

TwitNet : 트위터 사용자들의 관계를 시각적으로 나타내는 Cytoscape 플러그인 개발

박지혜^{0*}, 김보현^{**}, 이명준^{***}, 권영근^{****}

^{*}울산대학교 컴퓨터·정보통신공학부

{^{0*}amunata, ^{**}panicist} @ mail.ulsan.ac.kr, {^{***}mjlee, ^{****}kwonyk} @ulsan.ac.kr

TwitNet : Cytoscape Plugin for Visualizing Relation between Twitter Users

Ji Hye Park^{0*}, Bo Hyun Kim^{**}, Myung Joon Lee^{***}, Yung Keun Kwon^{****}

^{*}University Of Ulsan Computer Engineering&Information Technology

요 약

웹 2.0의 기술이 보급됨에 따라 소셜 네트워크 서비스에 대한 관심이 증가하였다. 국내에서는 싸이월드, 미투데이 등과 같은 서비스가 널리 사용되고 있으며 최근 급부상한 트위터는 여러 분야에서 관심을 받고 있다. 트위터는 팔로워나 트윗 등 활동 정도에 따라 랭킹 서비스가 제공되고 있지만 랭킹은 그들 사이의 관계를 세부적으로 나타내지 못한다.

본 논문에서는 트위터의 사용자들 사이에 존재하는 관계를 시각적으로 나타내는 도구에 대해 개발한다. 국내 사용자 중 팔로워의 랭킹에 따른 사용자를 이용하고, 시각화를 위해 생물학적 데이터를 네트워크로 나타내는 Cytoscape 플랫폼을 사용한다. 사용자 간의 관계를 나타내는 네트워크를 통하여 온라인상에서 영향력 있는 사용자들의 관계를 나타내고 그들의 관계를 수치로 분석한다. 또한 복잡한 네트워크로부터 선택된 노드와 관련된 연결만을 추출하는 기능을 제공하여 온라인상의 관계를 상세하게 나타낸다.

1. 서 론

최근 웹 2.0, 웹의 기술이 소개 되었다. 웹 1.0은 정보를 제공하기 위한 기능만을 제공하였다면 웹 2.0에서는 사용자가 데이터를 직접 다룰 수 있도록 정보를 더 쉽게 공유하고 서비스할 수 있도록 구성되어 있다[1].

웹 2.0의 기술이 보급되고 소셜 네트워크 서비스(SNS : Social Network Service)에 관련된 관심이 널리 알려짐에 따라 페이스북과 마이스페이스를 이을 서비스로 최근 트위터(Twitter)[2]가 관심을 받고 있다. 트위터는 마이크로 블로그 서비스를 제공하는 무료 SNS이다. 블로그는 많은 글과 사진을 담을 수 있는 형식을 가지고 있는 반면 트위터는 짧은 메시지를 업로드 하는 형식이다. 또한 기존의 서비스는 웹을 통한 접근만을 제공하였다면 트위터는 웹뿐만 아니라 모바일 기기와 인스턴스 메시징으로도 이용이 가능해 실시간을 보장할 수 있는 장점이 존재한다[3]. 이러한 특성이 기존의 다양한 SNS가 존재함에도 불구하고 트위터가 유행되는 원인으로 볼 수 있다.

트위터의 유행으로 이를 이용한 각종 서비스들이 제공되고 있다. 트위터에 사진이나 동영상을 업로드 할 수 있는 트윗픽(twitpic)[4]와 트윗비드(twitvid)[5], 사용자들의 순위를 제공해 주는 트위터 카운터

