



트위터는 소셜 네트워크인가? 네트워크 구조와 정보 전파의 관점

곽해운

KAIST 전산학과 박사과정 haewoon@an.kaist.ac.kr

이창현

KAIST 전산학과 박사과정 chlee@an.kaist.ac.kr

박호성

KAIST 전산학과 박사과정 hosung@an.kaist.ac.kr

문수복

KAIST 전산학과 교수 sbmoon@kaist.edu

이 연구에서는 4100만명 이상의 트위터 사용자 정보와 14억 7천만개의 팔로(Follow) 관계, 그리고 사용자들이 남긴 1억개 이상의 트윗들을 수집, 분석하여 트위터 사용자 네트워크의 구조적 특성과 정보 전파의 특성을 분석하였다. 단방향 팔로 관계의 상호성(Reciprocity)은 일반적인 사회적 관계 및 여타 온라인 소셜 네트워크에서 관찰되는 상호성보다 상당히 낮은 22.1%로 관찰되었으며 이는 트위터에서의 팔로 관계가 친밀한 사회적 관계에 기반하고 있지만은 않다는 것을 보여준다. 트위터의 사용자들은 오프라인에서의 이슈와 관련된 트윗들을 활발히 기록하며, 많은 팔로어를 갖는 허브(Hub) 사용자들이 풍부하게 존재하고, 팔로어를 많이 갖지 못한 사용자들도 리트윗을 통해 정보를 빠르고 넓게 퍼뜨릴 수 있다는 것을 정량적으로 밝혔다. 관계의 단방향성과 낮은 상호성, 그리고 빠르고 넓은 정보 전파는 다른 소셜 네트워크 서비스에서는 찾아볼 수 없는 트위터의 고유한 특성으로서, 새로운 정보 전파 매체로서의 가능성을 보여준다.

KEYWORDS 트위터 • 온라인 소셜 네트워크 • 네트워크 구조 • 정보 전파

1. 서론

2006년 처음 서비스를 시작한 트위터(Twitter)는 2010년 10월 전세계 1억 7,500만명의 사용자를 확보한 온라인 웹 서비스가 되었다. 2010년 4월 시점의 발표에 따르면 매일 약 30만명의 사용자가 새롭게 트위터를 사용하기 시작하였는데, 10월 시점에는 매일 약 50만명의 새로운 사용자가 트위터에 가입하는 것으로 더욱 그 성장세가 가속되고 있다¹⁾. 전세계 5억명의 사용자를 확보하고 있는 온라인 소셜 네트워크 서비스인 페이스북(Facebook)의 시장 확대와 맞물려 마이스페이스(MySpace)나 오르컷(Orkut) 등의 다른 온라인 소셜 네트워크 서비스들의 규모가 축소되고 있는 것²⁾과는 달리 트위터는 지속적으로 그 서비스의 규모가 확대되고 있다. 이것은 트위터가 제공하는 유저 경험이 페이스북 또는 여타 소셜 네트워크 서비스들이 제공하는 것과는 다른 것임을 암시하지만, 트위터가 제공하는 다른 사용자들과 온라인 관계를 맺고, 짧은 글을 남길 수 있고, 다른 사용자의 글을 확인할 수 있는 기본적인 기능은 다른 온라인 소셜 네트워크의 기능들과 일견 유사하다. SMS나 스마트폰을 통해 글을 남길 수 있는 기능 역시 최근의 온라인 소셜 네트워크 서비스들 대부분이 제공하는 기능이다. 그렇다면 어쩌서 트위터만이 페이스북의 성장에 크게 영향을 받지 않고 있는 것일까?

최근 미디어에 자주 보도되고 있는 트위터의 모습을 살펴보면 그 답을 간접적으로 찾아볼 수 있다. 2009년 1월 15일 허드슨강에 불시착한 비행기의 사진³⁾이 트위터를 통해 널리 알려진다든가, 같은 2009년 6월 이란 부정 선거의 시위 소식도 거의 실시간으로 트위터를 통해 전세계로 퍼져 나갔다. 한국에서도 2010년 9월 21일 서울에 많은 비가 내렸을 때

1) <http://www.nytimes.com/2010/10/31/technology/31ev.html>

2) <http://www.vincos.it/world-map-of-social-networks>

3) <http://twitter.com/#!/jkrums/status/1121915133>

트위터 사용자들이 실시간으로 각 지역의 홍수 상황을 사진과 함께 업데이트함으로써 기존의 뉴스 미디어보다 더 빠르고 효율적인 정보 전달을 시도한 바 있다(WikiTree, 2010). 이런 정보의 전파는 기존의 온라인 소셜 네트워크를 통해서는 얻지 못했던, 또는 얻기 어려웠던 효용이었다.

이러한 트위터의 특징을 아래의 질문에 대한 대답을 통해 단계적으로 분석하려는 것이 이 연구의 목적이다.

- 1) 트위터의 단방향적인 팔로 관계의 특성은 어떠한가?
- 2) 트위터에서 활발히 전달되는 정보의 종류와 성격은 어떠한가?
- 3) 입소문 방식의 리트윗을 통해 정보는 얼마나 빨리, 넓게 퍼져 나가는가?

먼저 페이스북을 비롯한 여타 온라인 소셜 네트워크 서비스들이 기본적으로 친한 친구들, 또는 오프라인에서 원래 서로 알고 있던 사이의 연결을 위한 서비스 제공에 초점이 맞춰져 있다면 트위터는 그보다 느슨하고 긴 연결을 중심으로 서비스를 제공하고 있다는 것을 밝히고, 그것을 가능케 한 단방향적 팔로 관계의 특성을 알아본다. 다음으로 다른 온라인 소셜 네트워크와의 비교를 통해 트위터의 네트워크 구조가 갖는 특성을 알아본다. 또한 트위터에서 활발히 전달되는 정보의 종류와 시간적 특성에 따른 분류를 수행하였다. 마지막으로 트위터 네트워크 위에서 정보를 널리 알리기 위해 사용되는 리트윗(Retweet)의 기능을 알아보고, 리트윗에 의한 정보 전파의 특성을 알아보려고 한다.

이 연구는 저자의 기존 연구(Kwak, Lee, Park & Moon, 2010)에서 수집된 데이터 및 실험 결과의 일부를 트위터의 미디어적 특성이라는 관점에서 재해석한 결과가 포함되어 있음을 알려둔다.

2. 이론적 배경

1) 트위터의 기능 소개

이 장에서는 트위터가 제공하는 기능들 중 트위터가 정보 전파에 유리한 특성을 갖게 하는 핵심 기능들을 간략하게 소개한다. 먼저 일반적인 온라인 소셜 네트워크의 양방향 친구 관계와 다른 트위터의 단방향 팔로 관계를 언급하고, 단방향 온라인 관계가 트위터에서의 메시지 전달 기능과 밀접하게 결합되어 있음을 설명한다. 또한 사용자들이 어떤 식으로 메시지를 주고 받으며 인터랙션할 수 있는지를 알아본다. 다음으로 트위터의 정보에 다양한 방법으로 접근을 가능하게 하는 API(Application Programming Interface)에 대해 알아보고, 마지막으로 리트윗이라 불리는 메시지의 공유 기능에 대해서 적으려고 한다.

