 대한 이해의 폭을 넓히는 것이다. 구체적으로는 트위터 사용 특성을 설명하는 소수의 핵심 요인들을 도출하고, 성별 및 나이에 따른 사용 특성의 차이를 분석하며, 사용 특성과 브랜드 선호 간의 관계를 파악하고자 하였다.

본 연구의 구체적인 연구 방법은 다음과 같다.

첫째, 국내 트위터 사용자들을 대상으로 나이, 성별, 특정 브랜드들에 대한 선호도를 설문하며, 이와 동시에 트위터가 제공하는 API를 활용하여 응답자들의 트위터 활용 특징에 대한 데이터를 수집한다.

둘째, 수집한 변수들을 요인분석하여 소수의 주요한 사용 형태 요인들을 추출한다. 이를 통해 전체 데이터를 설명할 수 있는 간결한 모형을 구축한다.

셋째, 응답자들의 성별 및 나이가 추출된 트위터 사용 형태 요인들과 어떤 관계를 갖는지 분석한다.

넷째, 추출된 요인들과 브랜드 선호도 간의 관계를 분석한다. 이를 통해 트위터 사용 특성이 브랜드 선호도의 예측에 활용될 수 있는지 분석한다.

2. Data collection

본 연구의 대상은 SNS를 활발히 사용하는 국내의 20 ~ 30대로 하였다. 트위터와 같은 SNS의 경우 모집단의 분포를 잘 반영하는 연구 표본을 수집하기 어려운 것으로 알려져 있다. 트위터의 공개 API가 제공하는 공개 스트림(public stream)에서 무작위로 사용자 표본을 선택하는 경우 특정 시점의 특정 사건이나 이벤트에 관심이 있는 사용자로 표본이 편향될 수 있고, 동시에 트윗(Tweet) 메시지 작성을 많이 하는 적극 사용자층으로 표본이 편향될 수 있다[15]. 네트워크의 특정 노드(node)들에서 출발하여 샘플을 확장해 가는 스노우볼(snowball) 표본 방식의 경우[16], 시작 지점이 되는 사용자에 따라 표본의 특성이 많이 달라질 수 있으며, 특정 유명인 혹은 기업의 트위터 계정을 시작점으로 할 경우 마찬가지로 그들에 대한 선호 특성에서 기인하는 편향이 발생할 수 있다. 특히 본 연구에서는 트위터 사용 특성과 브랜드 선호도에 대한 관계 또한 분석하고자 하였으므로 그러한 표본 방식은 적합하지 않다고 판단하였다.

따라서 본 연구에서는 주로 20~30대의 대학 재학생, 졸업생 등 SNS 주 사용 계층이 많이 분포해 있으며 특정 브랜드에 대한 선호와는 독립적인 것으로 볼 수 있는 국내 대학들의 트위터 계정의 Follower들을 분석 대상으로 하였다. 인터넷을 대상으로 하는 연구에서 인터넷의 주 사용 계층인 대학생들을 표본으로 하는 방식이 널리 사용되는데, 본 연구의 표본 방식도 이와 유사하다고 볼 수 있다[17, 18]. 설문은 2014년 10월부터 약 1개월간 Google Docs를 이용하여 온라인으로 진행하였다. Follower들 중 스팸이나 상업적 목적의 계정은 제외하였고, 또한 휴면 계정을 배제하기 위해 조사 시점에서 2개월 이내에 트윗 메시지가 있는 계정들만을 설문 대상으로 하였다. 온라인 설문을 통해 나이, 성별 정보와 더불어 젊은 층에게 인지도가 높은 총 8개의 캐주얼 패션 브랜드에 대한 선호도 정보를 5점 척도로 수집하였다. 또한 설문에 응답한 참가자들을 대상으로 트위터가 제공하는 공개 API를 이용하여 트위터 활용 특성 정보를 수집하였다. 이를 위해 Python 프로그래밍을 활용하였다. 총 167명의 사용자에 대한 데이터를 수집하였으며, 사용자 당 최대 200개까지의 게시글을 수집하였다. 200개 이하의 글만

게시한 사용자의 경우에는 전체 게시글을 분석 대상으로 하였다.

3. Analysis of emotion words and emoticons

트위터 데이터 중 감정 어휘 추출을 위해서는 감정어 사전 및 형태소 분석기를 활용하였다. 감정어 사전 구축을 위해서는 기존 연구에서 제시된 감정 어휘 목록을 활용하고[19], 이를 포털 사이트 사전의 유의어 검색을 통해 확장하였으며, 결과적으로 약 1,400여개의 감정 어휘를 분석에 활용하였다. 형태소 분석기는 오픈 소스 소프트웨어인 꼬꼬마 형태소 분석기를 활용하였다[20]. 형태소 분석기는 어절에서 의미 단위들을 추출해 줌과 동시에 각각의 품사를 지정해 주기 때문에 이를 활용하여 트윗 메시지 내의 감정 어휘들을 집계하였다.

