

네트워크 분석과 시각화를 통한 트위터 팔로우십 분석

Twitter Following Relationship Analysis through Network Analysis and Visualization

송등주(Deungjoo Song)**, 이창수(Changsoo Lee)**,
박찬권(Chankwon Park)***, 신기태(Kitae Shin)****

초 록

전 세계적으로 SNS(social network service) 사용자와 사용 시간이 매년 증가하고 있으며, SNS의 영향력 또한 점점 커져가고 있다. 이에 따라 SNS는 일상적인 의사결정에서부터 기업의 경영활동에 이르기까지 광범위하게 영향을 미치고 있다. 따라서 SNS를 적절히 분석하는 것은 매우 의미 있는 작업이 될 수 있는데, 이에 많은 연구들이 SNS에서의 다양한 활동과 관계를 들여다보려는 많은 노력들을 하고 있다. 본 연구에서는 대표적인 SNS 서비스 가운데 하나인 트위터를 이용해서 SNS 팔로잉 관계를 분석하고자 한다. 즉, 기존의 SNS 분석과는 달리 두 개 계정이 공통으로 팔로잉하는 계정들을 추출하고 시각화 함으로써 해당 계정들의 관심사를 분석하고자 한다. 이를 위해서 마이크로소프트 엑셀 매크로를 사용해 공통 팔로잉 계정을 추출하였으며, 인접행렬을 이용해 추출된 계정들 간의 관계를 정의하였다. 또한 팔로잉 관계 분석을 용이하게 하기 위해 방향 그래프를 이용해 시각화 하였으며, 이 같은 시각화에는 R 프로그래밍을 사용하였다.

ABSTRACT

The numbers of SNS (Social Network Service) users and usage amounts are increasing every year. The influence of SNS is increasing also. SNS has a wide range of influences from daily decision-making to corporate management activities. Therefore, proper analysis of SNS can be a very meaningful work, and many studies are making a lot of effort to look into various activities and relationships in SNS. In this study, we analyze the SNS following relationships using Twitter, one of the representative SNS services. In other words, unlike the existing SNS analysis, our intention is to analyze the interests of the

* First Author, Bachelor of Engineering, Dept. of Industrial and Management Engineering, Gangneung-Wonju National University(shadowweb@naver.com)

** Co-Author, Professor, Dept. of Industrial and Management Engineering, Gangneung-Wonju National University(cslee@gwnu.ac.kr)

*** Co-Author, Professor, Dept of Operations Management, Logistics and Distribution, Hanyang Cyber University(chankwon@hycu.ac.kr)

**** Corresponding Author, Professor, Dept. of Industrial and Management Engineering, Daejin University(ktshin@daejin.ac.kr)

Received: 2020-08-18, Review completed: 2020-08-24, Accepted: 2020-08-27

accounts by extracting and visualizing the accounts that two accounts follow in common. For this, a common following account was extracted using Microsoft Excel macros, and the relationship between the extracted accounts was defined using an adjacency matrix. In addition, to facilitate the analysis of the following relationships, a direction graph was used for visualization, and R programming was used for such visualization.

키워드 : 트위터, 복수 계정, 관계, 시각화

Twitter, Multiple Accounts, Relationship, Visualization

1. 서 론

페이스북(Facebook), 인스타그램(Instagram) 그리고 트위터(Twitter) 등으로 대표되는 SNS (social networking service or sites)는 인터넷과 모바일기기의 발달과 함께 급속도로 발전되었다. 독일의 온라인 통계포털 Statista의 2018년 11월의 통계에 따르면, 전 세계 SNS 이용자는 2010년부터 2017년 까지 9.7억 명에서 24.6억 명으로 몇 년 사이에 2배가 넘게 증가하였으며 2021년에는 30억 명을 넘어설 것으로 예측하고 있다. 전 세계 인구가 2017년 기준 약 75.1억 명인 것을 생각해 본다면, 인구의 약 3분의 1은 SNS를 이용한다고 할 수 있다[14]. 또한, Statista의 2019년 5월 통계에서는 2018년 전 세계 인터넷 이용자의 일일 평균 SNS이용 시간은 136분이라고 한다[13]. 이 두 통계는 SNS가 우리의 일상에 깊숙이 침투하였고, SNS를 통한 마케팅(SNS marketing)이나 오바마 미 전 대통령의 선거 공약으로 사용된 마이크로타겟팅(microtargeting) 등의 사례에서 보듯 SNS는 우리의 일상과 의사결정에 크고 작은 영향을 미친다[11, 20].

본 연구는 SNS 중 트위터를 중심으로 한 연구이다. 트위터의 여러 기능 중에 팔로잉(following) 기능을 중점적으로 다루었다. 트위

터 상에서 팔로잉은 관심이 있는 대상의 트윗들(tweets)을 -게시물과 메시지에 대한 상호작용 등- 내 화면에서 받아 볼 수 있는 기능이 다[12]. SNS 마다 팔로잉 기능을 뜻하는 명칭에는 차이가 있으나, 본질적으로 SNS 상에서 어느 한 대상을 팔로잉하고 있다는 것은 그 대상의 사소한 것 또는 중대한 것에 관심을 가지고 있다는 것을 의미한다. 또한 관계는 서로간의 상호작용 보다 더 오래 지속되는 정보이다. 그렇기에 팔로잉 상태를 분석하는 것으로 분석하는 대상의 과거와 현재의 관심사를 알 수 있고 더 나아가 미래의 동향도 예측할 수 있다[15].

본 연구는 한 계정의 관계를 분석하는 것이 아닌, 두 개의 계정이 공통으로 팔로잉한 교집합 계정들과 교집합 계정들 간의 관계를 엑셀의 매크로를 이용해 계정이 서로 팔로잉으로 연결되어 있으면 1을 아니면 0으로 연결 상태를 정의해주는 인접행렬(adjacency matrix)의 형태로 정리해 주고[17], R 프로그래밍의 igraph 패키지를 통해서 방향그래프(directed graph)로 시각화해 분석하고 이를 바탕으로 특성을 도출할 수 있도록 지원하는 것을 목표로 하고 있다[1, 7].