(1) 단방향(One-way) 온라인 관계 맺기

싸이월드를 시작으로 마이스페이스, 페이스북 등의 온라인 소셜 네트워크 서비스의 온라인 관계는 ‘친구 초대(또는 신청)’와 ‘초대(신청) 수락’의 두 단계로 형성되는 것이 일반적이다. 예를 들어 사용자 A와 온라인 친구 관계를 맺고 싶은 사용자 B가 있을 때, B가 A에게 먼저 친구 관계 신청을 하게 되고 A는 B의 친구 신청을 수락하는 순서로 온라인 친구 관계가 형성되는 것이다. 다른 사용자로부터의 친구 신청을 수락하는 행위는 대부분의 온라인 소셜 네트워크 서비스에서 필수적인 단계로, 두 사용자가 모두 새로운 온라인 관계의 형성에 신청과 수락이라는 형태로 동의하는 과정이 필요하다. 이러한 신청과 동의의 두 단계는 온라인 소셜 네트워크 서비스에서의 친구 관계가 대칭적이기 때문이다. 즉, 일반적인 온라인 소셜 네트워크 서비스들에서는 어느 사용자가 먼저 친구 신청을 했는지는 중요하지 않으며 사용자 A가 사용자 B의 친구라면, 사용자 B 역시 사용자 A의 친구라는 대칭적이고 동등한 온라인 관계를 형성한다.

이와 달리 트위터에서는 단방향의 온라인 관계를 형성할 수 있다. 사용자 A가 사용자 B와 온라인 관계를 맺고 싶을 때, 사용자 B의 동의가 없어도 사용자 A가 일방적으로 사용자 B를 팔로(Follow)하는 단방향의 관계를 맺을 수 있는 것이다. 이 때 사용자 A를 사용자 B의 팔로어라고 하고, 사용자 B를 사용자 A의 팔로이라고 한다.

온라인 소셜 네트워크에서의 온라인 관계 형성 과정과 트위터에서의 온라인 관계 형성 과정을 비교해 보면 두 가지의 큰 차이를 발견할 수 있다. 하나는 온라인 관계의 형성에 있어 상대의 동의가 필요 없다는 점이고 다른 하나는 형성된 온라인 관계가 대칭적이지 않다는 것이다. 온라인 관계의 형성에 있어 상대의 동의가 필요하지 않다는 것은 트위터에서의 온라인 관계가 다른 온라인 소셜 네트워크에서보다 쉽게 형성될 수 있다는 것을 의미한다. 온라인 소셜 네트워크에서 친구 신청을 할 때에는 상대방이 신청을 수락할지에 대한 고려가 선행되기 때문에 온라인에서만 새로운 관계를 형성하는데 노력하기보다는, 기존의 오프라인에서 알고 있던 친구들을 중심으로 온라인 친구 관계를 형성하게 되는 것이 보통이다. 또한 많은 온라인 소셜 네트워크에서 온라인 친구 관계에 따라 사진 등의 콘텐츠를 열람할 수 있는 권한을 조정하기 때문에 친구 신청/수락의 과정이 조심스럽게 일어나는 편이다. 그러나 트위터의 경우에는 한 사용자가 다른 사용자를 팔로할 때 상대방의 동의 없이 자신의 의사만으로 온라인 관계를 형성할 수 있고 관계가 맺어진 이후에도 팔로어와 팔로이의 명확한 구분이 존재하기 때문에 쉽게 온라인 관계가 형성될 수 있다. 결과적으로 온라인 친구 신청에 대한 심리적인 부담이 전혀 없기 때문에 다양한 계층의 사람들과 보다 쉽게 온라인 연결을 맺게 되는데, 이런 방식으로 맺어진 관계들은 주변의 또래 집단 안에 한정되는 관계가 아니라 기존의 집단을 크게 벗어나는 긴 연결들이 되어 정보 전파에 있어 중요한 역할을 할 수 있게 된다(Granovetter, 1973).

(2) 단방향 온라인 관계와 메시지의 전달

트위터에서의 단방향 온라인 관계는 단순히 다른 사용자에 대한 지지나 애정만을 나타내는 것이 아니라 트위터에서의 메시지 전달과 밀접한 관련이 있다. 트위터에서 각 사용자는 140자 길이의 짧은 메시지를 남길 수 있다. 이 메시지는 트윗(Tweet)이라 부른다. 트위터에서는 로그인을 통해 인증된 사용자의 기본 화면에서 그 사용자가 팔로하고 있는 모든 사용자들의 최근 트윗들을 시간의 역순으로 정렬해서 보여준다. 이 화면은 타임라인(Timeline)이라고 부르는데 시간이 지나가면 계속해서 자동으로 새로운 트윗들이 업데이트되게 된다. 즉, 한 사용자가 다른 사용자를 팔로하는 것은 해당 사용자의 트윗을 타임라인에서 지속적으로 구독(subscribe)하겠다는 의미를 갖는다. 이것은 블로그의 RSS 주소를 RSS 리더에 등록하는 경우와 비슷하다. 트위터에서 팔로 관계를 맺을 때 상대방의 동의가 필요하지 않듯이, 블로그의 RSS 주소를 RSS 리더기에 등록하는데도 블로그 관리자의 동의는 필요하지 않다. 또한 팔로 관계를 맺고 나면 상대방의 새 트윗이 사용자의 타임라인에 업데이트되게 되는데 이것 역시 블로그의 RSS 주소를 등록하고 나면 블로그에 새 글이 올라올 때마다 RSS 리더에 새 글이 업데이트되는 것과 같다. 즉 팔로는 단순히 친교를 나타내는 온라인 관계만이 아니라 트윗이라는 콘텐츠의 구독을 하는 기능적인 역할도 담당하고 있다.

직접 특정 사용자를 지칭해서 트윗을 남길 수도 있는데 이 때는 상대방의 아이디(ID) 앞에 @를 붙이고 트윗을 남기면 된다. @를 사용해서 사용자에게 직접 트윗을 남기는 경우 두 사용자 간의 온라인 관계가 반드시 맺어져 있어야 하는 것은 아니며, 사용자는 타임라인과는 다른 별도의 페이지에서 자신에게 남겨진 모든 메시지를 확인할 수 있다. 즉 많은 팔로이를 갖고 있어서 그들의 최신 트윗에 의해 타임라인이 빠른 속도로 갱신되는 사용자의 경우에도 다른 사용자들이 @를 이용해 자신에게 남긴 트윗은 쉽게 확인할 수 있다. 이 때 @의 위치가 트윗의 맨 앞인지 아닌지에 따라 그 트윗의 전달이 달라지므로 구별해서 사용해야 한

다. @의 위치가 맨 앞인 경우 해당 트윗은 리플라이라 불리며(예: A가 작성한 트윗이 @B 트윗 내용인 경우) 사용자 A와 사용자 B의 커뮤니케이션으로 고려되어, 사용자 A와 사용자 B를 모두 팔로하고 있는 팔로어들에게만 전달이 되게 된다. 그러나 @의 위치가 맨 앞이 아닌 경우에는 해당 트윗을 남긴 사용자의 모든 팔로어의 타임라인에 전달된다. 이 경우를 멘션이라고 한다. 트윗 스레드(thread)의 관점에서 트윗을 분류하는 경우에는 @가 트윗의 맨 앞에 나온 경우에도 해당 트윗이 다른 트윗에 대한 대답이 아닌 경우 리플라이라 부르지 않기도 하지만, 이 연구에서는 트윗이 어느 사용자에게 전달되는지를 기준으로 트윗의 종류를 구분하였다.