이모티콘 처리를 위해서는 SNS에서 많이 활용되는 이모티콘들을 정규 표현식(regular expression) 방식으로 정의하여 집계하였다. 이를 통해 ‘^_’, ‘^__^’, ‘^____^’ 등과 같이 유사하지만 다르게 표현되는 이모티콘인 경우에는 동일하게 검색되도록 하였다. 이러한 방식으로 ‘:’, ‘πππ’, ‘ㅋㅋ’, ‘OTL’ 등 총 50 종의 이모티콘을 분석에 활용하였다.

IV. Analysis Results

1. Descriptive statistics of Twitter usage

Table 1은 조사된 자료의 기술통계를 보여주고 있다. 총 457명에게 트위터의 멘션(Mention) 기능을 이용해 설문 요청을 보냈으며, 이 중 167명이 응답하여 응답률은 37%로 비교적 높게 나타났다. 응답자들의 계정에서 수집한 총 트윗 메시지의 개수는 26,358개로 이를 분석에 활용하였다.

응답자 중 여성이 98명으로 약 59%를 차지하고 있으며, 남성은 69명으로 약 41%를 차지하고 있다. 트위터 사용자의 성별 분포에 대한 정확한 통계는 알려져 있지 않으나, 조사에 따라 여성 사용자가 약 40~60% 사이로 나타나고 있으며[21], 본 연구 대상의 성별 분포도 그와 같은 범위에서 벗어나지 않는다고 할 수 있다. 응답자 나이의 평균은 약 27세, 표준편차는 8.04로 나타났다. 구체적으로 20대는 전체의 59%, 30대가 21%를 차지해서 대부분의 응답자가 20~30대에 분포하고 있음을 알 수 있었다. 그러나 남성 응답자와 여성 응답자의 나이 평균은 각각 31.6세와 23.8세로 상당한 차이를 보이고 있었다. 트위터를 사용하는 정도는 응답자 간에 큰 차이를 보이고 있었다. 트위터 사용 정도에 비례한다고 볼 수 있는 Following, Follower, 트윗 수, Faved, List 등의 통계를 보면 최대치와 최소치가 평균과 큰 차이가 나며, 특히 표준편차도 평균에 비해 매우 큰 것을 알 수 있다. 한편 사용자들의 멘션 횟수는 트윗 당 평균 0.44회, Hash Tag 사용은 0.36회, URL을 포함하는

경우는 0.26회로 나타났다.

한편 트윗의 텍스트 관련 변수들을 살펴 보면, 긍정 감정어와 부정 감정어가 각각 트윗 당 평균 0.27, 0.25회 사용된 것으로 나타나 비슷한 비율로 사용되고 있음을 알 수 있었다. 이모티콘의 경우 긍정 이모티콘이 평균 0.28회, 부정 이모티콘이 평균 0.21회 사용되어서, 긍정 이모티콘의 사용 비중이 더 높음을 알 수 있었다. 문장 부호는 트윗 당 평균 2.34회 사용된 것으로 나타났다.

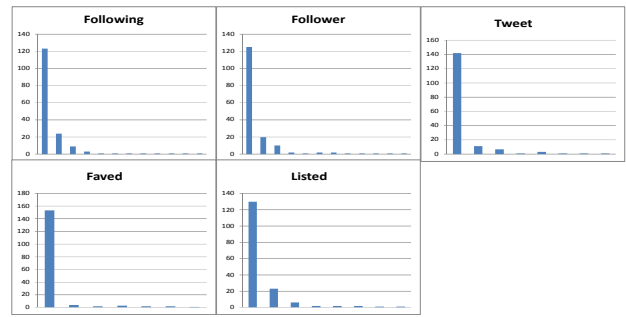


Fig. 1. Skewed Distribution of the Variables

Table 1. Descriptive Statistics

Variables	Min	Max	Average	Std. Dev.
Gender	Female 98명 Male 69명		-	-
Age	15	67	27.04	8.04
Following	18	36948	947.16	3549.42
Followers	6	34074	895.09	3322.47
Tweets	64	574526	18315.65	52209.6
Faved	0	20618	514.78	1816
List	0	90	7.01	12.81
Mentions	0	1.66	0.44	0.31
Hash Tag	0	1.95	0.36	0.48
URL	0	1.08	0.23	0.28
Positive emotion words	0.01	1.32	0.27	0.18
Negative emotion words	0	0.86	0.25	0.15
Numbers	0.02	2.57	0.34	0.37
Positive emoticons	0	2.02	0.28	0.26
Negative emoticons	0	1.33	0.21	0.25
Punctuations	0.35	9.69	2.34	1.30

이때 Fig. 1을 보면 일부 변수들, 즉 Following, Follower, 트윗 수, Faved 수, List 수 등이 매우 편향된 분포를 보이고 있음을 알 수 있다. 실제 이 변수들의 SPSS 분석에 의한 첨도(kurtosis) 및 왜도(skewness) 값은 각각 3.8 이상, 17.6 이상으로 매우 높게 나타나고 있다. 따라서 이후 분석에서의 설명력을 높이기 위해 해당 5개 변수에 대해서는 10을 밑으로 하는 로그 변환을 수행하였다. 이 결과는 Fig. 2에 나타나 있다. 변환 후 5개 변수들의 첨도는 모두 1.74 이하, 왜도는 0.85 이하로 양호하게 나타났다. 따라서 변형된 변수들을 활용하면 해당 변수의 변화에 따른 다른 변수들과의 관계를 더욱 명확하게 파악할 수 있다. 본 연구의 향후 분석에서는 해당 변수들에 대해 모두 로그 변형 후의 값을 사용하였다.