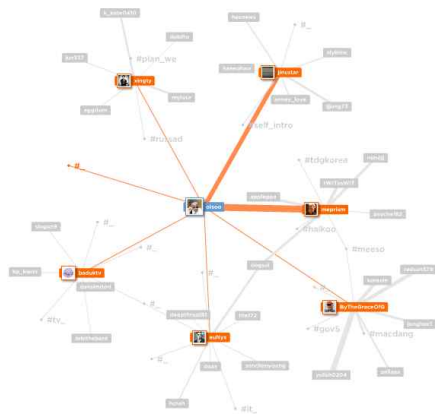
(twittercounter)[6]와 트위터 그레이더(twittergrader)[7] 등이 널리 사용되는 서비스이다. 이들 서비스는 트위터를 사용하는 사람들 중 영향력 있는 사용자가 누구인가를 파악할 수 있는 자료가 된다. 또한 트위터 사용자들 사이를 분석하고 있는 Mentionmap[8] 서비스도 있다. Mentionmap은 트위터 아이디 검색으로 친구들 사이의 댓글 빈도를 시각적으로 제공하고 있다. 그러나 기존의 서비스는 친구들 사이의 활동정도만을 나타내거나 팔로워의 수나 트윗의 개수, 활동 정도에 따른 순위를 제공한다. 영향력 있는 사람들 사이에 어떠한 관계가 존재하는지에 대해서는 알 수 없다.

본 논문에서는 트위터 사용자 중 영향력 있는 사용자들이 온라인상에서 어떻게 관계되고 있는가에 대해 시각적인 네트워크를 제공하고 분석하는 TwitNet을 구현한다. 이를 위해 트위터 카운터에서 서비스하고 있는 랭크를 이용하고 팔로워들 사이의 효과적인 시각화를 위해 생물학적 데이터를 시각적으로 보여주는 플랫폼 Cytoscape[9]가 도구로 사용된다. 본 논문의 구성은 2장에서 관련 연구를 설명하고 3장에서는 사용자들의 관계를 나타내고 분석하는 TwitNet의 설계와 구현에 대해 설명한다. 4장에서는 TwitNet을 이용하여 상위 100명들 사이의 관계를 분석하고 5장에서 결론 및 향후 연구에 대해 논한다.

2. 관련 연구

2.1 Mentionmap

트위터를 하다보면 팔로워들과 reply 또는 리트윗(retweet)으로 서로 트윗을 주고받게 된다. Mentionmap은 이러한 트윗들을 누구와 주고받는지를 실시간으로 보여주는 서비스이다. 내가 관계를 맺고 있는 사람이 또 다른 어떤 이와 관계를 맺고 있는지 볼 수 있고 리트윗을 주고받은 수치를 예지의 두께로 나타내어 준다. [그림 1]은 사용자 oisoo에 대한 Mentionmap을 나타낸다.



[그림 1] Mentionmap 사용의 예

Mentionmap은 검색하려는 사용자 이름을 입력 값으로 하여 보여준다. 이것은 검색하는 사용자와 관계되어 있는 사용자의 관계만 보여준다는 한계점이 있다.

2.2 Twitter Counter

트위터에는 자신이 알고자 하는 사람에게 팔로워를 맺는 '친구(friend)', 자신에게 관계를 맺은 '팔로워(follower)'의 관계가 있다. 이러한 트위터에서 생성되는 관계나 활동 정도에 따라 랭킹을 제공해 주는 서비스들이 나타났다. 그 중 트위터 카운터는 트위터에서 발생한 팔로워나 친구, 트윗의 수치에 따라 통계를 그래프로 표현해 주고 랭킹 서비스를 제공한다. 랭킹 서비스는 전세계 사용자를 대상으로 제공하기도 하고, 각 국가별 사용자를 대상으로 제공하기도 한다.

[그림 2]는 트위터 카운터에서 제공되는 통계의 예시를 보여준다.