(3) 트위터의 정보에 접근할 수 있는 다양한 API

몇 년 전 Web 2.0이라는 용어가 사용되기 시작했을 기반 기술들 중 하나로 Open API⁴⁾를 꼽는 경우가 많았다. Open API는 웹 서비스에 기록되는 사용자 데이터를 웹 서비스에 접근해야만 확인할 수 있는 것이 아니라, 인터페이스를 공개해서 웹 서비스의 외부에서도 서비스 내부의 데이터를 읽고 쓰게 해주는 기술들의 총칭이다. 특히 트위터는 다양한 종류의 API를 제공하고 있어서⁵⁾ 다른 서비스와의 연계가 매우 활발한 편이다. New York Times, Huffington Post, CNN 등의 언론 사이트에서는 각 뉴스 기사에 트위터 API를 연동해놓아서 각 사용자들이 간단한 버튼 클릭을 통해 해당 기사에 대한 링크를 남길 수 있게 되어 있다. 또한 iPhone 등의 모바일 애플리케이션 등에서도 듣고 있는 음악에 대한 정보, 현재 방문하고 있는 레스토랑의 정보 등을 버튼 클릭 한 번만으로 트윗을 쉽게 남길 수 있게 되어 있다. 이는 트위터 외부에서 트위터로의 정보 유입이 쉽게 일어날 수 있다는 것을 의미하며, 앞에서 말했듯이 이러한 트윗들은 해당 사용자를

4) http://en.wikipedia.org/wiki/Open_API

5) <http://dev.twitter.com> 에 다양한 API 들의 정의와 설명, 용례 등이 정리되어 있다

팔로하고 있는 모든 팔로어들에게 자동으로 전달된다.

(4) 트윗의 재전송 기능, 리트윗

앞서 말한 것처럼 한 사용자가 트윗을 작성하는 경우 그 사용자를 팔로하고 있는 팔로어들에게 해당 트윗이 전달되는데, 트윗을 전달 받는 팔로어들은 다양한 목적으로 해당 트윗을 다시 자신들의 팔로어들에게 재전송할 수 있다. 이 경우 마치 입소문이 퍼지듯이 트윗의 작성자로부터 그 팔로어로, 다시 그 팔로어의 팔로어로 트윗이 전달되어 갈 수 있다. 이렇게 트윗을 재전달하는 행위를 리트윗(Re + Tweet)이라고 하는데 재전달하는 횟수의 제한이 없기 때문에 재전달이 반복되면 점점 많은 사람이 트윗을 접하게 된다. 버튼 클릭 한 번만으로 자신의 팔로어의 트윗을 팔로어들에게 재전달 하는 것이 가능하며, 이런 편의성은 트위터에서의 빠른 정보 전파를 가능하게 한다. 홍수 속보나 교통 상황 등의 소식은 리트윗을 통해 트위터 내에서 빠르고 넓게 전달될 수 있다. 이 연구에서는 수집한 트윗을 바탕으로 리트윗의 효과를 속도와 범위의 측면에서 분석하고 정보 전파의 매체로서의 트위터의 가능성을 보이려고 한다.

3. 연구 방법론

이 장에서는 트위터의 미디어적 특성을 정량적으로 분석하기 위해 데이터를 수집한 방식을 소개하고, 수집된 데이터의 상세를 설명한다. 이 연구에서 사용된 데이터는 사용자들의 팔로 관계, 사용자들이 공개해놓은 개인 정보, 그리고 사용자들이 남긴 트윗이다.

1) 사용자들의 온라인 관계

트위터의 소셜 그래프 API(Social Graph API)⁶⁾를 이용하면 한 사용자의 팔로이 리스트와 팔로어 리스트를 얻을 수 있다. 트위터의 사용자 ID는

단조 증가하는 정수⁷⁾이기 때문에 스노우볼 샘플링 등의 방법을 통하지 않고도 사용자들의 관계 정보를 쉽게 수집할 수 있다. 이 연구에서는 2009년 6월부터 3개월간 약 4200만명의 트위터 사용자들의 14억 7천만 개의 팔로 관계를 수집하였다. 데이터 수집에는 20대의 서버를 사용했으며 각 서버는 트위터의 화이트 리스트(White list)⁸⁾에 등록되어 시간당 최대 2만 번의 쿼리를 트위터 서버측으로 전송하였다. 화이트 리스트에 서버가 등록되지 않는 경우에는 서버당 한 시간에 150~350 번의 쿼리 밖에 전송할 수 없으므로 이 연구와 마찬가지로 정보를 대량으로 수집하려는 경우에는 화이트 리스트 등록 과정이 필수적이다.

2) 사용자들의 개인 정보

사용자들의 개인 정보는 트위터의 유저 API(User API)⁹⁾를 사용하여 수집하였다. 사용자들이 트위터에 등록해 놓은 개인 정보는 사용자들의 트위터 페이지의 배경색, 배경 이미지, 사용자 이름, 짧은 프로필, 사용자의 사진, 사용자의 위치, 사용자가 있는 시간대 등을 포함한다. 이 때 사용자의 위치는 사용자가 입력한 정보에 기반하고 있어 특정 공통 포맷을

6) <http://dev.twitter.com/doc/get/friends/ids> 와 <http://dev.twitter.com/doc/get/followers/ids>에 정의되어 있는 Social Graph API의 형식에 맞춰 팔로이와 팔로어 리스트를 조회하기 원하는 사용자의 아이디(ID)를 HTTP 질의의 인자(parameter)로 넘기면 HTTP 응답으로 받을 수 있다.

7) 모든 사용자는 내부적으로 관리되는 정수(integer) 아이디를 갖게 되는데 이 아이디는 시간에 따라 1씩 차례로 증가하게 된다. 그러므로 모든 사용자들의 관계 정보를 얻기 위해서는 Social Graph API의 인자로 사용자 ID를 1씩 증가시키며 질의를 하면 된다. 모든 사용자에 대해 중복 없이 팔로어와 팔로이 리스트를 질의할 수 있다.

8) 트위터는 자사의 white list에 등록되어 있지 않은 서버들에 대해서는 서버당 질의의 수를 제한하는 방법으로 트래픽을 조절하고 있다. White list에 서버를 등록하기 위해서는 API를 사용하는 이유 등을 기재한 신청서를 트위터에 온라인으로 제출해야 하며 트위터 내부 규정에 따른 심사 과정을 통해 white list에 등록되게 된다. 자세한 정보는 <http://dev.twitter.com/pages/rate-limiting>에서 얻을 수 있다.

9) <http://dev.twitter.com/doc/get/users/show> 사용자의 어떤 개인 정보를 얻을 수 있는지 소개되어 있다.

따르고 있지 않으므로 정보를 해석하는데 일반적으로 어려움이 있으므로 시간대 정보만으로 충분한 경우에는 그 방법이 편리하다. 또한 사용자의 트위터 가입일, 남긴 트윗의 수, 즐겨찾기의 수 등 사용자의 트위터 활용 상황 정보도 포함한다. 이러한 정보들은 사용자들의 온라인 관계에 더해 사용자 개인의 특성을 보다 명확하게 파악하는데 도움을 준다.

3) 사용자들이 남긴 트윗

트위터에서 사용자들이 남기는 트윗의 많은 부분은 요점이 없는 잡담(Pointless babbles)이라는 연구가 2009년에 발표된 바 있다(PearAnalytics, 2009). 그런 트윗을 모두 포함한 대용량의 데이터를 통해 정보의 전파를 연구하는 것은 데이터의 처리 시간과 저장 용량의 문제가 발생할 수 있으므로, 이 연구에서는 그런 종류의 잡담을 제외하 나머지 트윗들로 대상을 한정하려고 한다. 트윗의 수집 범위를 한정하는데 사용한 방법은 트위터에서 발표하는 트렌딩 토픽(Trending topic) 리스트이다. 이 리스트는 현재 트위터에서 가장 많이 언급되고 있는 상위 10개의 단어 또는 연속된 몇 개의 단어들을 트위터에서 짧은 시간 간격으로(약 5분) 집계하여 발표하는 것이다. 즉 이 리스트에 포함되어 있는 단어들을 포함한 트윗은 단순히 요점이 없는 잡담이라기보다는 현재 많은 사람들의 관심의 대상이 되는 특정 토픽에 대한 느낌이나 의견 등을 담고 있으므로 해당 토픽에 대한 정보가 어떤 식으로 전파되는지에 관한 연구를 위해 보다 유용하게 사용될 수 있다.