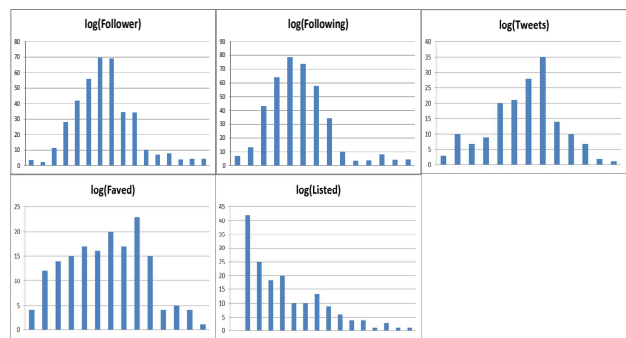


Fig. 2. Variables after Logarithmic Transformation

2. Factor analysis of Twitter usage variables

앞서 수집한 16개의 변수들은 그 수가 많고 상당수가 상호 간 높은 상관관계를 보이고 있다. 따라서 이들을 소수의 공통 요인(factor)으로 축소하여 그러한 요인들만으로 16개의 변수들을 잘 설명할 수 있다면 이를 통해 현상에 대한 간결한 모형을 구축할 수 있다. 이를 위해 수집된 16개의 변수들에 대해 요인분석을 실시하였다. 구체적으로는 카이제 정규화를 이용한 VARIMAX 회전을 통해 주성분 분석을 실시하였으며, 고유 값(eigenvalue) 1을 기준으로 요인 수를 결정하였다. 해당 데이터가 요인분석에 적합한지를 테스트하는 KMO 검정 및 Bartlett 검정 결과 각각 0.68 및 $p < 0.001$ 로 나타나 충분히 기준을 충족함을 확인하였다. 분석 결과 Table 2과 같이 총 5개의 요인을 추출하였으며, 이 다섯 요인이 모든 16개 변수들의 변량의 69.3%를 설명함을 확인할 수 있었다. 또한 각 변수와 요인과의 상관관계로 간주할 수 있는 요인적재량(factor loading)을 기준으로 변수들을 구분하였을 때(Table 2의 굵은 글씨) 모든 변수들의 요인적재량이 0.5 이상으로 보통 요인분석에서 요구되는 기준을 충족하는 것으로 나타났다.

Table 2. Factor Analysis of Twitter Usage Features

	Factors				
	1	2	3	4	5
Follower	0.946	0.069	0.070	0.030	-0.019
Following	0.880	-0.010	-0.052	0.038	0.060
List	0.849	-0.114	0.079	-0.006	0.007
Tweets	0.564	0.289	0.284	0.506	-0.040
Positive emoticons	0.019	0.798	0.159	-0.080	-0.088
Mentions	0.194	0.744	-0.111	-0.019	0.162
Negative emoticons	-0.216	0.522	0.017	0.373	0.026
Positive emotion words	0.040	-0.022	-0.727	-0.199	0.064
Negative emotion words	-0.082	-0.190	-0.650	0.015	-0.278
Hash Tag	0.136	-0.420	0.558	-0.313	-0.025
URL	0.239	-0.538	0.545	-0.253	0.043
Faved	0.112	-0.026	0.000	0.887	0.069
Punctuations	0.058	0.134	-0.086	-0.010	0.884
Numbers	-0.060	-0.144	0.428	0.124	0.704
누적 변량(%)	20.70	35.80	49.20	59.30	69.30

참고로, 5개의 변수에 대해 로그 값을 취하지 않은 상태로 요인분석을 실시했을 때는 고유 값 1을 기준으로 총 6개의 요인이 추출되어 약 72.6%의 변량을 설명하는 것을 알 수 있었다. 즉, 로그를 취한 경우가 1개 더 적은 요인의 수로 비슷한 크기의 변량을 설명할 수 있기 때문에 분석 대상 문제를 더 간결하게 설명할 수 있는 모형이라고 간주할 수 있다. 따라서 본 연구의 나머지 부분에서도 모두 Table 2의 5요인 모형을 활용하였다.

각 요인에 대해 공통적으로 높은 적재량을 보이는 변수들은 상관관계가 높은 변수들이라고 할 수 있다. 따라서 이러한 변수들을 살펴봄으로써 각 요인이 의미하는 바를 추측할 수 있다. 첫째 요인에는 Follower, Following, List, 트윗 수 등이 포함되어 있다. 이러한 변수들의 값은 대체로 트윗터를 오래, 활발하게, 많이 사용할수록 커진다고 볼 수 있으므로 첫 번째 요인은 트윗터 사용의 정도를 의미한다고 추측할 수 있다.

