Course Report of Wireless Communication and Mobile Network

Wenyu Zhang
5130309496

Jun 26, 2016

Contents

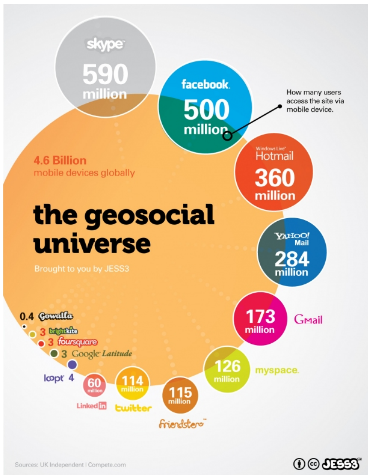
1 Introduction

This is the course report of the class wireless communication and mobile network. I focus on the project with a topic 'Optimization of Information Diffusion in Mobile Social Network Using Community Structure'. This report specifically formulates the work I have done in this course. I will divide my report into several sections such as Background and Motivations, Model and Problem Formulation, One Reasonable Method, Summary and Future Work.

2 Background and Motivations

2.1 Background

Nowadays, social networks have been evolving to online social networks(OSNs) such as Facebook, Twitter. Recent years have witnessed a dramatic growth of user population of online social networks. For example, according to the report in March 2013, Facebook has 1.11 billion people using the site each month, which represents a 23 percent growth from a year earlier.
What's more, with the proliferation of smart mobile devices, geosocial networks become a major kind of OSN. As a result, the mobility and geolocation should be taken into consideration when analyzing geosocial networks.
From fig.1 we can see more and more people using online social network to communicate with each other. Fig.2 shows the three layers of a geosocial network.



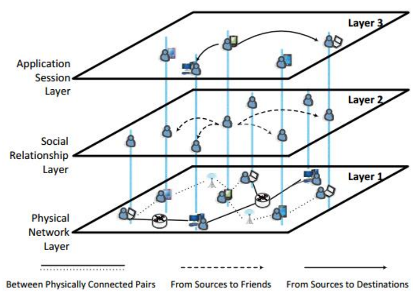


Fig.1 User Scale of Some Geosocial Networks & Fig.2 Layered System Model of OSNs

2.2 Motivation: A Marketing Motivation

Suppose that we have data on a social network, with estimates for the extent to which individuals influence one another, and we would like to market a new product that we hope will be adopted by a large fraction of the network. The premise of viral marketing is that by initially targeting a few influential members of the network and giving them free samples of the product, we can trigger a cascade of influence by which friends will recommend the product to other friends, and many individuals will ultimately try it. But how should we choose the few key individuals to use for seeding this process?
This is a meaningful question business cares because if more people receive the message, then more profits can be achieved.
Fig.3 shows a strategy of choosing influential members, which are colored with red. White nodes are those can be influenced by red nodes.

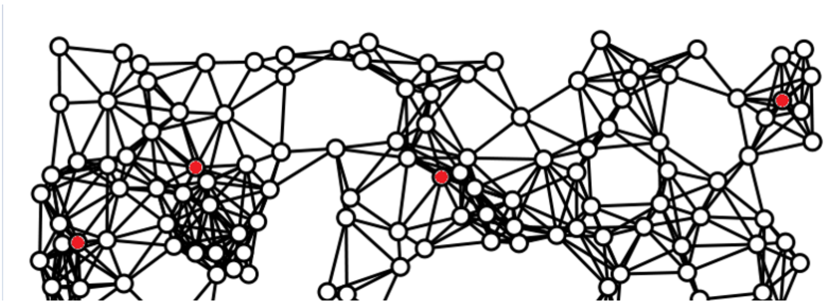


Fig.3 A Marketing Diffusion Model

3 Model and Problem Formulation

3.1 Mobile Social Network Model

First, we introduce a mobile social network model. G_t is a dynamic geo-social graph at time step t.

$$G_t = (V_t, E_t)$$

w_{uv} is the weight of edge between node u and v, then we have

$$w_{uv} = w_{vu}$$

The value of the edge weight can be calculated as

$$w_{uv} = a \cdot O_{uv} + b \cdot \beta^l$$

Note that this is a new definition of weight in geo-social network which takes both social tie strength and impedance in geo-transmission into consideration.
β is the damping parameter, l is the geographic distance between node u, v, O_{uv} is the strength of social tie, which represents the frequency of contacts between two users.As the result of big data analysis, the stronger the tie between two users, the more their friends overlap, a correlation that is valid for about 95% of the links.
O_{uv} can be formulated as

$$O_{uv} = \frac{n_{uv}}{(k_u - 1) + (k_v - 1) - n_{uv}}$$

where n_{uv} is the number of common neighbors of u and v, and $k_u(k_v)$ denotes the degree of node $u(v)$.

3.2 Independent Cascade Diffusion Model

Here comes our diffusion model.
G is the graph of the network.

$$G = (V, E)$$

p_{uv} is the probability of diffusing success of node u, v.

$$p_{uv} = normalized(w_{uv})$$

S_{ut} is the state of node u at time t, which can be either 0 or 1. 0 represents inactive and 1 represents active.

$$S_{ut} \in 0, 1, \forall u \in V \ at \ time \ step \ t$$

$$A \subseteq V, S_{u0} = 1, \forall u \in A \ and \ S_{v0} = 1, \forall v \in V \setminus A$$

Then we make an iteration: When node u first becomes active in step t, it is given a single chance to activate each currently inactive neighbor v; it succeeds with a probability p_{uv}.
Finally, the target function is $\sigma(A)$, which represents the influence extent

$$\sigma(A) = number \ of \ active \ nodes \ ultimately$$

Fig.4 is part of the diffusion process. The arrow represents the direction of message transmission. And one node may have more than one chance to be activated.