본 연구의 제2장에서는 두 계정의 팔로잉 페이지에서 팔로잉한 데이터를 추출해 교집합 부분만 추출해 그들의 관계를 인접행렬의 형태로 정리해 주고 방향그래프로 시각화 하는

방법론을 간단한 예시와 함께 소개한다. 제3장에서는 본 연구의 적용사례로써 전 세계적인 스타 방탄소년단(BTS)의 멤버들이 운영하는 계정과 소속사가 운영하는 계정, 대한민국 현대통령 문재인과 대한민국 청와대의 계정, 삼성(Samsung)과 엘지(LG) 계정을 앞서 소개한 방법론을 통한 분석을 기술하였다. 제4장에서는 본 연구의 결론과 추후 연구과제를 기술하였다.

2. 본 론

2.1 SNS 분석의 기존연구

소셜 네트워크(social network) 분석 프로그램의 공통적인 틀/framework)은 행렬 데이터를 그래프로 표현하는 그래프 이론의 수학적 접근 방식이다. 그래프 이론에서 개인 또는 집단, 생각 등은 ‘점’으로 관계는 ‘선’으로 표현된다. 그래프 이론을 기반으로 그래프의 점에 연결된 선의 수를 계산하거나 점과 점사이의 최단 경로를 계산하는 등의 방법, 그래프의 밀도(density)를 계산하는 방법 등이 있다[12]. 그러나 SNS로 인해 소셜 네트워크의 규모는 더욱 커지고 복잡해 졌다. 과거부터 이러한 SNS 상의 네트워크를 분석하는 많은 연구들이 진행되었다.

손중수 외 3명은 규모가 큰 온라인 SNS에서 최단 경로를 빠르게 계산하고 분석하기 위해 최상위선 탐색 방법에 전처리 과정을 추가해 개선된 소셜 네트워크 분석 방법을 제시하여 SNS에서 임의의 사용자를 탐색 할 때 연결이 많은 사용자를 먼저 탐색해 최단거리를 빠

르게 찾을 수 있게 하였다. 이 방법을 통해 정확도는 너비우선 탐색 방법에 비해 80% 수준이지만 7배 이상 더 빠르다는 장점을 가지고 있다. 그리고 근접 중심성과 매개 중심성 분석의 속도는 각각 1.8배, 6.8배 향상 시킬 수 있다[16]. 또한, 소현수 외 2명은 Python의 twython library를 이용한 두 가지 방법을 통해 트위터에서 각각 정보를 얻는다. 첫 번째는 특정 키워드가 포함된 게시물을 검색해 사용자에게 전송 하는 방법, 두 번째는 특정 GPS 정보를 통해 특정 장소에서 발생하는 게시물을 전송 하는 방법이다. 또한, 연관관계 분석을 통해 정보를 보다 정확하게 분석할 수 있는 방법을 제시하였다[14]. 제외동포 관련 언론보도 및 SNS 분석에서는 다음(www.daum.net)의 트위터 메시지 검색기능을 이용해 한 달간의 기간 동안 특정 키워드(제일동포, 세종교포 등)의 메시지를 긍정, 중립, 부정의 이미지와 5자 이내의 주제어로 정리한 이슈를 분석한 것을 제시하였다[13]. Zi Lu 외 3명은 SNS 커뮤니티에서 개인 노드 간 공간 분포 특성을 연구하기 위해 그래프 이론을 퍼지 접근법과 결합해 퍼지 SNS 커뮤니티 중심성 분석 방법을 제시하였다[8]. 최종산은 외국인의 한식에 대한 인식을 분석하기 위해서 트위터에 등장하는 텍스트들 중 ‘Korean Food’라는 핵심어가 포함된 텍스트들을 텍스트 마이닝 기법으로 추출해 텍스트에 담겨있는 긍정 또는 부정적 인식정도를 파악하는 감정분석, 단어의 출현빈도가 많은 핵심단어를 시각적으로 표현한 워드 클라우드(word cloud), 단어 간의 구조관계를 살펴보기 위해 텍스트에 등장하는 단어를 하나의 개체로 간주해 노드로 보고 소셜 네트워크 분석을 실시하였다[4]. 오선주는 소셜 네트워크에서

구성원간 관계를 효과적으로 찾기 위한 랭킹 기준과 랭킹 함수를 제시하였다[10].

이전의 연구들은 SNS 상의 메시지나 게시물 등이 가지고 있는 키워드를 중심으로 분석하거나 네트워크 내의 거리, 밀도 및 차수 등을 분석하고 있다. 그러나 본 논문에서는 네트워크와 네트워크 내에서 노드 간의 관계에 초점을 맞춘다. SNS상의 게시물과 메시지는 단발적인 정보를 가지고 있다. 그러나 서로간의 관계의 형성은 비교적 오래 지속된다. 그렇기에 경우에 따라서 서로간의 관계가 더 많은 정보를 제공해 주기도 한다. 그러나 SNS상에서 거대해진 네트워크를 한, 두 가지의 범주로 분류하고 관계의 특성을 파악하는 것은 매우 복잡하고 어렵다. 그래서 본 논문에서는 두 계정의 팔로잉 데이터를 통해 한 계정의 팔로잉 집단의 네트워크와 다른 한 계정의 팔로잉 데이터 집단의 네트워크의 교집합인 네트워크를 도출해 그 집단이 가지고 있는 특성을 분석하는 연구 방법을 제시한다.

