**Project Proposal**

**-- Mining the Microblog User’s Social Circles**

**Zhang, Zhi 1155039997 cheungzeecn@gmail.com**

**He, Qingfan 1155044621 kenjdmlfo@163.com**

**Background**

Since the Web 2.0 era has come, techniques such as Wiki, SNS, etc, are widely used and gradually become an inseparable part of people's life, including microblog. Microblog is a platform that users can follow from any famous individuals to their schoolfellow or colleagues, as well as read their tweets without any permission. Microblog users tend to follow some notables while these celebrities usually might not fellow back, which will make these celebrities differ from the real friends of the follower. According to this reality, our project aim to find out the real social circles of the microblog users, dig out the hidden information of each circle, and try to do some business operations base on the social circles we mined, as up to the end of December 2012, Sina Weibo has more than 500 million users, which contains infinite business opportunities.

**Data Source**

We obtain data from Sina Weibo, there are a lot of APIs and crawlers already. Our data consists of two parts, one is user’s personal information, another is user’s bilateral follower list.

(1) User’s personal information

UserName, Id, Gender, birthday, general information, etc.

Contact information

Education background

User’s tag about himself/herself

V titles if the user is a V user

(2) User’s bilateral follower

Follower list and the number of follower

Followee list and the number of followee

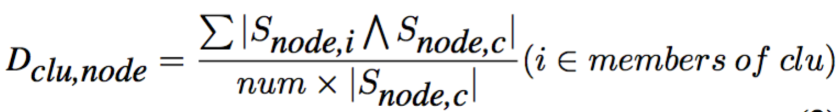
**Analysis Methods**

(1) Circle discovering

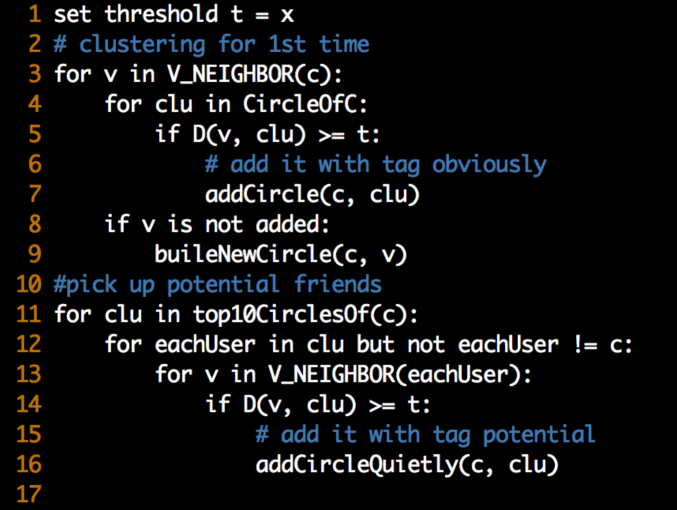
1. For a certain user, we collect all his bilateral friends as neighbours
2. For a certain user c, we iterate all his neighbours v to cluster them into circles
3. For every circles of c, we find out the users should be in the circle but are not the neighbor of c yet.

(2) Method of clustering

1. Define V\_NEIGHBOR(a) = all the bilateral followees of a
2. Define Sa,b = V\_NEIGHBOR(a) ∩ V\_NEIGHBOR(b)
3. Define the distance between a node named *node* and a cluster(circle) named *clu*, *c* is the center user we want to analyse, *num* is the number of existing member of the cluster *clu*.



1. Clustering:



(3)Circle analyzing

Naïve analyzing using the basic information for every person in the circle to label a circle, it maybe, polling method is used here.

**Applications**

(1) Analysis someone’s social circles, to know more about this person in Recruitment.

(2) Effective promotion. We can accurately do the promotion in the circles with certain label if we can tag the circles automatically

(3) Recommend a potential friend exists in a certain circle but the user having follow him/her.