[그림 2] 트위터 카운터의 팔로워 통계

2.3 Cytoscape

Cytoscape는 분자 관계 네트워크와 유전자 표현을 시각적으로 나타내는 생물정보학을 위한 오픈 소스의 소프트웨어이다. Cytoscape는 생물학적 데이터를 가지는 GML, XGMML, SBML 등의 마크업 언어(Markup Language)를 지원하며 텍스트파일 및 엑셀 파일도 지원한다. 추가적으로 플러그인을 제공하여 Cytoscape에서 추가적인 기능이 가능하도록 한다. 플러그인 개발은 Cytoscape 오픈 자바 소프트웨어 아키텍처를 이용하여 개발할 수 있다. Cytoscape 플러그인의 기능을 이용하면 생물학적 데이터 외에 추가적인 데이터를 시각화하고 데이터 분석을 위한 도구로서 사용할 수 있다.

3. TwitNet의 설계와 구현

트위터에서는 많은 사용자가 온라인상에서 관계를 맺고 상호작용한다. 사용자들의 활동정도나 친구 수에 관련된 많은 랭킹 서비스가 제공되고 있지만 관계를 네트워크로 나타낸 서비스는 많지 않다.

3장에서는 영향력 있는 사용자를 대상으로 그들 사이에 나타나는 친구 관계를 네트워크로 나타내고 분석하는 도구 TwitNet을 설계하고 구현한다.

3.1 트위터 API 제한 정책

트위터 API는 클라이언트에게 주어진 시간동안 제한된 횟수의 요청이 가능하도록 한다. 트위터 API에는 두 가지 API가 존재한다. REST API와 Search API이다. REST API는 계정기반, IP기반의 제한을 실행하며 REST API를 위한 기본적인 제한은 시간당 150개의 요청이다. 이러한 제한을 피하기 위해서는 트위터에서 제공하는 Whitelisting을 통해 시간당 2만개의 요청을 허락받아야 한다.

3.2 데이터 수집

트위터는 세계적으로 하루 30만 명이 사용자로 신규 가입하고, 전 세계적으로 사용자가 1억 명을 돌파하였다. 최근 샌프란시스코에서 열린 트위터 개발자 회의(Chirp)[10]에서 밝혔다.

네트워크로 표현될 트위터 사용자는 많은 사용자들 사이에서 영향력 있고 활동적인 사람이어야 한다. 영향력 있는 사용자는 팔로워가 많은 사용자라고 할 수 있다. 초기의 설계는 팔로워를 기준으로 랭크된 상위 100명과 그들의 팔로워, 팔로워의 팔로워 사이의 관계를 표현하려 하였다. 그러나 데이터 수집에서 트위터 API 제한 정책의 한계에 의해 불가능하였다. 우리는 트위터 카운터에서 팔로워의 수에 따른 랭킹 중 상위 100명 사이의 팔로워, 친구의 관계를 나타내고 이를 분석하고자 한다. 또한 본 논문에서는 한국 사용자만을 대상으로 한다. 상위 100명은 4월 14일 트위터 카운터에 랭크된 순위로 1위로 소셜가 이외수부터 100위에 가수 이효리까지 이다 [11].

[그림 3]은 트위터 카운터에서 한국 사용자 중 팔로워의 수에 의한 랭킹을 보여준다.

The 1000 most popular Twitter users for Seoul				
Ranked by followers count and refreshed daily.				
		Followers	Following	Tweets
1	oisoo Joined 10 months ago Hwacheon 38,166458,127,516781 Seoul	114,990	398	2,129
2	soranet Joined 10 months ago soranet Seoul	92,850	102	301
3	leeminho Joined 1 year ago Korea Seoul	73,093	0	8
4	kimjuha Joined 9 months ago KOREA Seoul	60,913	6	4,959
5	biobyblo Joined 10 months ago the mind Seoul	50,297	37	1,199
6	kimsangbum Joined 12 months ago Seoul	44,208	1	10
7	chanjin	43,705	338	24,592

[그림 3] 트위터 카운터에서 제공하는 한국 사용자 중 팔로워에 대한 랭킹

관계는 트위터 API에서 제공하는 메소드의 호출에 의해 확인 가능하다. [표 1]은 "oisoo"라는 사용자가 "50a"라는 사용자에게 팔로워를 하고 있는지를 확인할

수 있는 요청의 예이다.