트렌딩 토픽 리스트에 포함된 단어들을 언급하고 있는 트윗들을 모으기 위해 두 단계로 동작하는 데이터 수집기(Crawler)를 구현하였다. 먼저, 트렌딩 토픽 API¹⁰⁾를 이용해 5분 간격으로 트렌딩 토픽을 수집하였다. 그리고 수집한 트렌딩 토픽들을 포함하는 트윗을 검색 API(Search API)¹¹⁾로

10) <http://dev.twitter.com/doc/get/trends/current>

지속적으로 검색하고 수집하였다(현재는 트위터의 스트리밍 API(Streaming API)¹²)를 사용하는 것이 더 편리하다). 이 때 마지막으로 트렌딩 토픽 리스트에 포함된 시점으로부터 2주일 간 계속해서 토픽을 언급하는 트윗들을 수집하였다. 결과적으로 4,262개의 트렌딩 토픽에 대해 1억 6백만개의 트윗을 수집할 수 있었다. 수집된 자료 중 프라이버시 문제로 부터 자유로운 데이터들은 연구 목적을 위해 프로젝트 홈페이지¹³)를 통해 공유가 이루어지고 있다.

4. 연구 결과: 트위터 네트워크의 구조적 특성

이 장에서는 단방향 팔로 관계로 구축된 트위터 네트워크의 구조적 특성을 알아본다. 먼저 단방향 관계의 상호성(Reciprocity)을 알아보고, 다음으로 네트워크에서 각 사용자들이 갖는 팔로어 수의 분포를 살펴본 뒤 이러한 특성들이 트위터 내에서의 정보 전파에 어떻게 작용하는지를 알아본다.

1) 단방향 관계의 상호성

트위터 사용자들 사이의 팔로 관계가 일반적인 온라인 소셜 네트워크의 양방향 관계와 다른 단방향 관계라는 것은 앞에서 서술한 바 있다. 하지만 두 사용자 간에 서로 팔로를 하는 한 쌍의 단방향 관계는 온라인 소셜 네트워크의 양방향 관계와 비슷한 특성을 갖게 된다는 것을 알 수

11) <http://dev.twitter.com/doc/get/search>

12) http://dev.twitter.com/pages/streaming_api 트위터의 스트리밍 API는 특정 쿼리를 일정 시간 간격으로 반복해서 검색 결과를 얻을 때 유용하게 사용할 수 있는데, 데이터로 수집하는 시점에는 트위터의 스트리밍 API가 공개되어 있지 않았다.

13) <http://an.kaist.ac.kr/traces/WWW2010.html>

있다. 극단적으로 트위터의 모든 단방향 관계에 대해 쌍을 이루는 대칭적인 관계가 존재한다면 트위터 네트워크는 일반적인 온라인 소셜 네트워크의 구조와 비슷하게 된다. 그렇다면 실제 트위터 네트워크는 얼마나 양방향적인 특성을 갖고 있을까?

이것을 정량적으로 측정하는 방법이 네트워크의 상호성을 측정하는 것이다(Wasserman and Faust, 1994). 단방향 네트워크의 상호성은 다음과 같이 계산한다.

$$r = \frac{\text{두 단방향 관계가 짝을 이루고 있는 사용자 쌍의 수}}{\text{관계가 존재하는 사용자 쌍의 수}}$$

예를 들어 A (->) B -> C 의 단방향 관계들을 갖는 네트워크가 있다면 관계가 존재하는 사용자 쌍의 수는 (A, B)와 (B, C)의 2이고, 두 단방향 관계가 짝을 이루고 있는 사용자 쌍의 수는 (A, B)의 1이므로 이 네트워크의 상호성 값은 $r=0.5$ 가 된다.

이 연구에서 수집한 14억 7천만 개의 트위터의 단방향 팔로 관계로 이루어진 네트워크에서 상호성 값을 계산한 결과 0.221 이라는 낮은 값을 얻을 수 있었다(Kwak et al., 2010). 이것은 온라인 관계가 형성되어 있는 사용자 쌍을 고르면 그 중 22.1%만 서로 팔로를 하고 있음을 의미한다. 이 값은 다른 온라인 소셜 네트워크의 커뮤니케이션에서 보고된 상호성 값들, 0.68의 플리커(Flickr)(Cha, Mislove & Gummadi, 2009), 0.84의 야후! 360(Yahoo! 360)(Kumar, Novak & Tomkins, 2006), 0.77의 싸이월드 방명록(Chun, Kwak, Eom, Ahn, Moon & Jeong, 2008)과 비교하면 매우 낮은 수치이다. 한국인 사용자에게 익숙한 싸이월드 서비스에서 제공되는 방명록을 사용한 사용자들의 메시지 교환은 트위터의 팔로 관계처럼 한 사용자가 다른 사용자에게 일방적으로 메시지를 남길 수 있는 방식임에도 불구하고 트위터보다 3배 이상 더 높은 상호성이 관찰되었다. 이는 트위터의 팔로 관계가 실제로 다른 온라인 소셜 네트워크에서 사용자들이 맺는 양방향성 관계와 다른 것은 물론이고, 싸이월드의

방명록에서 관찰되는 커뮤니케이션 관계와도 다름을 보여준다. 트위터의 단방향적인 관계가 서로 짝을 이루지 않고 대부분은 비대칭적으로만 형성되어 있다는 사실은, 트위터의 팔로 관계가 단순히 친구를 나타내는 사회적인 관계가 아니라, 트윗을 구독하기 위한 기능적인 용도로 많이 사용되고 있음을 보여준다고 할 수 있겠다.

2) 팔로어 수의 분포

다음으로 알아볼 트위터 네트워크의 구조적 특성은 사용자가 갖는 팔로어 수의 분포이다. 일반적인 온라인 소셜 네트워크에서 한 사용자와 관계를 맺고 있는 이웃(neighbor) 수의 분포를 그려보면 멱함수(power-law)를 따른다는 것은 잘 알려져 있다(Mislove, Marcon, Gummadi, Druschel & Bhattacharjee, 2007). 이웃 수의 분포가 멱함수 분포 함수를 따른다는 것은 이웃 수가(d)인 사용자의 수와 이웃 수가 $10x(d)$ 인 사용자의 수의 비율을 $1/k$ 라 할 때, 이웃이 $10x(d)$ 인 사용자의 수와 이웃이 $100x(d)$ 인 사용자의 수의 비율 역시 $1/k$ 로 유지되는 것을 의미한다. 이웃 수의 분포가 멱함수 분포를 따르는 경우에는 극단적인(매우 많거나, 매우 적은) 값을 갖는 사용자가 normal 분포의 경우보다 더 많이 존재하게 된다. 이웃 수의 분포를 그래프로 그리는 경우 이웃의 수가 많은 사용자가 그래프의 꼬리 부분에 위치하게 되기 때문에 일반적인 normal 분포의 꼬리보다 멱함수 분포의 꼬리가 두껍게 나타나게 되는데 이러한 현상을 두꺼운 꼬리 분포(heavy tail 또는 fat tail)이라고 한다.

그렇다면 트위터에서 사용자가 갖는 팔로어 수의 분포는 어떤 분포를 따르고 있을까? 이 질문에 대한 답을 위해 각 사용자가 갖는 팔로어 수를 <그림 1>에 표시하였다. 그래프의 y축은 상보 누적 함수(Complementary cumulative density function, CCDF)으로 표시되었다. 이것은, 어떤 한 점의 x 값 이상의 모든 x 값들이 갖는 비율을 누적한 값이며, 해당 x 값의 누적 분포(Cumulative density function, CDF) 값을 1에서 뺀 값과 같다. 예를 들어 그래프의 빨간 선이 $x=10^5$ 일 때 $y=10^{-5}$ 를

통과하고 있는데 이것은 팔로어를 10만명(10^5) 이상 갖는 트위터 사용자의 비율이 전체의 10^{-5} 를 차지한다는 뜻이다. 멱함수 그래프는 로그-로그(log-log) 평면에서 직선으로 표시되므로 네트워크의 이웃 수를 분포할 때는 일반적으로 로그-로그(log-log) 평면에서 그래프를 그리게 된다.