두 번째 요인에는 이모티콘과 멘션이 포함되어 있다. 즉, 이모티콘과 멘션을 많이 사용할수록 이 요인의 값은 커지게 된다. 트윗터에서 멘션은 다른 사용자의 트윗을 언급하거나 또는 사용자 간의 대화에 사용되며, 특히 이모티콘을 많이 동반하는 경우에는 개인적인 친밀한 대화일 가능성이 크다. 따라서 두 번째 요인은 사적이며 친근한 대화를 의미한다고 볼 수 있다. 두 번째 요인으로 구분되지는 않았으나 Hash Tag 및 URL 변수와도 높은 음의 상관관계를 보이는 점도 이러한 해석을 뒷받침한다. 즉, Hash Tag나 URL이 포함된 메시지가 공개적 정보 공유의 목적에 많이 사용됨을 감안하면, 반대로 사적인 대화는 이들과 음의 관계를 갖는 것이 자연스럽다고 할 수 있다.

세 번째 요인은 감정어와 Hash Tag, 그리고 URL의 사용과 관련되어 있으며, 이때 감정어와는 음의 적재량을 보이고 있다. 따라서 감정어를 사용하지 않을수록, Hash Tag와 URL을 많이 포함할수록 세 번째 요인의 값은 커지게 된다. Hash Tag는 보통 자신의 메시지에 검색을 용이하게 하는 특정 키워드를 추가할 때, URL은 외부에 있는 정보 소스에 연결할 때 사용된다는 점을 가정하면, 세 번째 요인은 대체로 감정적이지 않은 공개적 정보 제공 및 공유 용도의 트윗터 사용을 나타냄을 알 수 있다. 또한 이 요인이 정보의 객관성을 나타낼 수 있는 숫자 변수와 높은 요인적재량을 보이는 점도 이러한 해석을 뒷받침한다.

네 번째 요인의 경우는 단일 변수인 'Faved' 수와 관련이 있다. 이 요인은 트윗 수와는 비례하지만 Following 및 Follower 수와는 상관관계가 높지 않은 것으로 나타났다. Faved는 주로 다른 사용자의 트윗에 호감을 표시하거나, 혹은 트윗터에서 발견된 유용한 정보를 향후 참고하고자 저장할 때 사용된다. 따라서 네 번째 요인은 그와 같은 용도의 사용 성향을 설명하는 요인이라고 할 수 있다.

마지막 요인은 문장 부호 및 숫자 변수와 높은 상관관계를 보이고 있다. 또한 이 요인은 다른 변수들과는 모두 매우 낮은 상관관계를 보여주고 있다. 이를 통해 이 요인은 숫자와 구두점을 포함한 보다 객관적인 정보 및 사실을 표현하는 트윗들과 관련이 있음을 추측할 수 있다. Table 3에는 위와 같은 해석들에 대한 요약이 정리되어 있다.

그러나 이러한 해석은 요인분석 결과에 나타난 제한적인 정보에 바탕을 둔 것으로, 이를 검증하기 위해서는 각 해석에 해당하는 추가적인 자료를 수집하여 분석할 필요가 있으며, 이는 본 연구의 범위에는 포함되지 않는다. 따라서 이러한 해석의 결과는 잠정적인 것이며, 본 연구의 나머지 부분의 이해를 위한 참고의 용도로 활용되어야 한다.

Table 3. Interpretation of the 5 Factors

Factor	Highly loaded variables	Proposed interpretation
1	Follower, Following, Tweets, List	The level of Twitter use
2	Emoticons (positive and negative), mentions, Hash Tag(-), URL(-)	Personal and friendly messages
3	Emotion words (-) (positive and negative) Hash Tag, URL	Public sharing or provision of information
4	Faved, Tweets	Showing preference or collecting information
5	Numbers, Punctuations	Presenting objective information or facts

3. Gender, age, and Twitter usage behavior

다음으로, 성별 및 나이와 트윗터 사용 특성의 관계에 대한 분석을 실시하였다. 이를 위하여 앞서 추출된 5개 요인을 활용하였다. 첫째, 나이와 5개 요인의 상관관계 분석에 대한 결과는 Table 4에 나타난 바와 같다. 이 결과에 의하면 나이는 요인 1 및 요인 2와 유의한 상관관계를 보이고 있다. 구체적으로 요인 1과는 양의, 요인 2와는 음의 상관관계를 보이고 있다. 즉, 앞

서 5개 요인에 대한 해석에 의하면 나이가 많을수록 트위터 사용 정도는 증가하며, 반면 트위터를 사적 대화에 사용하는 정도는 감소하는 것을 확인할 수 있다. 이러한 차이는 Fig. 3을 통해서도 확인할 수 있다. Fig. 3은 나이의 중간 값(median)인 25를 기준으로 응답자를 두 그룹으로 나눈 후 비교한 그래프로, 요인 1과 요인 2에서 큰 차이를, 요인 4에서도 작은 차이를 확인할 수 있다. 다만 요인 4의 차이는 Table 4에서 확인할 수 있듯이 유의하지 않게 나타났다.