URL은 user_a에 팔로우를 하는 대상의 ID나 이름을, user_b에 팔로우를 받는 대상의 ID나 이름을 입력 값으로 받는다.

[표 1] 관계를 확인할 수 있는 URL의 예

```
http://api.twitter.com/1/friendships/exists.XML?user_a=oisoo&user_b=50a
```

결과 값으로 user_a가 user_b의 팔로워이면 true를, 팔로워가 아니면 false를 되돌려준다. 이러한 URL의 요청을 통해 상위 100명에 대한 팔로워 관계를 XML(eXtensible Markup Language) 형식으로 저장하였다. [표 2]는 XML로 저장된 팔로워 사이에 나타나는 관계 중 일부를 보여준다.

[표 2] XML로 저장된 사용자 간의 관계

```
<?xml version="1.0" encoding="UTF-8" ?>
<Relation>
  <follow user1="oisoo" user2="soranet" type="false" />
  <follow user1="oisoo" user2="leeminho" type="false" />
  <follow user1="oisoo" user2="kimjuha" type="true" />
  <follow user1="chanjin" user2="gyuri88" type="true" />
  <follow user1="chanjin" user2="hur" type="true" />
  .
  .
</Relation>
```

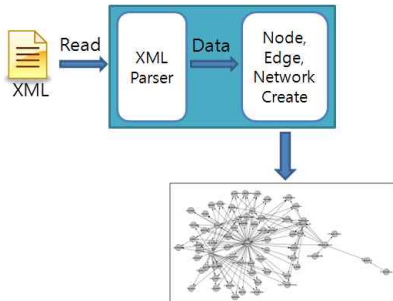
상위 100명 중 100위인 'HyoriiLee' 사용자는 자신의 트윗들을 외부 비공개로 설정하였다. 'HyoriiLee' 사용자에게 접근할 수 없으므로 본 논문에서는 Top 100 중 1인을 제외한 99명의 사용자를 대상으로 구현한다.

3.3 네트워크 생성

네트워크는 저장된 XML의 정보를 사용하여 사용자들 사이의 관계를 시각적으로 나타낸다. [그림 4]는 네트워크에 노드와 에지를 그리는 네트워크 생성에 대한 시스템 구성도이다.

TwitNet에서 XML의 follow 속성 user1과 user2는 네트워크에서 노드로 구성되고 속성 type은 user1과 user2를 연결하는 에지로 구성된다. 속성 type에서

'true'는 user1과 user2를 연결하는 에지가 존재하고 'false'는 에지가 존재하지 않는다. user1과 user2를 연결하는 에지가 존재하는 것은 user2는 user1을 팔로워로 가진다는 의미이다.



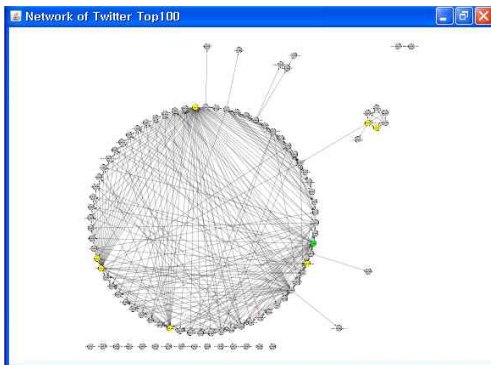
[그림 4] 네트워크 생성에 대한 시스템 구성도

[그림 5]는 네트워크를 생성하기 위해 선택하는 메뉴를 보여준다. "Create Network" 메뉴는 사용자들 사이의 관계를 나타내는 XML을 읽고, 이를 분석하여 네트워크를 만든다.



[그림 5] 네트워크 생성을 위한 인터페이스

[그림 6]는 TwitNet의 "Create Network" 메뉴로 생성된 네트워크이다.