본 연구는 두 계정의 교집합을 추출하는 방법을 사용하므로 기존의 연구인 두 계정의 관심사를 별도로 분석하는 것에 비해 노드와 관계의 양은 줄어든다. 또한 교집합 안에서의 네트워크 상태에 따라서 두 계정의 관계를 파악할 수 있다. 두 계정의 교집합의 네트워크가 복잡하다면 두 계정은 교집합 네트워크의 내부 계정이라 할 수 있다. 예를 들어, 임의의 계정 A와 B의 교집합의 네트워크가 영화라는 큰 범주로 대표될 수 있다고 가정하면 그 범주안의

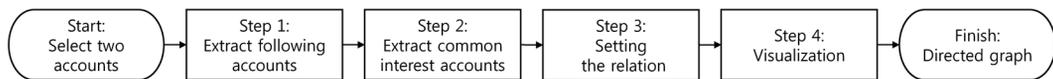
계정들은 영화감독, 배우, 배급사, 촬영관계자 등이 나타날 수 있다. 이때, 네트워크가 복잡하다면 A와 B는 강한 관계가 있음을 알 수 있다.

2.2 연구 방법론

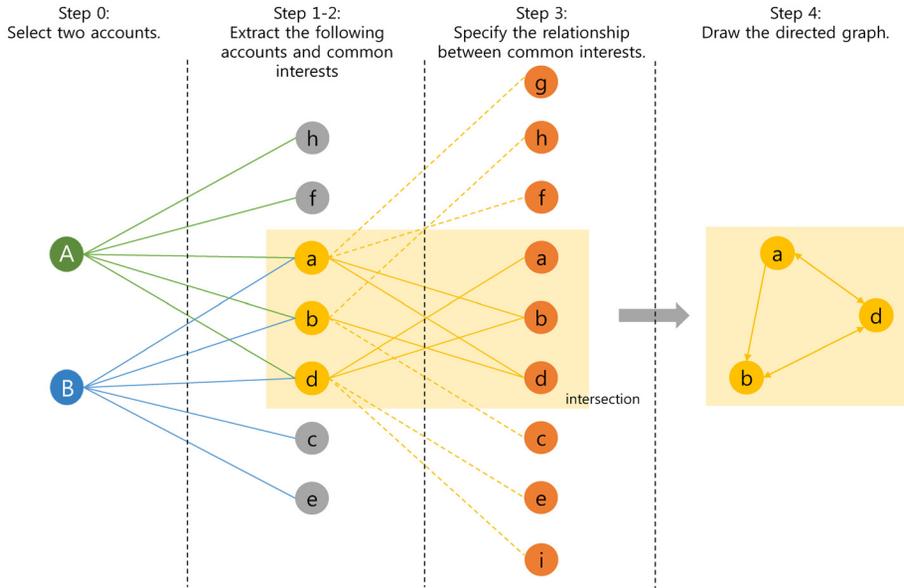
본 연구는 두 트위터 계정이 팔로잉한 계정들의 교집합 부분과 교집합 내의 네트워크를 구성해 관계를 파악하기 위한 연구이다. 전체적인 단계는 다음과 같으며 이를 순서도로 나타내면 <Figure 1>과 같다.

- 단계 0. 두 계정 선택: 분석자는 분석할 두 계정을 선택해 준다.
- 단계 1. 팔로잉 계정 추출: 선택된 두 계정의 팔로잉 페이지에서 계정 ID들을 각각 추출한다.
- 단계 2. 공통관심사 계정 추출: 추출된 계정 ID들 중 두 계정이 동시에 팔로잉한 교집합 계정들만 추출한다.
- 단계 3. 공통 관심사 간의 관계 지정: 교집합 계정들의 팔로잉 페이지에서 계정 ID들을 추출해 교집합 계정들의 팔로잉관계를 인접행렬의 형태로 지정해 준다.
- 단계 4. 시각화: 인접행렬을 방향그래프로 시각화한다.

<Figure 1>의 시작 단계(단계 0)과 같이 우선 분석할 두 계정을 선택해준다. 본 연구에서



<Figure 1> A Flow Diagram of the Research Process



<Figure 2> A Flow Diagram of the Research Process in Terms of SNS Accounts

는 A와 B로 가정한다. 다음으로 단계 1에서 두 계정(A, B)의 전체 팔로잉 계정들을 추출한 뒤, 단계 2에서 A와 B가 동시에 팔로잉한 교집합 계정들을 탐색한다. 그리고 단계 3에서 A와 B가 동시에 팔로잉한 교집합 계정들의 팔로잉 페이지에서 팔로잉한 계정들을 추출해 교집합 내에서 서로 간의 관계를 지정해 인접행렬을 만들어 준다. 마지막으로 단계 4에서 인접행렬을 R프로그래밍으로 시각화 한다. 이를 그림으로 나타내면 <Figure 2>와 같다.

단계 0. 우선 목표가 되는 두 계정을 선택해준다. 제2.3절에서는 임의의 계정 A와 B(이하 A, B)를 선택했다고 가정한다. 단계 1. A와 B의 팔로잉 페이지에서 A와 B가 팔로잉한 계정들을 각각 행과 열에 나열해 준다. A와 B의 팔로잉 페이지에서 추출한 계정 ID들을 각각 a, b, d, f, h와 a, b, c, d, e라고 가정한다면, 단계 1이 완료되었을 때의 행렬의 모습은 <Figure 3>와 같다. 편의상 h와 e는 생략하였다.

2.3 구현

분석자는 분석을 위해 두 개의 계정을 선택하여 계정 A, 계정 B로 표현하며, 여기서 추출한 팔로잉 계정은 소문자 a, b, c 등으로 표현한다. 즉, 계정 A의 팔로잉 페이지에서 추출한 팔로잉 계정을 a라고 하면 A-a로 표시한다.

	B-a	B-b	B-c	B-d
A-a				
A-b				
A-d				
A-f				

<Figure 3> The Matrix after the Step 1

단계 2. 공통 관심사 계정 추출: 단계 1에서 정리된 행렬을 바탕으로 행렬의 각 원소들의 상태를 결정해 준다. 이때, 행렬의 각 원소들은 첫 행과 첫 열의 ID에 의해서 0과 1의 값 중 한 가지를 갖는다. 여기서 1은 ID들 간의 ‘일치’를 의미하고 0은 ‘불일치’를 의미한다. 즉, A와 B 모두 a라는 계정을 팔로잉해 A-a와 B-a로 나타났다면, A-a의 행과 B-a의 열이 교차하는 곳의 값은 1이 된다. 이 방법으로 <Figure 3>의 행렬을 완성시키면 <Figure 4>와 같다.