Table 4. Age and the 5 Factors

	Factor				
	1	2	3	4	5
Correlation	.435**	-.235*	-0.107	-0.121	0.14

** p < 0.01

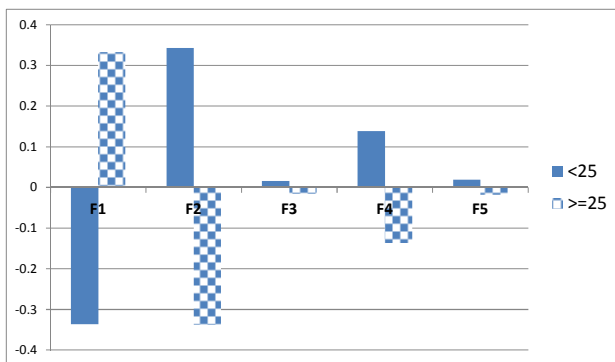


Fig. 3. 5 Comparison of the 5 Factors by Age

다음으로, 성별과 5요인과의 관계에 대한 분산분석(ANOVA) 결과가 Table 5에 나와 있다. 이 결과에 의하면 성별은 요인 1, 요인 2, 요인 4에 유의한 영향을 끼치는 것을 알 수 있다. Fig. 4는 이 차이를 그래프로 보여주고 있다. 이에 따르면 여자에 비해 남자 응답자의 요인 1의 값이 높게 나타나고 있다. 즉, 남자의 트위터 사용 정도가 전반적으로 높음을 알 수 있다. 요인 2의 경우 여자가 남자보다 높게 나타나, 개인적인 친밀한 용도의 트위터 활용은 여자 응답자에게서 더 많이 발견되는 것을 알 수 있다. 요인 4의 경우도 여자 응답자가 더 높게 나타나서 여자 사용자들이 'Favorite' 기능을 훨씬 더 많이 사

용하는 것을 알 수 있다. 그 외, 요인 3과 요인 5의 경우 성별 간 차이가 유의하지 않았다.

Table 5. Gender and the 5 Factors (ANOVA)

	F value	p value
Factor 1	16.801	0.000***
Factor 2	5.047	0.026**
Factor 3	1.191	0.277
Factor 4	6.956	0.009**
Factor 5	0.242	0.623

** p < 0.01, *** p < 0.001

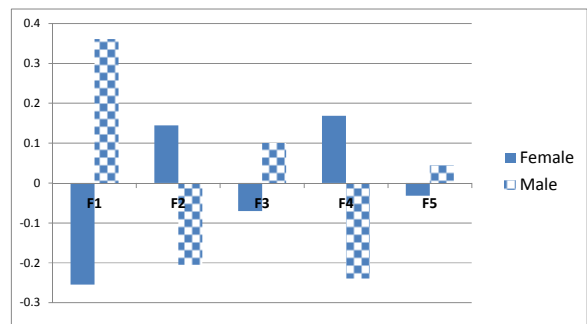


Fig. 4. Comparison of the 5 Factors by Gender

그러나 이와 같은 차이는 앞서 기술통계에서 살펴본 바와 같이 남성 응답자의 평균 나이가 여성 응답자보다 크기 때문일 수도 있다. 이에 따라 나이와 성별, 그리고 나이와 성별의 상호작용 효과를 고려한 회귀분석을 각 요인에 대해 실시하였다. 그 결과는 Table 6과 같다. 첫째, 나이와 성별을 동시에 고려했을 때 성별의 효과는 대체로 유의하지 않으며, 대신 나이의 효과가 요인 1과 요인 2에 대해 매우 유의한 것으로 나타났다. 나이는 두 요인에 대해 모두 p < 0.001의 수준에서 유의한 것을 확인할 수 있으며, 회귀 모형의 F값도 역시 각각 유의한 것으로 나타났다. 또한 요인 1의 경우 R² 값이 0.229로 나타나 요인 1의 변량의 상당부분인 약 23%가 회귀모형에 의해 설명됨을 알 수 있었다. 즉, 이 결과를 통해 나이는 트위터 사용 정도에 매우 큰 영향을 끼침을 확인할 수 있었다.

Table 6. Regressions analysis for Age, Gender, and the 5 Factors

	Gender		Age		Gender×Age(Interaction)		F	R ²
	β	t	β	t	β	t		
Factor 1	.102	1.303	.482	5.298***	-.159	-1.953	16.120***	0.229
Factor 2	-.032	-.385	-.374	-3.833***	.219	2.513*	6.853***	0.112
Factor 3	.165	1.871	-.166	-1.623	-.009	-.100	1.677	0.03
Factor 4	-.168	-1.920	-.092	-.908	.060	.665	2.597	0.046
Factor 5	-.030	-.333	.150	1.463	-.021	-.226	0.916	0.017

Note: * p < .05, *** p < .001

추가로, 요인 2의 경우 나이와 성별의 상호작용 효과 또한 유의한 것으로 나타나고 있다. 이 결과는 성별에 따라 나이의 영향 정도가 달라질 수 있음을 의미한다. 이를 그래프로 나타내면 Fig. 5와 같다. Fig. 5는 위의 요인 2에 대한 회귀분석의 결