〈그림 1〉에서 10만명에서 100만명 사이의 팔로어를 갖는 사용자들이 멱함수로 예측한 녹색 실선의 값보다 더 많이 존재함을 알 수 있다. 즉, 다시 말해 이 사용자들의 트윗은 자신의 팔로어로 전달되는 것만으로도 한 번에 10만명에서 100만명에게 전달된다는 것을 의미하고, 이렇게 많은 팔로어를 갖는 사용자들의 비율이 일반적인 온라인 소셜 네트워크의 특성에 따라 예측한 것보다 훨씬 높게 나타나는 것이다. 이것은 앞서 알아본 트위터의 단방향 관계가 갖는 특성이 가져온 결과라고 보는 것이 타당하다. 상대적으로 쉽게 온라인 관계를 형성할 수 있는 팔로어의 특성에 따라 트위터의 사용자들이 기존에 관심이 있었던 정치가(예: @BarackObama: Barack Obama, 2011년 1월 1일 현재 팔로어 620만명, @Number10gov: UK Prime minister, 2011년 1월 1일 현재 팔로어 176

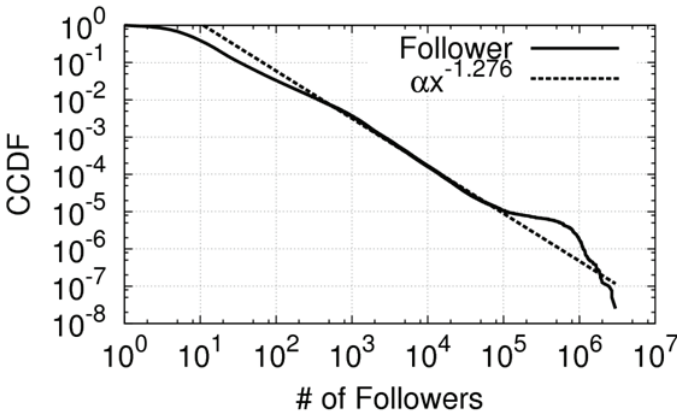


그림 1. 팔로어 수의 분포: 직선으로 표시된 멱함수 분포보다 더 많은 수의 허브가 존재함을 알 수 있다

만명) 또는 스타(예: @aplusk, Ashton Kutcher, 2011년 1월 1일 현재 팔로어 615만명, @britneyspears, Britney Spears, 2011년 1월 1일 현재 팔로어 648만명) 등을 팔로하고 그들의 트윗을 구독하고 있는 것이다. 이런 스타들 뿐만이 아니라 뉴스 기사를 트윗으로 남기는 매스미디어 계정들(예: @nytimes, New York Times, 현재 팔로어 282만명)도 많은 팔로어를 갖는 계정 중에 하나이다. 즉, 팔로어를 많이 가진 허브 유저에 의한 정보 전달이, 더 많은 허브 유저에 의해 일어날 수 있는 것이 트위터의 특징이다.

3) 트위터의 네트워크 구조적 특성 요약

트위터의 팔로 관계는 일반적인 온라인 소셜 네트워크의 양방향 친구 관계와는 다르게 관계의 형성을 위한 상대의 동의가 필요 없고, 또한 관계 역시 비대칭적이라는 특징이 있다. 이러한 특징들은 일반적인 온라인 소셜 네트워크와는 차별되는 트위터 네트워크의 구조적 특징을 가져왔다. 하나는 낮은 값의 상호성이고 다른 하나는 많은 수의 허브(hub) 유저들이다. 두 구조적 특징 모두 트위터가 기존 온라인 소셜 네트워크 서비스와는 다른 특성을 갖고 있음을 보여준다. 기존의 온라인 소셜 네트워크 서비스들이 사용자들 사이의 친한(social) 관계를 바탕으로 오프라인에서 잘 알던 사용자들 간의 관계 유지에 초점을 두고 있다고 한다면, 트위터에서는 자신이 기존에 관심이 있었던 사용자 또는 기업, 언론 계정들을 팔로하며 트윗을 구독하는 특성을 보인다. 트위터에서의 이러한 단방향 연결은 많은 팔로어를 가짐으로서 직접 정보 전파에 큰 역할을 할 수 있는 사용자들을 만들어낼 뿐만이 아니라, 오프라인에서의 또래 집단과 또래 집단 사이를 길게 이어주는 가늘고 긴 연결(Granovetter, 1973)의 특성을 가지게 되어 빠르고 넓은 정보 전파를 가능하게 한다.

5. 연구 결과: 트위터에서의 정보 전파

이 장에서는 세 달 동안 수집한 트렌딩 토픽들을 언급한 트윗들을 대상으로 트위터에서 활발히 전달되는 정보의 종류와 성격을 알아보려고 한다. 그리고 리트윗의 형태로 입소문(word-of-mouth) 형태로 퍼지는 정보 전파의 규모와 속도에 관해 논의할 것이다.

1) 트렌딩 토픽의 유형

세 달 동안 수집한 4200 여개의 트렌딩 토픽 중에서 해당 트렌딩 토픽을 언급한 트윗이 많은 순서대로 100개의 트렌딩 토픽을 골라 New York Times의 기사 분류를 따라 토픽을 분류한 결과를 <표 1>에 정리하였다. 그 결과 영화나 연예인들의 가십 기사를 포함한 arts 분류가 35개의 트렌딩 토픽을 포함하는 가장 큰 분류였다. 이하 토픽이 포함된 순서를 따라 technology(19), sports(11), world(9), style(7), U.S. region(7), politics(4), health(3), weather(2)의 분류에 트렌딩 토픽이 속하는 것을 알 수 있었다. 아무런 분류에도 속하지 않은 트렌딩 토픽은 'RT'와 같은 트위터에서 기능적인 요소를 담당하는 단어로서 정보 전파의 역할을 하는 단어가 아니었으므로 제외하였다.

트렌딩 토픽은 단어 한 개 내지는 두 개로 이루어져있기 때문에 동일하게 지역을 지칭하고 있더라도 문맥에 따라 분류가 달라지게 된다. 그러므로 해당 토픽이 왜 트렌딩 토픽으로 선정되었는지를 충분히 이해할 필요가 있는데, 트렌딩 토픽과 기간, 그리고 트렌딩 토픽으로 선정된 이유를 잘 정리하고 있는 웹 사이트¹⁴⁾의 자료를 참고하여 분류를 수행하였다.

이러한 분류 과정을 통해 트위터에서 사용자들이 많이 이야기하는 주제들은 대부분 오프라인의 이슈와 밀접한 관련이 있음을 알 수 있었

14) <http://www.whatthetrend.com> 해당 토픽이 왜 트렌딩 토픽이 되었는지를 위키 형식으로 누구나 편집할 수 있게 정리하고 있는 서비스이다.