	B-a	B-b	B-c	B-d
A-a	1	0	0	0
A-b	0	1	0	0
A-d	0	0	0	1
A-f	0	0	0	0

<Figure 4> The Example of Constructing the Common Interest Matrix

완성된 <Figure 4>의 행렬에서 각각의 행과 열에 대해서 행 합계와 열 합계를 구한 뒤, 행 합계 및 열 합계가 0일 경우 A와 B가 동시에 팔로잉한 계정이 아니므로 행렬에서 제외해서 두 계정의 교집합 계정들만을 추출한다. 여기서 교집합 계정들만 남은 행렬을 공통 관심사 행렬(common interest matrix)이라고 정의한다. <Figure 4>의 예시에서 행 합계는 위부터 1, 1, 1, 0이고 열 합계는 왼쪽부터 1, 1, 0, 1이다. 따라서 ID ‘f’를 포함하는 행과 ID ‘c’를 포함하는 열 합계가 0 이므로 A과 B의 공통 관심사가 아니기에 행렬에서 제외하고 단계 2를 마무리

한다. 본 단계의 결과로써 얻은 공통 관심사 행렬은 <Figure 5>와 같다. 단계 2의 마지막 과정에서 공통 관심사 행렬은 대칭행렬로 재조정된다.

	B-a	B-b	B-d
A-a	1	0	0
A-b	0	1	0
A-d	0	0	1

<Figure 5> The Common Interest Matrix

단계 3. 관심사간 관계 지정: 공통 관심사 행렬에서 첫 열의 모든 계정들의 팔로잉 페이지에 접근해 계정 ID들을 추출해 준다. 본문의 예시에서는 A-a, A-b, A-d의 팔로잉 페이지에서 계정 ID들을 추출해 준다. 이 추출된 ID들을 행렬의 옆에 목록을 만들어서 첫 행의 모든 계정 ID들과 비교해 준다. 본문의 예시에서는 B-a, B-b, B-d와 비교해준다. 목록을 만들어 비교해 주는 과정을 그림으로 표현하면 <Figure 6>과 같고, <Figure 6>에서의 목록은 A-a의 팔로잉 페이지에서 추출한 계정 ID들이다.

	B-a	B-b	B-d	a-b
A-a	0	1	1	a-c
A-b				a-f
A-d				a-d

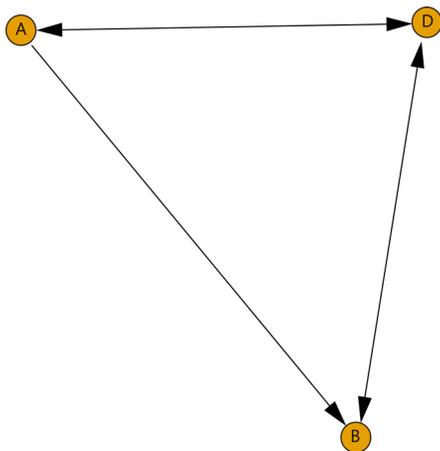
<Figure 6> The Common Interest Matrix and the List of Following Data

A-a의 팔로잉 페이지에서 추출한 계정 ID들을 b, c, f, d라고 하고, 이를 첫 행과 비교해 주면 B-b와 B-d의 b, d가 목록의 계정 ID와 동일하므로 A-a와 B-b, B-d가 교차하는 곳의 값은 1이 된다. 이 과정을 마지막까지 반복해 공통 관심사 행렬의 관계를 모두 지정해 <Figure 7>과 같은 인접행렬을 얻는다.

인접행렬에서 1은 팔로잉하고 있다는 것을 의미한다. <Figure 7>의 예시에서는 A-a가 B-b와 B-d를 팔로잉하고 있다. 마찬가지로 A-b는 B-d를 A-d는 B-a와 B-b를 각각 팔로잉하고 있다.

	B-a	B-b	B-d
A-a	0	1	1
A-b	0	0	1
A-d	1	1	0

<Figure 7> The Adjacency Matrix of Two Target Accounts



<Figure 8> A Directed Graph for the Accounts

단계 4. 시각화: 완성된 인접행렬을 R 프로그램

래밍의 igraph 패키지를 이용해 시각화 하면 다음의 <Figure 8>과 같은 방향그래프를 얻을 수 있다. 단계 3까지의 결과물을 CSV파일로 저장한다. 그 파일을 R studio에서 행렬 형식으로 불러온 뒤, 방향그래프를 작성할 수 있다.

3. 적용 결과

제3장에서는 전 세계적인 K-pop 스타 방탄소년단(BTS)의 두 계정, 대한민국 현 대통령과 청와대 그리고 삼성과 엘지의 계정을 제2장의 방법론을 통해 3가지 결과물을 도출하고 그에 대한 분석을 실시하였다. 분석에 앞서, 2019년 6월을 기준으로 작성되었다. 분석은 시기에 따라서 다소 차이가 있을 수 있음을 먼저 알린다. 또한, 웹 스크래핑 단계에서 3,000개 이상의 팔로잉 데이터 이상은 추출이 불가능해 각 계정당 3,000개 이하의 팔로잉 데이터에 대한 분석을 실시하였다.