과를 나이를 X축으로, 요인 값을 Y축으로 하여 보여주고 있으며, 이때 나이의 범위는 평균을 중심으로 표준편차를 빼고 더한 만큼으로 하였다. 이 그림에 따르면 20대 초반 여성의 경우 트위터를 친밀한 개인적 대화의 목적으로 쓰는 경향이 뚜렷함을 알 수 있다. 그러나 이러한 경향은 남성, 여성 모두 나이에 따라 줄어들음을 알 수 있다. 그러나 나이에 따른 변화는 남성이 훨씬 원만함을 알 수 있고, 이는 나이와 성별의 상호작용 효과에 의한 것으로 볼 수 있다.

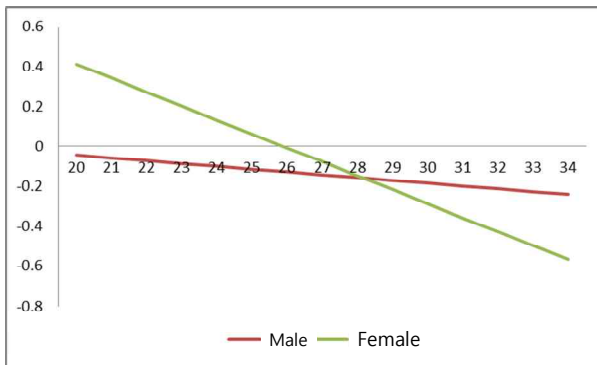


Fig. 5. Interaction Effect between Age and Gender
Note: X - Age, Y - Factor 2

4. Twitter usage and brand preference

다음으로, 트위터 사용 특성의 5개 요인과 브랜드 선호도 간의 관계를 분석해 보았다. 이는 앞서 분석한 5개의 요인들이 개인적 특성을 반영하는 경우 그러한 요인들을 제품, 서비스, 브랜드 등의 선호도 추측에도 활용될 수 있기 때문이다. 따라서 그러한 가능성을 검증해 보기 위해 5개의 요인과 8개의 패션

브랜드에 대한 선호도 간의 상관관계 분석을 실시하였다. 이때 사용된 8개의 브랜드는 국내의 젊은 층에게 인지도가 높은 다양한 종류의 캐주얼 브랜드들로 선정하였으며, 각각에 대한 간략한 설명은 Table 7과 같다.

Table 7. Description of the 8 Brands

Brand	Description
A	Well-known global SPA brand. Characterful style.
B	Global high class casual brand. Classic style.
C	Well-known global SPA brand. Basic and practical style.
D	Global casual brand. Unique and fitting style.
E	Korean high class casual brand. Classic style.
F	Global casual brand. Well known for its jeans in Korea. Characterful style.
G	Global budget casual brand. Basic and practical style.
H	Global high class casual brand. Mostly sports style.

Table 8의 상관관계 분석을 보면 총 5개의 유의한 상관관계를 관찰할 수 있다. 특히 요인 2가 2개의 브랜드와 유의한 상관관계를 보였으며, 한편 요인 4의 경우는 유의한 상관관계를 발견할 수 없었다. 브랜드 측면에서는 총 4개의 브랜드에서 유의한 상관관계가 관찰되었다. 이러한 결과를 종합해 보면 트위터 사용 특성은 브랜드 선호도의 예측에 어느 정도 유용하게 사용할 수 있을 것으로 보이나, 동시에 그 예측력은 제한적임을 알 수 있다.

Table 8. Correlation between the 5 Factors and Brand Preference

	A	B	C	D	E	F	G	H
Factor 1	-0.002	-0.003	-0.127	0.072	-0.136	-0.041	-.178*	-0.083
Factor 2	-0.033	0.116	-0.009	0.064	.290**	0.115	0.125	.170*
Factor 3	-.166*	-0.059	0.024	0.018	0.028	-0.066	0.037	0.057
Factor 4	0.029	-0.07	-0.003	-0.064	-0.069	-0.141	-0.031	-0.118
Factor 5	0.09	0.133	0.058	.270**	0.129	0.064	0.073	0.066

* p < 0.05, ** p < 0.01

V. Discussion and Conclusion

1. Summary and discussion

본 연구에서는 국내의 트위터 사용자들의 SNS 사용 특성에 대한 다양한 분석을 통해 그에 대한 이해를 높이고자 하였다. 구체적으로는 사용 특성을 설명하는 핵심 요인을 도출하고, 성별 및 나이에 따른 사용 특성의 차이를 분석하며, 사용 특성과 브랜드 선호도 간의 상관관계를 파악하고자 하였다.

이를 위해 국내 트위터 사용자들을 대상으로 설문을 실시하였으며 응답자들의 트위터 이용 정보를 공개 API를 통해 수집하여 분석하였다.