표 1. 트렌딩 토픽의 분류

분류	분류에 속한 토픽의 수	대표 토픽의 예
Arts	35	Harry potter, Michael Jackson, True blood, American idol
Technology	19	Apple, iphone, Blackberry, Gmail, Bing, Facebook
Sports	11	Football, Brazil, England, Lakers, Miami
World	9	Iran, #iranelection, France, Air France, Kim, Tehran
Style	7	Father's day, Starbucks, Breakfast
U.S. region	7	New York, Texas
Politics	4	Obama, President, President Obama
Health	3	Mexico, Atlanta(*신종플루)
Weather	2	Rain

다. 이전 연구(Kwak et al., 2010)에서 트렌딩 토픽들 중 85.8%는 해당 토픽을 언급한 전체 트윗의 50% 이상이 트렌딩 토픽으로 떠오른 날 쓰여짐을 보였다. 이 결과를 바탕으로 크레인과 소넷(Crane & Sornette, 2008)의 연구내용에 따라 트렌딩 토픽의 인기 곡선을 기준으로 토픽들을 분류한 결과 뉴스 속보와 같이 갑자기 등장하여 빠르게 인기를 얻고 다시 빠르게 인기를 잃는 트렌딩 토픽(headline news)이나 꾸준한 인기를 얻고 오래 지속되는 트렌딩 토픽(persistent news), 순간적인 피크를 갖는 트렌딩 토픽(ephemeral) 등을 골라낼 수 있었다.

이 연구에서는 트위터와 같은 소셜 미디어 서비스에서 트렌딩 토픽에 대한 사용자들의 상호 작용을 관찰한 결과를 덧붙이려고 한다. 해당 트렌딩 토픽을 포함한 트윗을 남긴 사람 수를 기준으로 상위 20개의 토픽에 대해 각 트윗들의 목적에 따른 분류(리플라이, 리트윗, 멘션)의 세 분류와, 이 세 분류에 포함되지 않은 나머지는 모두 싱글톤(Singleton)이라는 분류로 정의)한 결과가 <그림 2>이다.

<그림 2>에서 눈에 띄는 것은 높은 비율의 싱글톤 트윗들이다. 이

트윗들은 리플라이(Reply)나 멘션(Mention)처럼 다른 사용자와의 인터랙션을 위해 쓰여진 것이 아니라 트윗을 작성한 사용자의 팔로어에게 전달되는 역할을 수행한다. 친교 목적이 아니라 트윗 구독을 위한 온라인 관계가 형성되는 것으로 미루어볼 때 사용자간의 인터랙션이 아니라 트렌딩 토픽에 대한 정보 전달을 위한 트윗, 또는 정보 전달 역할을 하는 트윗이 많이 작성되는 것은 예측 가능한 현상이다. 재미있는 것은 높은 비율의 리플라이 트윗이다. 리플라이는 전적으로 상대방과의 의견 교환 등의 인터랙션을 위한 트윗이다. 이것은 대부분의 트윗이 요점이 없는 잡담(Pointless babbles)이라고 했던 PearAnalytics의 연구(2009)에서는 부각되지 않았던, 오프라인에서의 이슈와 같은 트렌딩 토픽 들에 대해서는 사용자 사이에서의 상호작용을 하기 위한 트윗도 많이 작성됨을 의미한다. 다음으로 주목해서 살펴볼 것은 리트윗을 하는 트윗들이 차지하고 있는 비율이다. 각 토픽들에 대해서 약 5~15% 정도의 비율을 차지하고 있는 리트윗들은 각각 원래의 트윗을 자신들의 팔로어들에게 다시 한 번 재전송하는 목적을 갖고 작성된 것이다. 또한 각 토픽의 성격에 따라 트윗의 내용을 재전송하려는 리트윗의 비율이 다른 것이 발견되었다.

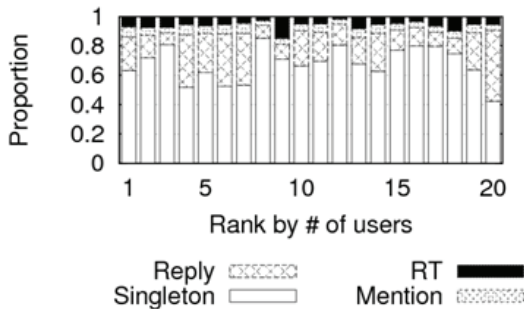


그림 2. 상위 20개의 트렌딩 토픽을 포함하는 트윗들의 목적에 따른 분류 비율

2) 리트윗을 통한 빠른 정보 전파

그렇다면 리트윗은 실제로 얼마나 빠르고 넓게 정보를 전파할까? 우선 리트윗에 의한 정보 전파 네트워크의 모양을 파악해보자. <그림 3>은 'hcr'(Health care reform, 2009년에 있었던 오바마 대통령의 의료 보험 개혁과 관련한 트윗)을 포함하는 트윗들이 리트윗을 통해 전파된 네트워크 구조이다. 네트워크의 각 노드는 서로 다른 사용자를 나타내고 노드를 연결하는 링크는 리트윗을 통해 정보가 전달된 것을 의미한다. 링크의 방향은 정보의 전파 방향을 나타내며 링크의 색은 각각 다른 트윗을 의미한다. 높은 비율의 최대 연결 성분(Giant Connect Component, GCC)과, 많은 수의 짧은 리트윗 트리는 이전 연구(Kwak et al., 2010)에서 보였던 'air france' 리트윗 네트워크와 비슷한 형태이다.

<그림 3>의 아래에 위치하고 있는, 최대 연결 성분(GCC)에 포함되지 않은 connected component들 중에는 길이가 1짜리 리트윗 tree들을 많이 찾아볼 수 있다. 즉, 정보의 소스(source)가 트윗을 남겼고, 그 소스의 팔로어중 하나가 해당 트윗을 리트윗한 경우인데 이 경우 리트윗 tree의 길이는 1이지만 해당 트윗에 노출되는 사용자의 수는 두 사용자의 팔로어 수의 합이 된다.

이전 연구(Kwak et al., 2010)에서 사용자가 갖는 팔로어가 1,000명 이하인 사용자들에 대해 팔로어 수와 무관하게 그들의 트윗이 리트윗되는 경우 수백명의 추가적인 사용자들에게 트윗이 전달됨을 밝혔다. 이는 많은 수의 팔로어를 갖지 못한 일반 사용자들도 입소문(Word-of-mouth)을 통하면 정보를 널리 퍼뜨릴 수 있는 환경이 갖추어져 있다는 것을 의미한다.

리트윗에 의한 정보 전파 역시 굉장히 빨리 일어나는데 모든 트렌딩 토픽들에 대해, 50%의 리트윗은 원래의 트윗이 쓰여진 시간으로부터 1시간 이내에, 그리고 75%의 리트윗이 하루 이내에 일어남을 밝혔다(Kwak et al., 2010). 이렇게 대부분의 리트윗이 원래의 트윗이 쓰인 지 얼마 안 되는 시간 안에 활발히 일어나는 이유에 대해서는 두 가지 가능

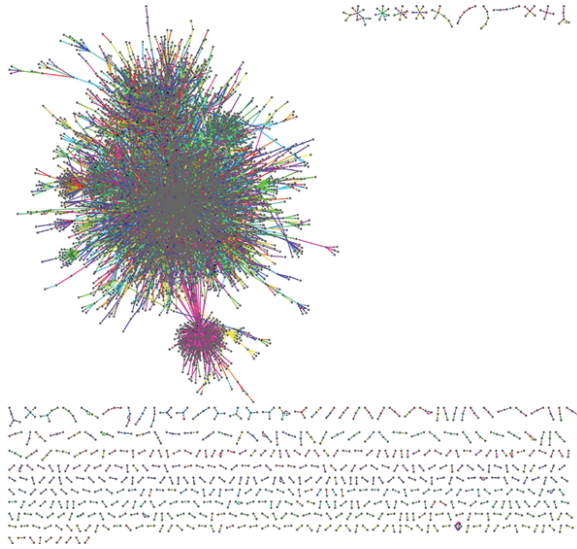


그림 3. '#hcr'을 포함하는 트윗들의 리트윗 시각화: 각 노드는 유저를 나타내고 노드를 연결하는 링크는 리트윗에 의한 정보의 전파를 나타낸다. 링크의 색(color)은 각기 다른 트윗을 나타낸다