3.1 방탄소년단(BTS) 계정 분석

분석은 방탄소년단의 두 트위터 계정인 A: https://twitter.com/BTS_twt(방탄소년단 멤버들이 직접 관리하는 계정, 이하 BTS1)과 B: https://twitter.com/bts_bighit(소속사에서 운영하는 계정, 이하 BTS2)를 이용하였다. 2019년 6월을 기준으로 BTS1과 BTS2가 팔로잉한 계정 수는 각각 143개, 22개로, 단계 1과 단계 2를 통해 두 계정의 공통 관심사 행렬을 도출한 것을 <Figure 9>에 나타내었다.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	@ARMYPEDIA/s_love_my/Vliveofficial/OMME_bigtuncle_bangGHITAudit/TS_jp_official/joohee9>theyyun12AMjinwo@hitmanb>BigHitEnt												
2	ARMYPEDIA	1	0	0	0	0	0	0	0	0	0	0	0
3	s_love_my	0	1	0	0	0	0	0	0	0	0	0	0
4	Vliveofficial	0	0	1	0	0	0	0	0	0	0	0	0
5	OMME_bigt	0	0	0	1	0	0	0	0	0	0	0	0
6	tuncle_ban	0	0	0	0	1	0	0	0	0	0	0	0
7	GHITAudit	0	0	0	0	0	0	1	0	0	0	0	0
8	TS_jp_offic	0	0	0	0	0	0	0	1	0	0	0	0
9	2AMjinwo	0	0	0	0	0	0	0	0	1	0	0	0
10	@joohee9	0	0	0	0	0	0	0	1	0	0	0	0
11	@hitmanb	0	0	0	0	0	0	0	0	0	1	0	0
12	GHITAudit	0	0	0	0	0	1	0	0	0	0	0	0
13	>theyyun1	0	0	0	0	0	0	0	0	1	0	0	0
14	@BigHitEnt	0	0	0	0	0	0	0	0	0	0	0	1

<Figure 9> The Common Interest Matrix of BTS1 and BTS2

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ARMYPEDIA/s_love_mys/liveofficial/OMME_bigtuncle_bangGHITAudit/TS_jp_official/joohee9 theyyun11AMjinwoo hitmanb BigHitEnt												
2	ARMYPEDIA	0	0	0	0	0	0	1	0	0	0	0	1
3	s_love_mys	0	0	0	0	0	0	0	0	0	0	0	1
4	Vliveofficial	0	0	0	0	0	0	0	0	0	0	0	1
5	OMME_bigt	0	0	0	0	1	1	0	0	1	0	1	1
6	tuncle_ban	0	0	0	0	0	1	0	0	1	0	1	1
7	GHITAudit	0	0	0	0	0	0	0	1	1	1	1	1
8	TS_jp_offic	1	1	0	0	0	0	0	0	0	0	0	1
9	joohee9	0	0	0	0	0	0	0	0	1	1	1	0
10	theyyun11	0	0	0	1	0	0	0	1	0	1	1	1
11	AMjinwoo	0	0	0	0	0	0	0	1	1	0	1	0
12	hitmanb	1	1	0	1	1	1	1	1	1	1	0	1
13	BigHitEnt	1	1	0	1	1	1	0	1	1	1	1	0

<Figure 10> The Adjacency Matrix of BTS1 and BTS2

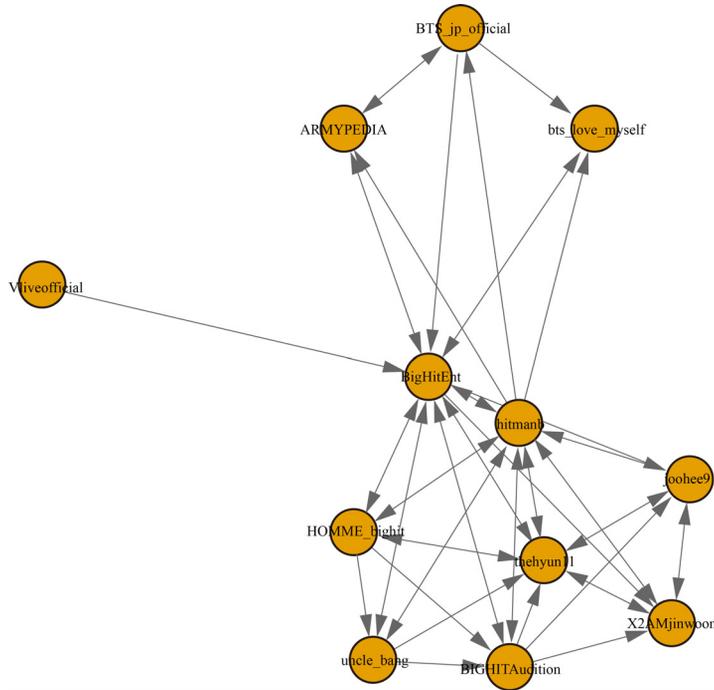
BTS1의 143개의 팔로잉 계정들과 BTS2의 22개의 팔로잉 계정들 중 12개의 계정이 BTS1과 BTS2가 공통으로 팔로잉한 계정들임을 <Figure 9>의 공통 관심사 행렬을 통해 알 수 있다. 이 행렬을 단계 3을 통해 관심사간의 관계를 지정해 주면 다음의 <Figure 10>과 같은 인접행렬을 얻을 수 있다.

<Figure 10>의 인접행렬에서 BigHitEnt가 다른 계정들로부터 가장 많이 팔로잉 당하고 있으며, hitmanb는 다른 계정들을 가장 많이 팔로잉하고 있음을 알 수 있다. 이 인접행렬을 R 프로그래밍으로 시각화 하면 다음의 <Figure 11>과 같은 방향 그래프를 얻을 수 있다.

방향그래프에서 다른 계정들이 가장 많이 팔로잉한 계정은 hitmanb와 BigHitEnt이다. 이들은 각각 BTS의 소속사인 빅히트 엔터테인먼트(Big Hit Entertainment)의 대표 방시혁과 빅히트 엔터테인먼트의 계정이다. 다음으로 그룹 에이트

(eight)와 음므(HOMME)의 이현(thehyun11)이 있다. 두 계정 주위로 맨 위부터 시계 방향으로 BTS의 일본 공식계정인 BTS_jp_official, 유니세프(UNICEF)와 파트너십을 맺고 자선활동을 홍보하는 bts_love_myself, 그룹 에이트의 멤버 주희(joohee9), 그룹 2AM의 멤버 진운(X2AMjinwoon), 빅히트 엔터테인먼트의 신인 계발팀의 계정인 BIGHITAudition, 빅히트 엔터테인먼트의 대표 방시혁의 어린이 동요 사업인 잉클뱅크(uncle_bang), 그룹 음므의 계정(HOMME_bright), 연예인들의 일상을 동영상으로 팬들과 소통하는 Vlive(Vliveofficial)이 있으며, 마지막으로 방탄소년단의 팬클럽인 아미(ARMY)와 함께 만드는 BTS의 기억저장소인 ARMYPEDIA가 있다.