[그림 6] 생성된 네트워크

3.4 네트워크 분석

네트워크는 상위 100명 사이의 관계를 시각적으로 보여준다. 그러나 이러한 네트워크는 노드와 에지의 수가 많아 네트워크상에서 관계를 많이 가지고 있는 사용자를 쉽게 찾기 힘들다. TwitNet은 네트워크상에 존재하는 사용자들이 가지고 있는 관계에 대한 분석 기능을 제공한다. 네트워크 분석은 두 부분의 컴포넌트로 구성된다. 각각의 컴포넌트가 하는 기능은 다음과 같다.

● Degree of Node

네트워크에서 에지는 노드로 들어오는 in-direction과 노드에서 나가는 out-direction이 존재한다. in-direction은 다른 노드로부터 도착하는 방향성을 가지는 경로이고, out-direction은 다른 노드로 나가는 방향성을 가지는 경로이다. TwitNet에서 in-direction은 팔로워를 받는 방향성을, out-direction은 팔로워를 하는 방향성을 가진다.

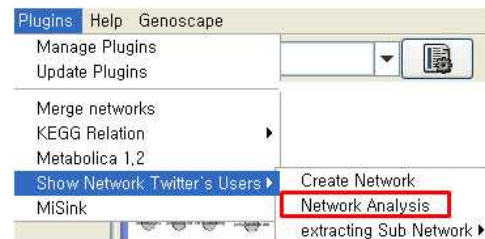
Degree of Node는 방향성을 분석하여 팔로워를 맺은 사람의 수, 팔로워를 신청한 사람의 수를 수치로 나타낸다.

● Loop between two Nodes

두 노드 사이에 발생하는 루프는 두 노드가 서로 밀접한 관계를 맺고 있음을 알 수 있다. TwitNet에서 루프는 서로 팔로워가 되는 존재이다.

TwitNet은 상위 100명 사이에 발생하는 루프를 분석하여 그들 사이에 얼마나 밀접한 관련이 있는가를 수치로써 나타낸다.

[그림 7]은 네트워크를 분석하기 위해 선택하는 메뉴를 보여준다.



[그림 7]네트워크 분석을 위한 인터페이스

분석된 수치는 노드의 속성 데이터로 추가되고 Cytoscape 하단의 Data Panel에서 확인 가능하다. [그림 8]은 분석된 데이터를 나타내는 Data Panel의 인터페이스와 선택된 사용자에 대한 분석의 예를 보여준다.

TwitNet은 노드에 대해 분석하므로 Node Attribute

Browser에서 노드에 추가된 속성 값을 확인할 수 있다. Degree:In과 Degree:Out은 에지의 in-direction과 out-direction을 수치로서 나타낸다. 길이(length)가 2로 제한된 Loop는 상대방과 서로 팔로워 관계에 있는 정도를 수치로서 나타낸다.

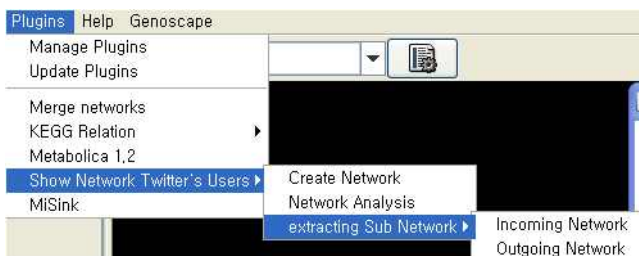
ID	# Loop	Degree:In	Degree:Out	Total Degree	Type
dogsul	0	3	54	57	user
QuoVadisKorea	0	8	0	8	user
sisain_editor	0	2	32	34	user
jewelguy	0	2	5	7	user
dosanim	0	3	23	26	user
twipl	0	2	29	31	user
soccersu	0	3	0	3	user
pumpl	0	4	0	4	user

[그림 8] Data Panel 인터페이스

Network Attribute Browser는 Loop를 계산한 길이와 전체 네트워크에서 발생하는 Loop의 수를 확인할 수 있다. 각 속성 값에 따라 정렬 기능도 제공한다.