성을 생각해볼 수 있다. 첫 째는 정보의 가치가 시간이 감에 따라 빠르게 떨어지는 경우이다. 어떤 사용자가 자신의 팔로이들이 적은 트윗을 읽고 자신의 팔로어들에게도 그 트윗을 보여주고 싶을 때 리트윗을 하게 된다. 그런데 만약 트렌딩 토픽이 특히 속보성 뉴스에 관한 것이라면 그 사건이 발생한 시점에는 자신의 팔로어들이 아직 그 소식을 알지 못할 것이라는 생각으로 리트윗을 활발히 하게 되겠지만, 어느 정도 시간이 흘러 어느 정도 그 뉴스가 알려진 이후라는 생각이 들면 더 이상 그 트윗을 리트윗하는 일은 없어질 것이다. 이 시나리오는 트렌딩 토픽들 중 약 85%가 해당 토픽을 언급하는 트윗들이 빠르게 늘어나고 또 빠르게 줄어든다는 사실로부터 설득력을 얻는다. 두 번째 가능성은 타임라인의 한정된 길이 때문에 트윗이 노출되는 시간이 정해져 있다는 것이다. 사용자

들은 일반적으로 어느 정도의 시간을 사이에 두고 트위터에 로그인하여 타임라인에 쌓여 있는 트윗들을 보게 된다. 이 때 사용자가 얼마나 거슬러 올라가며 트윗을 확인하느냐에 따라 그 사용자가 과거 어느 정도 시간까지의 트윗을 볼 수 있느냐가 달려 있다. 사용자가 읽지 못하는 트윗은 리트윗될 수도 없으므로 시간이 지나감에 따라 한 트윗이 리트윗될 확률은 자연스럽게 감소한다 할 수 있다. 두 가능성 중 어느 쪽이 더 많은 경우를 설명하는지에 대해서는 좀 더 사용자 중심의 연구가 필요하다.

위의 결과는 모든 트렌딩 토픽에 대한 리트윗의 효용을 분석한 것이지만 앞에서 언급한 대로 트렌딩 토픽의 인기 곡선에 따른 분류를 적용하여 트렌딩 토픽의 종류와 정보의 전파의 관계에 관한 연구도 향후 수행할 예정이다.

3) 트위터에서의 정보 전파 특성 요약

사용자들이 많이 언급한 상위 100개의 트렌딩 토픽들의 분석을 통해 실제 오프라인에서의 이슈들이 트위터에서 많이 이야기되는 것을 밝혔다. 이전 연구(Kwak et al., 2010)를 통해 많은 트렌딩 토픽들이 빠르게 인기를 얻고 그만큼 빠르게 인기를 잃는 것을 밝힌 바 있으며 그것은 트위터 내에서의 화제가 지속적으로 빠르게 바뀐다는 것을 의미한다. 또한 트렌딩 토픽들을 언급하는 트윗들 중에서는 리플라이의 비중이 높은 것을 보였는데, 이것은 소셜 미디어 서비스들에서 사용자 간 인터랙션을 통해 콘텐츠가 소비되는 경향을 보여준다. 기존의 매스 미디어가 일방적인 정보 전달에 중심을 두던 것과는 달리, 트위터와 같은 온라인 관계를 기반으로 한 서비스들에서는 사용자 간의 메시지 교환을 통해 해당 정보에 대한 의견 교환이 일어날 수 있으며 이것이 소셜 미디어 서비스에서 정보가 소비되는 특성이라 할 수 있다.

리트윗을 통한 정보 전파 네트워크에서는 많은 사람들에게 리트윗을 당하는 몇몇 소스의 존재를 확인했고 짧은 길이의 리트윗 트리 역시 많이 확인할 수 있었다. 대형 언론 계정들의 정보가 아니라, 일반 사용자들의

트윗을 리트윗하는 짧은 길이의 리트윗 트리들은 실제로 트위터에서의 정보 전파가 허브 노드에만 의존하는 것이 아니라 지엽적인 많은 정보 소스로부터 동시 다발적으로 퍼져나간다는 것을 확인할 수 있게 한다. 이전 연구(Kwak et al., 2010)에서 사용자가 갖는 팔로어가 1,000명 이하인 사용자들에 대해 팔로어 수와 무관하게 그들의 트윗이 리트윗되는 경우 수백 명의 추가적인 사용자들에게 트윗이 전달됨을 밝혔다. 이는 많은 수의 팔로어를 갖지 못한 일반 사용자들도 입소문(Word-of-mouth)을 통하면 정보를 널리 퍼뜨릴 수 있는 환경이 갖추어져 있다는 것을 의미한다. 그리고 리트윗들에 의한 정보 전파는 굉장히 빨리 일어난다는 것을 밝혔는데, 이 연구에서 그러한 현상이 발생하는 두 가지 가능성에 대해 논의하고, 각 가능성들을 정확히 검증하기 위해서는 사용자 중심의 정성적인 트위터 연구의 노력이 더 필요하다는 제언을 하였다.

6. 향후 연구의 방향

정보 전파가 쉽게 일어나는 트위터를 이용해 어느 사용자가 정보 전파에 있어서 큰 영향력을 발휘하는지에 관한 연구가 활발히 진행되는 것은 자연스러운 일이다. 이전 연구(Kwak et al., 2010)에서 팔로어의 수와, 트위터 네트워킹으로부터 계산한 페이지랭크(PageRank)와, 리트윗된 횟수로 각각 계산한 사용자의 순위가 각기 다름을 보인 바 있다. 뒤이어 비슷한 시기에 트위터에서 사용자가 갖는 팔로어 수는 그 사용자의 트윗이 리트윗된 횟수와는 관련이 없다는 정량적인 연구도 발표된 바 있다(Cha, Haddadi, Benevenuto & Gummadi, 2010). 또한 사용자의 입장에서 어떤 팔로이로부터 전달되어 온 정보를 가장 신뢰하는지에 관한 화이트페이지(Vocus, 2010)들도 출간되어 있는 상태이다.