전체적으로 보면 빅히트 엔터테인먼트의 대표 방시혁과 빅히트 엔터테인먼트의 계정을 중심으로 회사의 관계자들과 사업들 그리고 BTS



<Figure 11> The Directed Graph of BTS1 and BTS2

의 캠페인 등이 주를 이루고 있다. 또한, 2AM의 다른 멤버들보다 진운이 더 활발한 관계를 맺고 있는 것을 알 수 있다. 그래프에서 특히 주목해 볼만한 점은 아이들이나 아이들을 키우는 부모가 아니면 알기 어려웠던 잉클뱅이라는 사업이다.

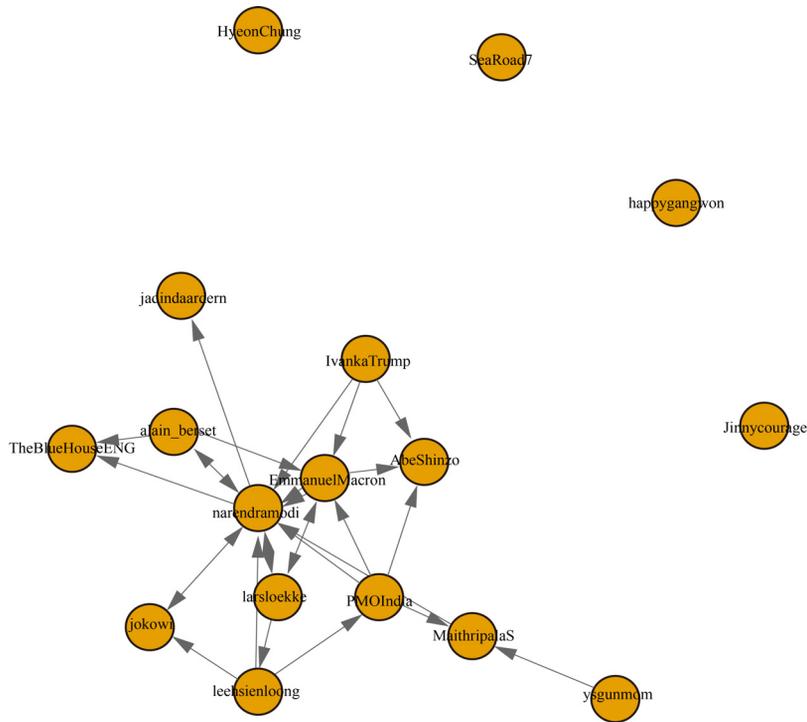
이 집단 내에서 발생하는 트윗은 BTS1과 BTS2에 전달이 되며, hitmanb와 BighitEnt의 트윗은 집단 내 거의 모든 계정들에게 전달할 수 있다.

3.2 문재인 대통령과 청와대 계정 분석

2019년 6월 기준 각각 153,000여 개, 32,700여 개의 계정을 팔로잉한 대한민국 19대 대통령 문재인 대통령의 계정(<https://twitter.com/moonriver365>)

과 청와대(<https://twitter.com/TheBlueHouseKR>) 계정의 공통 관심사 행렬을 처리하여 총 17개의 계정이 두 계정의 공통 관심사로 도출되었다. 이 공통 관심사 행렬을 단계 3을 통해 공통 관심사 간의 관계를 지정해 인접행렬로 만들었다.

문재인 대통령과 청와대 계정의 인접행렬에서 narendramodi가 다른 계정들이 가장 많이 팔로잉한 계정이며, 다른 계정들을 가장 많이 팔로잉하고 있음을 알 수 있다. 또한, 이 전의 분석과는 다르게 독립적인 계정들도 나타났다. 이는 문재인 대통령과 청와대의 계정은 이 독립적인 계정들을 팔로잉 했지만 행렬 내의 다른 계정들은 이들과 관계가 없기 때문이다. 이 인접행렬을 방향그래프로 시각화한 것을 다음의 <Figure 12>에 나타내었다.



<Figure 12> The Directed Graph of President Moon Jae-In and Blue House

<Figure 12>에서 가장 중심이 되는 계정은 앞서 공통 관심사 행렬과 같이 narendramodi로 타임지가 선정한 2019년 가장 영향력 있는 인물 100인에 선정된 인도의 열네 번째 총리 Narendra Modi이다. 두 계정의 주위로 위에서 부터 반시계 방향으로, 마흔 번째 뉴질랜드의 수상이며 노동당(Labour Party)의 리더 Jacinda Ardern(jacindaardern)이 있다. 다음으로 alain_berset은 스위스 대통령이자 유럽인 사회주의자들의 당인 스위스 사회 민주당의 Alain Berset, 청와대의 영문 계정(TheBlueHounseENG), 인도네시아의 일곱 번째 대통령 Joko Widodo(jokowi), 싱가포르의 세 번째 수상이며 중도 무파당인 인민행동당(People's Action Party(PAP))의 리더 李顯龍(Lee Hsien Loong, leehsienloong),

스물다섯 번째 덴마크 수상이며 2011년에 대한민국의 수교훈장을 수여받은 Lars Løkke Rasmussen(larsloekke)이 있다.

위에서 언급되지 않은 독립적인 4개의 계정과 ysgunmom계정을 살펴보면, 테니스선수 정현(HyeonChung), 친일청산과 적폐청산을 메인에 걸어둔 Seaload7, 경제전문미디어 이투데이의 김진희 기자로 추정되는 Jinnycourage, 강원도의 계정(happygangwon), 무비판적지지와 민주당권리당원이라고 소개되어 있는 ysgunmom계정이 있다.