3.5 네트워크 추출

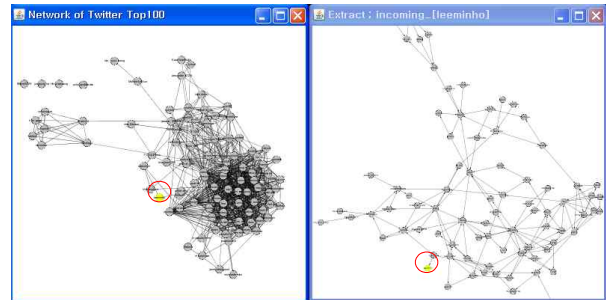
많은 데이터를 가지는 네트워크는 노드와 에지의 관계가 서로 중첩되어 있어 상관관계를 알아내기가 쉽지 않다. TwiNet은 선택된 노드에 대해 in-direction이나 out-direction만을 가지는 네트워크를 추출하는 기능을 제공한다. 노드와 노드 사이에 방향성을 발견하기 위해 그래프 탐색 알고리즘인 깊이 우선 탐색 알고리즘[12]을 사용하였다. [그림 9]는 네트워크 추출을 위해 선택하는 메뉴를 보여준다.



[그림 9] 네트워크 추출을 위한 인터페이스

"Incoming Network" 메뉴는 선택된 노드에 대한 in-direction을 발견하여 새로운 네트워크를 생성하고 "Outgoing Network" 메뉴는 out-direction에 대한 새로운 네트워크를 생성한다. [그림 10]은 네트워크 추출의 메뉴를 이용하여 전체 네트워크에서 선택된 노드에

관련된 Incoming Network의 추출이다.



[그림 12] 전체 네트워크(좌)와 선택된 노드 "leeminho"(이민호)에 대해 추출된 네트워크(우)

4. TwiNet을 이용한 분석

TwiNet은 XML로 저장된 트위터 사용자들의 관계를 일고 파싱하여 시각적인 네트워크로 나타내고 네트워크를 분석하여 관계를 수치적으로 보여주는 기능을 가진 Cytoscape의 플러그인으로 개발되었다. 상위 100명 중 트윗이 보호되어 있는 100위의 "HyoriiLee"(이효리) 사용자를 제외하여 노드는 99개를, 관계를 나타내는 에지는 900개가 생성되는 것을 볼 수 있다. 상위 100명 중 1위인 "oisoo"(이외수) 사용자는 상위 100명들 사이에서 22명의 팔로워를 가지고 11명에게 팔로워를 하였다. 상위 100명에서 그들과 가장 많은 관계를 가지고 있는 사용자는 93위의 "sitehis"(박선민)이며 30명의 팔로워를 가져 그들 사이에서 가장 많은 팔로워를 가진 사용자이다. 그러나 관계의 수치가 많다고 하여 팔로워나 팔로워의 수치가 모두 높지는 않다. [표 4]는 분석된 수치의 상위 10명을 보여준다. "dogsul"(독설닷컴) 사용자는 전체적인 관계의 수치가 높지만 자신을 팔로워 하는 사용자는 4명이다. 그 외에도 20위의 "kim_hyun_joong"(김현중) 사용자는 팔로워가 0, 친구가 1로 상위 100명과 관계는 단 한개이다. 또한 각각 81위, 64위, 38위인 "BBme2DAY"(Big Bang's me2DAY), "jungsungha"(정승하), "HeroJaejoong"(영웅재중) 사용자는 상위 100명들 사이에 팔로워, 친구의 관계를 하나도 가지지 않는 것으로 나타났다. 그러므로 데이터의 대상이 된 상위 100명은 많은 팔로워를 가지지만 그들 사이에서 항상 깊은 연결이 있다고 할 수 없다.

