Followee recommendation based on text analysis on micro-blogging activity[1]

Nazanin jafari

Noushin salek faramarzi

1: Armentano, Marcelo Gabriel, Daniela Godoy, and Analía A. Amandi. "Followee recommendation based on text analysis of micro-blogging activity." *Information systems* 38.8 (2013): 1116-1127.

Introduction

- In Twitter, users are interested in finding not only their close friends but also new relevant contacts not yet known to them. Therefore it is necessary to recommend relevant users to them to follow.
- Our recommendation phase consists of two steps:
  - Topology based approach
  - Content based approach

Topology based approach

- The method employed to explore the Twitter network with the goal of gathering candidate users to recommend a target user u_t is based on the following hypothesis:

- If user u_f follows a user that is also followed by u_t, then other people followed by u_f can be interest to u_t.

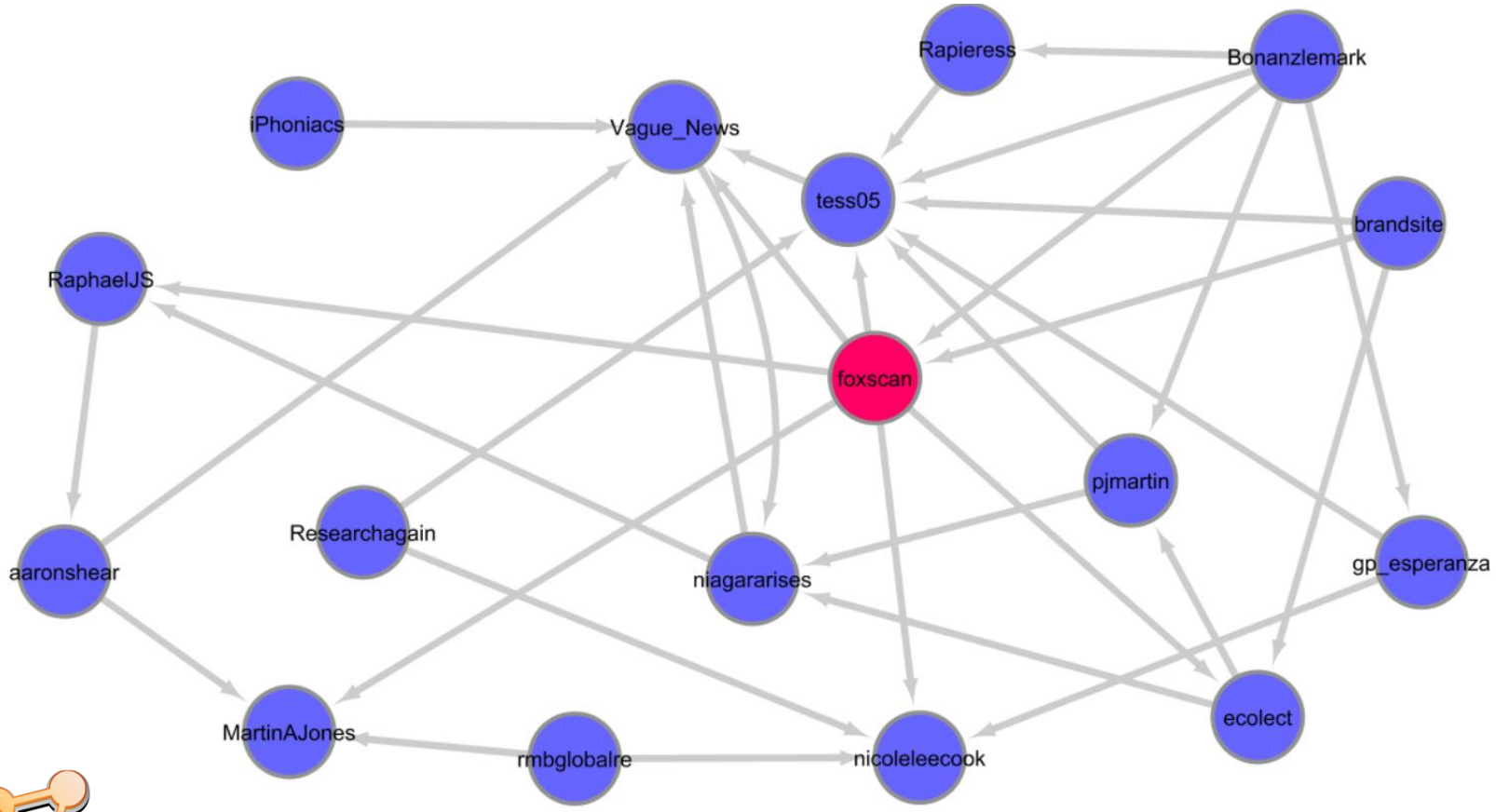
Cont'd

- To be more precise we will follow these four steps:
  1. Starting with the target user u_t, obtain the list of users he/she follows, let us call this list
  2. For each element in S get its followers, let us call the union of all these lists L,
  3. For each element in L obtain its followees, let us call the union of all these lists P
  4. Exclude from P those users that the target user is already following. Let us call the resulting list of candidates R

Input data

- Relationship of users in twitter and tweets they share are our data set.
- This dataset consists of 54,000 follower/followee relation between more than 200,000 users and their corresponding tweets which reach more than 1 million tweets at specific time span.
- We directly received data set database from the author of the paper and it has two tables :
  - Table "socialgraph": first column follows second column (directed links between usernames)
  - Table "tweets": first columm is the username, second column date posted, third column is the tweet content.

Implementation of topology based approach



Cytoscape

Results

Followee list of Target user, S = {tess05, nicoleleecook, MartinAJones, Vague_News, RaphaelJS, ecolect}

Followers of S, L = {foxscan, Researchagain, brandsite, Rapieress, gp_esperanza, Bonanzlemark, pjmartin, rmbglobalre, aaronshear, iPhoniacs, niagararises, tess05}

Followees of L, P = {tess05, nicoleleecook, MartinAJones, Vague_News, RaphaelJS, ecolect, foxscan, gp_esperanza, Rapieress, pjmartin, niagararises, nicoleleecook}

Candidates = {gp_esperanza, Rapieress, pjmartin, niagararises}

Experimental results

- Some target user followees are become hidden and then we tested if they were suggested as future followee.
- Set of followee of each user divided into 70% training and 30% testing.
- If followees in 30% appeared in candidate R set, then hypothesis satisfied.

Pre-Processing data for content based approach

- Our stop words list consists of 700 Default English and MySQL Stop words

Stopword list

a	been	get
about	before	getting
after	being	go
again	between	goes
age	but	going
all	by	gone
almost	came	got
also	can	gotte
am	cannot	had
an	come	has
and	could	ha

- Slang words have been taken from noslang.com/dictionary and it contains 5350 entries.

Future works

- In content based approach we will try to rank each of the candidate users that was obtained from topology based approach.
- We will implement four different profiling strategies in content based approach
- Different metrics will be considered to evaluate our final recommender system