전체적으로 인도의 Narendra Modi 총리를 중심으로 관계가 있음을 알 수 있다. Modi 총리를 제외하고는 서로간의 관계는 상당히 미미하다고 볼 수 있다.

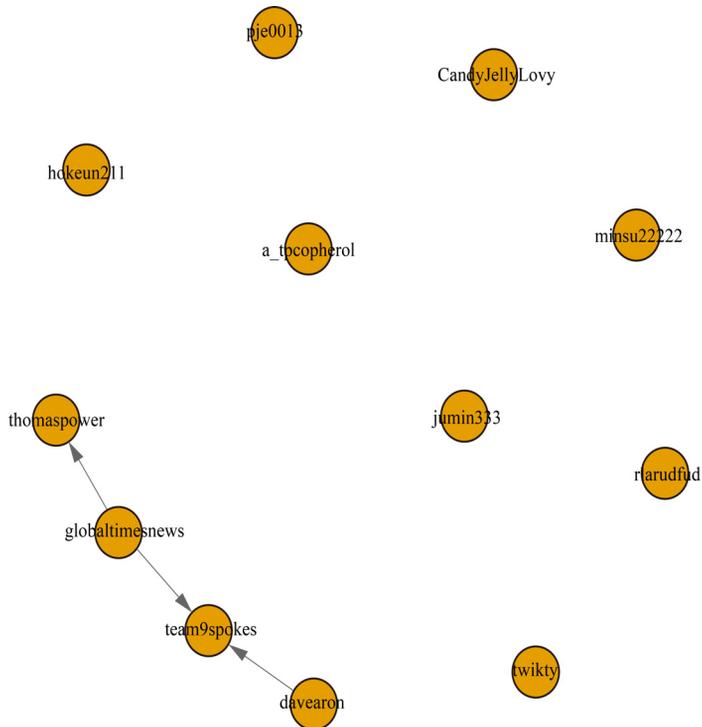
3.3 삼성(SAMSUNG)과 엘지(LG) 계정 분석

삼성(Samsung)과 엘지(LG)는 휴대폰, 컴퓨터, 가전 등 여러 분야에 걸쳐서 오랜 라이벌 관계이다. 이 전까지는 서로 방향성이 같은 두 개의 계정의 관심사를 도출하고 시각화 하였으나 이 절에서는 서로 경쟁관계일 때 어떠한 결과가 나올지 알아보고자 한다. 삼성은 105,000여 개의 계정을 팔로잉한 삼성전자의 최신 소식을 알려주는 삼성전자 뉴스룸(<https://twitter.com/SAMUNGtomorrow>)과 6.373개의 계정을 팔로잉한 엘지전자(<https://twitter.com/LGElectronics>)의 공통 관심사 행렬을 처리하여 삼성과 엘지의 공통 관심사로 총 12 개의 계정

이 도출되었다. 이 공통 관심사 행렬을 단계 3을 통해 관심사 간의 관계를 지정해 인접행렬로 정리하였다.

삼성과 엘지의 인접행렬을 보면 독립적인 계정들이 행렬의 대부분을 차지하는 것을 볼 수 있다. 이는 방향 그래프로 시각화 한 자료를 가지고 더 분석해 보겠다. 이 인접행렬을 방향 그래프로 시각화한 것을 다음의 <Figure 13>에 표시하였다.

<Figure 13>을 보면 서로 경쟁관계를 대변해 주듯이 관심사들 간의 관계가 미미한 것을 볼 수 있다. 관계가 있는 계정들을 먼저 살펴보면 맨 위부터 토큰(Token)과 이코노믹스(Economics)의 합성어이며 블록체인(Block Chain) 기술에 관한 책인 토크노믹스(Tokenomics)의 저



<Figure 13> The Directed Graph of Samsung and LG

자인 Thomas Power(thomaspower), 세계의 주요 사를 중국 정부의 관점에서 저술하는 중국의 환구시보(globaltimesnews), 대시보드 기술(Dashboard Technology)의 9Spokes(team9spokes), IT분야의 리서치 기업인 가트너 부사장 Dave Aron(davearon)이 있다. 관계가 없이 독립적인 계정들은 알파-토코페롤(Alpha-Tocopherol, a_tocopherol)을 제외하고는 공식적인 정보를 찾을 수 없었다.

종합적인 데이터를 분석해 보면 삼성과 엘지의 관심사들 간의 관계는 미미하지만 공통적으로 중국의 정세와 대시보드, 블록체인 기술에 관심을 가지고 있는 것을 알 수 있으며, 최근 ‘삼바’-삼성이 바이오 분야에 전폭적인 투자를 한 것-의 영향으로 알파-토코페롤이 공통 관심사로 도출 되었을 것으로 보인다.

5. 결 론

본 연구의 제2장에서는 두 트위터 계정의 팔로잉 페이지에서 엑셀 매크로를 이용해 두 계정이 공통적으로 팔로잉한 계정들을 공통 관심사 행렬로 도출하고 그 계정들 간의 관계를 인접행렬의 형태로 만들어 준다. 또한, 인접행렬을 R 프로그래밍을 이용해 방향그래프로 시각화하는 방법론을 제시하였다. 제3장에서는 방법론을 이용해 우리에게 친숙한 연예인, 정치인과 기업의 계정을 방법론을 이용해 분석해 보았다.

본 방법은 한 개의 계정을 사용해서 그래프를 그리고 분석하는 방법과는 다르게 동시에 두 개의 계정을 사용하기 때문에 계정 1과 계정 2의 공통된 관심사와 그렇지 않은 관심사를 구

별해 내고 관심사들 간의 관계를 파악할 수 있다. 또한, BTS의 두 계정 분석에서 방시혁 대표의 잉클뱅크 사업처럼 일반인들에게 잘 알려지지 않은 사실들을 시각화해서 보여줄 수 있다. 주기적으로 분석 시 계정들간의 방향성 또한 파악이 가능하다고 본다. 또한, 다른 지표들과 더불어서 사용할 수 있는 보조지표로서도 가치가 있다고 판단한다. 본 연구에서 제안된 방법론을 통하여 단일 계정에서 파악하지 못하였던 새로운 관계망의 특성을 도출할 수 있고 또한, 공통 관심사를 도출하여 분석의 복잡도를 줄일 수 있도록 하였다. 본 연구의 결과가 관계망의 특성을 자동으로 분석하지는 못하지만, 관계망의 시각화를 통하여 정성적이고 직관적인 분석이 가능하도록 만들어 준다는 점에서 연구의 가치를 살펴볼 수 있을 것이다.