분석 결과는 다음과 같이 요약할 수 있다. 요인분석 결과, 총 16개의 트위터 사용 특성 변수를 5개의 요인으로 축소할 수 있었으며, 추출된 5개의 요인은 전체 변량의 약 69%를 설명하였다. 각 요인들은 서로 뚜렷하게 구분되는 변수 군들과 높은 요인적재량을 보여주었다. 이러한 변수 군들에 근거하여

각 요인들의 잠재적 의미를 추측할 수 있었다. 첫째 요인은 전반적인 트위터 사용의 정도, 둘째 요인은 사적이며 친밀한 대화, 셋째 요인은 공적인 정보 공유, 넷째 요인은 선호의 표현이나 정보의 수집, 그리고 다섯째 요인은 객관적인 정보의 표현 등을 의미하는 것으로 해석되었다.

다음으로, 각 요인들은 성별 및 나이에 따라 뚜렷한 차이를 보이는 것으로 나타났다. 특히 요인 1과 요인 2는 성별 및 나이에 따라 큰 차이를 보이고 있었다. 그러나 성별과 나이를 동시에 고려한 회귀분석 결과, 성별의 영향은 제한적인 것으로 나타났으며, 대신 나이에 따른 요인 1과 요인 2의 차이가 매우 뚜렷하였다. 요인 1의 경우 나이의 효과가, 요인 2의 경우 나이의 효과와 더불어 나이와 성별 간의 상호작용 효과가 유의한 것으로 관찰되었다. 대체로 나이가 많을수록 트위터의 전반적 사용 정도는 증가하고, 반면 트위터를 사적 친밀한 대화를 위해 사용하는 정도는 줄어드는 것을 관찰할 수 있었다. 더불어 20대 초중반의 여성의 경우에는 트위터를 사적인 대화의 용도로 사용하는 경향이 뚜렷하였다.

그 다음, 각 요인들과 브랜드 선호도 간의 상관관계를 분석하였다. 분석 결과, 총 8개 브랜드 중 5개에 대해 유의한 상관관계를 발견할 수 있었다. 그러나 일부 브랜드에 대해서는 유의한 관계를 관찰할 수 없었으며, 각 요인들의 브랜드 선호도에 대한 예측력 또한 제한적이라고 볼 수 있었다.

본 연구의 의의는 다음과 같다. 첫째, 그 중요성에도 불구하고 국내 SNS 사용자, 즉 한국어 사용자를 대상으로 한 SNS 사용 특징에 대한 연구가 아직까지는 매우 부족하다. 본 연구는 국내 트위터 사용자들을 대상으로 그 사용 특성을 분석하고 주요한 요인들을 추출하여 그에 대한 이해를 높이고자 하였다. 이는 점에서 첫 번째 의의가 있다. 영어의 경우 그 범용성으로 인해 관련된 많은 연구 및 다양한 분석 도구가 존재하지만 한국어의 경우에는 그렇지 못하다. 아직까지 국내 SNS 활용에 대해서는 매우 제한적인 연구만 수행되어 왔으며 활용할 수 있는 연구 방법 및 도구 또한 매우 미흡한 실정이다. 이러한 관점에서 본 연구는 한국어를 대상으로 SNS의 사용자 분석 응용 연구를 수행하고 연구 방법에 대한 사례를 제시했다는 점에서 그 의의가 있다. 둘째, 분석 결과, 연령대에 따른 트위터 사용 특징의 뚜렷한 차이를 관찰할 수 있었다. 이는 연령대별 사용자들의 특정한 SNS에 대한 인식 및 사용 목적이 매우 다를 수 있음을 의미한다. 이는 국내 SNS 사용자를 이해하는데 큰 도움을 줄 수 있다. 셋째, 본 연구의 분석 결과, 트위터 사용 특징들이 일부 브랜드의 선호도와 유의한 상관관계를 보임을 확인할 수 있었다. 이는 트위터 사용 특징이 소비자의 특성을 반영함을 의미하며, 그러한 특징을 파악함으로써 더 나아가 소비자의 선호도를 추정할 수 있음을 의미한다. 이제까지 개인화에 대한 연구들은 대체로 협업필터링 방식의 응용 및 확장에 초점을 맞춰왔다. 협업필터링은 사용자들의 제품에 대한 평가 정보 혹은 구매 정보를 활용하는 방식이기 때문에, SNS의 경우처럼 제품 평가 혹은 구매 정보의 획득이 명시적

으로 가능하지 않은 상황에서는 적용하기 힘들다. 또한 협업필터링 방식은 기본적으로 전자상거래 사이트들을 그 대상으로 하기 때문에 더욱 다양한 방식의 사회적 상호작용이 이루어지는 소셜 미디어에 적용하는 것은 무리가 있다. 따라서 본 연구는 국내의 기업들이 SNS를 분석하여 소비자를 더 잘 파악하고 서비스 및 정보 추천 등의 다양한 개인화 서비스를 발전시켜 나가는 데 활용할 수 있는 결과를 제시했다는 점에서 의의가 있다.