[표 4] 분석된 수치의 상위 10명

사용자 이름	Degree :In	Degree :Out	Total Degree	순위
sitehis	30	35	65	1
dogsul	4	54	58	2
hcroh	6	50	56	3
sangjungsim	6	40	46	4
hiconcep	26	18	44	5
jamiepark	18	24	42	6
hongss	16	23	39	7
twit_korea	18	20	38	8
mirmong0	17	21	38	8
iSonics	26	11	37	10
coreacyd	9	28	37	10

4. 결론 및 향후 연구 과제

본 논문에서는 트위터 사용자들의 관계를 시각적으로 나타내는 Cytoscape 플러그인 TwitNet을 설계 및 구현하였다. 개발된 TwitNet은 한국 사용자 중 팔로워로 랭킹된 상위 100명을 대상으로 그들 사이의 관계를 네트워크로 나타내고 분석하였다. 네트워크는 상위 사용자들 사이의 관계를 시각적으로 나타내었으며 네트워크 분석을 통해 팔로워를 많이 가지는 상위 사용자들 사이에 생성되는 관계를 수치적으로 나타내었다. 또한 복잡한 전체 네트워크에서 필요한 관계만을 추출하여 볼 수 있는 네트워크 추출 기능을 이용해 해당 사용자에게 관계되어 있는 사용자만을 쉽게 확인할 수 있다.

분석된 결과는 최근 “Measuring User Influence in Twitter: The Million Follower Fallacy”[12]라는 제목으로 트위터 한 유저의 영향력이 과연 쓸만한가를 연구한 보고서와 같이 팔로워의 수가 많다고 해서 영향력이 반드시 크지 않다는 의미를 다시한번 입증하였다. 네트워크 생성과 분석을 통해 온라인상에서 영향력 있는 사용자들이 어떻게 관계되고 있는가에 대해 효과적으로 보여줄 수 있을 것으로 기대된다. 하지만 대상으로 적용된 상위 100명은 실시간으로 변할 수 있는 가능성을 가지고 있다. 또한, 본 논문에서는 상위 100명에 대해서만 관계를 보이고 있다.

향후 과제로는 영향력 있는 사용자에 대한 대상을 실시간으로 수집하여 시각적으로 보여줄 수 있는 기능을 추가하고 대상을 좀 더 포괄적으로 수용할 수 있는 방안

에 대해 구현할 계획이다. 또한 네트워크 분석에 있어 Loop의 길이가 유동적일 때 분석하는 방법과 다른 사용자들 거쳐 관계를 맺고 있는 관계에 대해 분석할 수 있는 기능을 추가할 계획이다.

5. 참고문헌

고상민, 황보환, 지용구, “소셜네트워크서비스와 온라인 사회적 자본”, 한국전자거래학회지, 제 15권, 1호, pp.103~118, 2010년 2월

[1] <http://twitter.com>, 트위터

[2] 설진아, “소셜 미디어의 진화양상과 사회적 영향”, 한국언론정보학회, 2009

[3] <http://twitpic.com>, 트윗픽

[4] <http://twitvid.com>, 트윗비드

[5] <http://twittercounter.com>, 트위터 카운터

[6] <http://twitter.grader.com>, 트위터 그레이더

[7] <http://asterisq.com/blog/2009/10/14/explore-your-twitter-network-with-mentionmap>, Mentionmap

[8] <http://cytoscape.org>, Cytoscape

[9] <http://chirp.twitter.com>, Chirp

[10] http://twittercounter.com/pages/country%26time_zone=Seoul, 트위터 카운터-서울

[11] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, Krishna P. Gumma야, “Measuring User Influence in Twitter: The Million Follower Fallacy”, Proc. International AAAI Conference on Weblogs and Social Media (ICWSM), 2010년 3월

[12]