본 연구의 제한사항으로는 트위터에서 팔로잉 데이터를 수집할 때, 트위터 팔로잉 페이지 내에서 일정 수준 이후로는 팔로잉한 계정들의 데이터가 로드되지 않아서 팔로잉의 수가 아무리 많아도 2,000에서 3,000개 사이의 계정밖에 수집할 수 없는 것이 제한사항이다.

추후 연구방향으로는 계정들 간의 상호작용의 정도를 표현하고, 팔로잉을 한 시점 등을 추가해 어떤 사건이나 행위에 다른 계정들이 관심을 가지게 되었는지에 대한 데이터를 추가해 볼 수 있겠다. 또한, 본문에서 간략히 언급했던 목록을 만들어서 계정간의 관계를 비교해주는 과정이 개선되어야 할 부분이다. 다른 방향에서는 본 연구에서는 팔로잉만을 다루었으나 SNS에서의 좋아요와 같은 활동의 분석에도 동일하게 적용될 수 있을 것으로 판단되므로 이러한 방향으로의 확장도 가능할 것으로 보인다. 이는 추후 연구에서 다루도록 한다.

References

- [1] Bang-Jensen, J. and Gutin, G. Z., "Digraphs: Theory, algorithms and applications," 2007. 8. 15.
- [2] Butler, S. K., "Eigenvalues and Structures of Graphs," University of California, San Diego, 2008.
- [3] Cho, M., "R programming for Big Data Analysis, 283," Information Publishing Group, 2016.
- [4] Choi, J. S., "Analysis on foreigners' perception of Korean food using social big data," *Asia-Pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 7, No. 8, pp. 427-437, 2017.
- [5] Clement, J., "Daily time spent on social networking by internet users worldwide from 2012 to 2018 (in minutes)," Statista, 2019. 5. 14. <https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/>.
- [6] Clement, J., "Number of social media users worldwide from 2010 to 2021 (in billions)," Statista, 2018. 11. 29. <https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>.
- [7] Csárdi, M. G., "Network analysis and visualization," 2019.4.22.
- [8] Lu, Z., Zhang, Q., Wu, D., and Gao, F., "A fuzzy SNS community centrality analysis method and a case study on Ren Ren Net in China," *Decision Making and Soft Computing*, pp. 530-535, 2014.
- [9] Ognyanova, K., "Network analysis and visualization with R and igraph," NetSciX 2016 School of Code Workshop, <https://kato.net/netscix2016.html>.
- [10] Oh, S., "A model for ranking semantic associations in a social network," *The Journal of Society for e-Business Studies*, Vol. 18, No. 3, pp. 93-105, 2013.8.
- [11] Pahwa, A., "What is Microtargeting?," Feedough, 2018. 2. 24. <https://www.feedough.com/what-is-microtargeting/>.
- [12] Scott, J., "Social network a handbook," Sage Publications, London, 1987.
- [13] Shin, H. C., Seo, S. Y., Park, H. Y., and Cho, Y. G., "Analysis of major media reports and SNS related to overseas Koreans: Focusing on content analysis of reports from 5 media companies from 2012 to 2015," *Conference Proceedings of Korean Academic Society for PR*, pp. 29-37, 2016.
- [14] So, H. S., Kang, S. S., Oh, S. W., "Cooccurrence relation analysis and visualization in tweet for food safety domain, 2016.
- [15] Solis, B., "A prediction: Twitter to predict the future," *SocialMediaToday*, 2010. 4. 7. <https://www.socialmediatoday.com/content/prediction-twitter-predict-future>.
- [16] Son, J. S., Jo, S. H., Kwon, K. R., and Jeong, I. J., "Improved social network service analysis method in SNS," *Journal of Intelligence and Information Systems*,

Vol. 18, No. 4, 2012.

- [17] Takuro Sasaki, Technology of Web Data Collextion, 252, Hanbit Media, 2017.
- [18] Twitter, Wikipedia, https://en.wikipedia.org/wiki/Twitter#Adding_and_following_content.
- [19] Web scraping, Wikipedia, https://en.wikipedia.org/wiki/Web_scraping.

- [20] Young, A., "How data and micro-targeting won the 2012 election for obama-Antony young-mindshare North America, Media Village, 2012. 11. 20. <https://www.mediavillage.com/article/how-data-and-micro-targeting-won-the-2012-election-for-obama-antony-young-mindshare-north-america/>.

저 자 소 개



송등주
2019
관심분야

(E-mail : shadowweb@naver.com)
강릉원주대학교 산업경영공학과 졸업
빅 데이터



이창수
1987
1994
1992년~현재
관심분야

(E-mail: cslee@gwnu.ac.kr)
서울대학교 산업공학과 (학사)
서울대학교 산업공학과 (석·박사)
강릉원주대학교 산업경영공학과 교수
데이터 품질



박찬권
1987
1996
2003년~현재
관심분야

(E-mail: chankwon@hycu.ac.kr)
서울대학교 산업공학과 (학사)
서울대학교 산업공학과 (석·박사)
한양사이버대학교 경영학부 교수
전자문서 표준화, ERP/SCM, 물류정보시스템



신기태
1987
1995
1995년~현재
관심분야

(E-mail: ktshin@daejin.ac.kr)
서울대학교 산업공학과 (학사)
서울대학교 산업공학과 (석·박사)
대진대학교 산업경영공학과 교수
기업정보시스템