특히 이 분류에 속하는 연구들을 수행할 때 주의해야 할 점은 영향력이라는 용어의 범위를 명확하게 정의하는 일이다. 이 때 몇 가지 요소

들을 고려할 수 있다. 우선 첫 째는 정보 전달의 현상을 관찰하는 방법에 관한 것이다. 트위터에서 가장 명시적으로 한 사용자에게서 다른 사용자로 정보가 퍼져 나가는 것을 보여주는 것은 리트윗이다. 하지만 리트윗은 정보의 전달을 측정하는 수단으로 사용하기에는 단점이 몇 가지 있는데, 첫 째로는 네트워크의 형태에 따라서 어떤 리트윗은 이미 정보를 다 알고 있는 사람들에게 중복된 정보를 전달하는 역할만을 할 수 있다는 것이다. 이 경우 그 리트윗은 정보 전달의 관점에서 새로운 사용자에게 정보를 전달하지 못한, 그야말로 있으나 없으나 한 트윗이 되게 되므로 이런 리트윗들로부터 활발한 정보 전달이 일어났다는 결론을 얻는 것은 어렵다고 할 수 있다. 두 번째 단점은 리트윗은 비록 트윗을 읽은 사용자들이 '더 널리 알려야 할 필요가 있는' 정보라고 수동적인 분류를 한 것이지만 그러한 과정이 항상 자신이 원래 팔로하고 있던 사용자들의 트윗에 대해서 이루어진다는 점은 편향된 분류가 일어날 가능성이 있다는 것이다. 예를 들면 어떤 유명 스타의 트윗은 중요한 정보를 담고 있지 않음에도 불구하고 그 스타의 팬에 의해 리트윗될 수 있다. 세 번째 단점은 트위터에는 수동적인(Passive) 사용자가 많음에도 불구하고 리트윗 여부를 관찰하여 정보가 퍼진 것으로 관찰한다면, 실제로 해당 트윗을 전달 받아 읽었지만 리트윗을 하지 않은 많은 수의 수동적인 독자(reader)들을 고려하지 않는 방법이 된다는 것이다. 리트윗을 통해 정보 전달을 측정하는 것은 명확한 방법이기는 하지만 위와 같은 단점이 있으므로 연구 목적에 맞게 정보 전달을 정의하고 그에 맞춰 어느 사용자가 그러한 유형의 정보 전달에 가장 큰 영향을 끼쳤는지를 분석하고, 명확하게 진술할 필요가 있다. 두 번째로 고려해야 할 요소는 전파되는 정보의 유형에 관한 것이다. 전파되는 정보가 어떤 분류에 속한 정보이냐에 따라 사용자의 영향력은 크게 달라질 수 있다. 물론 다양한 토픽에 대한 최신 소식을 골고루 잘 전달하는 New York Times와 같은 계정들도 있지만 IT 토픽에만 특성화된 테크크런치(@TechCrunch)나 매셔블(@Mashable)과 같은 계정들도 있다. 이것들은 쉽게 알 수 있는 경우이지만 같은 정치 뉴

스에 관해서라도 진보주의의 정보를 전달하는데 큰 역할을 하는 사용자와 보수주의의 정보를 전달하는데 큰 역할을 하는 사용자는 다를 수 있다. 이러한 세세한 정보의 분류에 따른 영향력의 변화를 정확히 알기 위해서는 트윗 콘텐츠의 기계적인 카테고리 분류만이 아니라 의미 분석도 필요하게 된다. 세 번째로 고려해야 할 요소는 영향력의 지속성에 관한 것이다. 사용자의 영향력을 평가한 결과를 마케팅 등의 다양한 응용 사례에 이용하기 위해서는 사용자의 영향력의 지속성에 관한 지식이 필요하게 된다. 이 지속성은 시간의 변화에 따른 지속성으로 생각하기 쉽지만 트위터 네트워크는 지속적으로 새로운 사용자가 유입되고 기존의 사용자가 빠져나가며, 동시에 새로운 온라인 관계도 생겨나고 기존의 온라인 관계가 없어지기도 하는 등 정보를 전달하는 경로 자체도 지속적으로 변화하기 때문에 트위터 내부 요소의 변화에 따른 영향력의 지속성도 중요하게 고려되어야 한다. 이러한 요소들을 고려하여 트위터의 정보 전파에 있어서 사용자의 역할, 팔로 관계의 역할과 같은 다양한 연구가 지속적으로 수행될 것이라 예상된다.

7. 결론 및 함의

이 연구에서는 싸이월드나 페이스북으로 대표되는 기존의 온라인 소셜 네트워크 서비스들과는 다른 트위터에서의 정보 전파 특성을 다루었다. 약 세 달의 기간 동안 사용자들의 온라인 관계, 사용자들이 트위터에서 공개해놓은 개인 정보, 그리고 사용자들이 남긴 트윗 데이터를 수집하였고, 개인 정보를 침해할 여지가 없는 범위 내에서 수집한 데이터의 일부를 연구 목적으로 공유하고 있음을 밝혔다.

먼저 트위터의 단방향 팔로 관계에 주목하였다. 단방향 팔로 관계는 트위터에서의 메시지 전달 기능과 밀접하게 결합되어 있는데, 실제로 트위터 사용자 간의 팔로 네트워크에서 상호성 값을 측정해 본 결과 여

타 온라인 소셜 네트워크 서비스에서 관찰되는 것보다 상당히 낮은 값이 관찰되었다. 이는 트위터 네트워크의 팔로 관계가 친교를 나타내고 기존의 관계를 유지하는데 사용되는 것보다, 트윗을 구독하는 용도로 사용되고 있음을 말해주는 정량적인 증거로 고려될 수 있다.

트위터의 단방향적인 팔로 관계는 다른 온라인 소셜 네트워크에서 공통적인 특성을 기반으로 예측한 것보다 더 많은 수의 허브(hub) 유저들을 탄생시켰다. 이러한 허브 유저들은 많은 팔로어를 가짐으로써 하나의 트윗을 작성하는 것만으로도 수십 수백만명에게 직접 정보를 전달할 수 있는 영향력을 갖게 되었다.

또한 많은 사용자들이 언급한 상위 100개의 트렌딩 토픽들은 실제 오프라인에서의 이슈들과 많은 관련이 있음을 보였다. 트렌딩 토픽을 언급하는 트윗들 중 높은 리플라이의 비중으로부터 소셜 미디어 서비스에서 사용자 간의 메시지 교환을 통해 기존의 매스미디어들의 단순 정보 배포와는 달리 해당 정보에 대한 의견 교환이 일어날 수 있으며 이것은 정보를 소비하는 방식의 변화로 보인다는 것을 보고했다.

리트윗을 통한 정보 전파 네트워크에서는 많은 사람들이 리트윗을 하는 몇몇 주요 정보 소스의 존재를 확인했고, 짧은 길이의 많은 리트윗 트리로부터 트위터에서의 정보 전파는 허브 노드에만 의존하는 것이 아니라 지엽적인 많은 정보 소스로부터 동시 다발적으로 퍼져나간다는 것을 확인할 수 있었다. 이전 연구를 통해 많은 수의 팔로어를 갖지 못한 일반 사용자들도 리트윗이라고 하는 입소문(Word-of-mouth)을 통하면 정보를 빠르고 넓게 퍼뜨릴 수 있는 환경이 갖추어져 있다는 것을 보였다.

즉, 이 연구를 통해 트위터에서는 오프라인의 이슈와 관련된 토픽들을 주제로 활발한 트윗이 기록되고, 많은 수의 허브는 그런 트윗을 몇십, 몇백만명에게 손쉽게 전달할 수 있으며, 비록 팔로어를 많이 갖지 못한 사용자들도 리트윗을 통하면 많은 사용자에게 자신의 트윗을 빠르고 넓게 퍼뜨릴 수 있다는 것을 정량적으로 설명하였다.

참고 문헌

- Cha, M., Mislove, A., & Gummadi, K. P. (2009). A measurement-driven analysis of information propagation in the Flickr social network. *Proceedings of the 18th international conference on World Wide Web*.
- Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. *Proceedings of international AAAI conference on weblogs and social media*.
- Chun, H., Kwak, H., Eom, Y-H., Ahn, Y-Y., Moon, S., & Jeong, H. (2008). Comparison of online social relations in terms of volume vs. interaction: A case study of Cyworld. *Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*.
- Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system, *Proceedings of the National Academy of Sciences*, 105(41), 15649~15653.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6), 1360~1380.
- Kumar, R., Novak, J., & Tomkins, A. (2006). Structure and evolution of online social networks. *Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining*.
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a new media? *Proceedings of the 19th international conference on World Wide Web*.
- Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*.
- PearAnalytics (2009). Twitter study - August 2009, white paper, <http://goo.gl/cCa2s>.
- Vocus (2010). Influencer grudge match: Lady Gaga versus Bono! - What makes an influencer: a survey by Vocus and Brian Soils, white paper, <http://goo.gl/xXGju>.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and*

applications. Cambridge University Press.

WikiTree (2010). 서울 ‘물난리’ 트위터 현장중계, <http://goo.gl/oWsg4>.

최초 투고일 • 2011. 01. 10

논문 수정일 • 2011. 02. 10

게재 확정일 • 2011. 02. 12