2. Limitations and issues for future study

본 연구의 한계는 다음과 같다. 첫째, 트위터 및 SNS 연구의 특성상 전체 모집단을 대표하는 표본을 추출하기는 어렵다. 본 연구에서도 주로 20대, 30대로 구성된 응답자들을 대상으로 데이터를 수집하였다. 따라서 본 연구 결과를 다른 인구통계학적 특성을 지닌 집단으로 일반화 할 때에는 주의가 필요하다. 둘째, 본 연구에서 제시한 트위터 사용의 5개 요인에 대한 해석의 객관적 타당성을 입증하기 위해서는 이에 대한 설문을 포함한 추가 연구가 필요하다. 셋째, 본 연구는 트위터 사용자들을 그 대상으로 하였다. 따라서 트위터와 다른 특징 및 기능을 가진 SNS의 경우에는 그 차이점들을 반영한 분석 및 응용 방식이 필요하며, 이에 대한 향후 연구가 필요하다. 넷째, SNS 사용 행태는 시간 등 다양한 환경 변수에 따라 변화할 수 있다. 따라서 본 연구의 결과는 고정적인 것으로 간주할 수 없으며 그러한 다양한 변수들에 따른 변화가 연구될 필요가 있다.

REFERENCES

- [1] P. R. Berthon, L. F. Pitt, K. Plangger, and D. Shapiro, "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy," *Business Horizons*, vol. 55, no. 3, pp. 261-271, May 2012.
- [2] L. R. Men and W.-H. S. Tsai, "How companies cultivate relationships with publics on social network sites: Evidence from China and the United States," *Public Relations Review*, vol. 38, no. 5, pp. 723-730, Dec. 2012.
- [3] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach," *PLoS ONE*, vol. 8, no. 9, p. e73791, Sep. 2013.
- [4] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting Dark Triad Personality Traits from Twitter

- Usage and a Linguistic Analysis of Tweets,” in 2012 11th International Conference on Machine Learning and Applications (ICMLA), 2012, vol. 2, pp. 386–393.
- [5] M. Pennacchiotti and A.-M. Popescu, “A Machine Learning Approach to Twitter User Classification,” *ICWSM*, vol. 11, pp. 281–288, 2011.
- [6] M. M. Mostafa, “More than words: Social networks’ text mining for consumer brand sentiments,” *Expert Systems with Applications*, vol. 40, no. 10, pp. 4241–4251, Aug. 2013.
- [7] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [8] M. Kandias, V. Stavrou, N. Bozovic, L. Mitrou, and D. Gritzalis, “Can We Trust This User? Predicting Insider’s Attitude via YouTube Usage Profiling,” 2013, pp. 347–354.
- [9] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, “Predicting Personality from Twitter,” in 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011, pp. 149–156.
- [10] D. Quercia, R. Lambiotte, D. Stillwell, M. Kosinski, and J. Crowcroft, “The personality of popular facebook users,” in *Proceedings of the ACM 2012 conference on computer supported cooperative work*, 2012, pp. 955–964.
- [11] Y. Shen and R. Jin, “Learning personal + social latent factor model for social recommendation,” 2012, p. 1303.
- [12] C. Biancalana, F. Gasparetti, A. Micarelli, and G. Sansonetti, “An approach to social recommendation for context-aware mobile services,” *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 1, pp. 1–31, Jan. 2013.
- [13] H. Ma, “An experimental study on implicit social recommendation,” in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, 2013, pp. 73–82.
- [14] H. J. Ahn, “A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem,” *Information Sciences*, vol. 178, no. 1, pp. 37–51, 2008.
- [15] P. G. Kelley and J. Cranshaw, “Conducting Research on Twitter: A Call for Guidelines and Metrics,” presented at the CSCW Measuring Networked Social Privacy Workshop 2013, 2013.
- [16] S. Wu, J. M. Hofman, W. A. Mason, and D. J. Watts, “Who says what to whom on twitter,” in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 705–714.
- [17] D. H. McKnight, V. Choudhury, and C. Kacmar, “Developing and validating trust measures for e-commerce: An integrative typology,” *Information systems research*, vol. 13, no. 3, pp. 334–359, 2002.
- [18] M. Ahuja, B. Gupta, and P. Raman, “An empirical investigation of online consumer purchasing behavior,” *Communications of the ACM*, vol. 46, no. 12, pp. 145–151, 2003.
- [19] I. Park and K. Min, “Making a List of Korean Emotion Terms and Exploring Dimensions Underlying Them,” *Korean Journal of Social and Personal Psychology*, vol. 19, no. 1, pp. 109–129, 2005.
- [20] D. Lee, J. Yeon, I. Hwang, and S. Lee, “KKMA : A Tool for Utilizing Sejong Corpus based on Relational Database,” *KIISE Transactions on Computing Practices (KTCP)*, vol. 16, no. 11, pp. 1046–1050, 2010.
- [21] A. Smith and J. Brenner, *Twitter use 2012*. “Pew Internet & American Life Project Washington”, DC, 2012.

Authors



Hyung Jun Ahn received his Ph.D. in Management Engineering from KAIST, Korea in 2004. Dr. Ahn has been a member of the faculty of the School of Business, Hongik University, Seoul, Korea since 2008. He is an Associate Professor now, and his main research interests include intelligent information systems, text mining, and social networks.