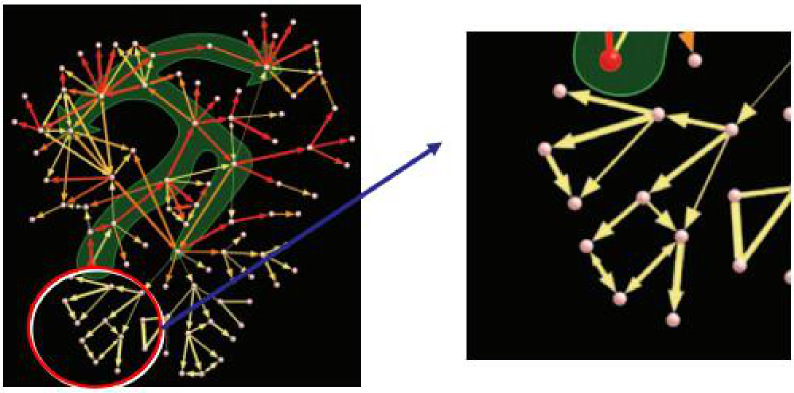


Fig.4 A Cascade Diffusion Model

3.3 Main Problem: Influence Maximization

Based on the model formulated before, our main problem is to target A in the whole network, and then maximize $\sigma(A)$, which can be formulated as
Given $G_t = (V_t, E_t)$ and $number\ of\ initially\ active\ nodes|A|$, targeting A, to maximize $\sigma(A)$.
This is a NP hard problem, which is very complicated.
There have been some approximation algorithms like greedy algorithms trying to significantly reduce the complexity of targeting on the premise of influence assurance.
Different from those algorithms. I come up with the idea that we can use community detection to optimize the influence maximization problem, which will be explained in the next section specifically.

4 One Reasonable Method

4.1 Brief Introduction of Community Detection

Generally speaking, Community detection is dividing the nodes into some communities, nodes in the same community have closer relationships. Below is the mathematical representation of community detection.

$$G = (E, V)$$

$$C = C_1, C_2, C_3, ..., C_k$$

$$C_i \in C\ and\ C_i \subseteq V, C_1 \cup C_2 \cup C_3 \cup ... \cup C_k = V$$

$$Q(C) = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta_{ij}$$

$$Q(C) \leq 1 \; and \; can \; be \; either \; positive \; or \; negative$$

Fig.5 is an example of community detection.

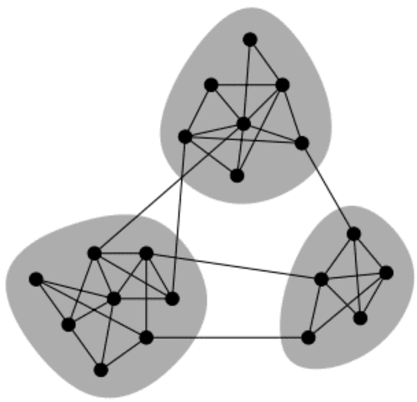


Fig.5 An Example of Community Detection

$Q(C)$ is called modularity, which is the evaluation index of the quality of community detection. The higher $Q(C)$ is, the better the detection is. Fig.6 shows the meaning of $Q(C)$

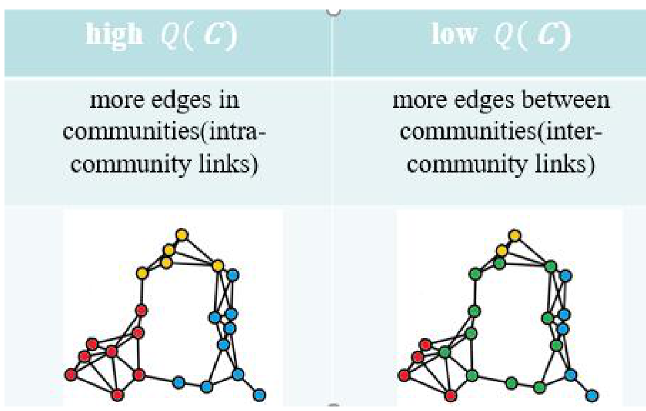


Fig.6 Meaning of Modularity

We should notice that in the mobile geo-social network, it's necessary to add the dynamic and geographic characteristic into the problem formulation. Fig.7 shows the way we detect communities in dynamic networks. We can update the community structure by the difference between two time steps instead of recalculate the whole network again, which is much faster as the difference of two adjacent time step is small.

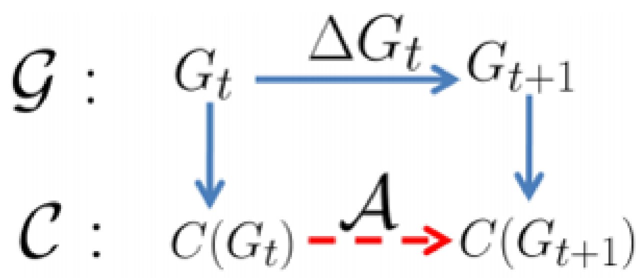


Fig.7 Strategy of Detecting Dynamic Networks

We also have to consider the weighted graph condition because in geographic situation, the bond strength of different edges is different. So we modify the representation of modularity as below

$$Q(C) = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) w_{ij} \delta_{ij}$$

4.2 Supportive Reasons

We now explain why we can use the community detection to solve diffusion problem. Below are some intuitive fact and criteria.

1. We can distribute seeds in different communities to achieve higher $\sigma(A)$.

Fig.8 is two seed choosing choices. In the left graph, all three seeds are in one community, as we already know, the ties between communities is weak ties because of low weight, so there is a high probability that message cant transfer to the other two communities, which means all the nodes in the left two communities is impossible to be activated. On the contrary, the targeting of the right graph is much more stable.

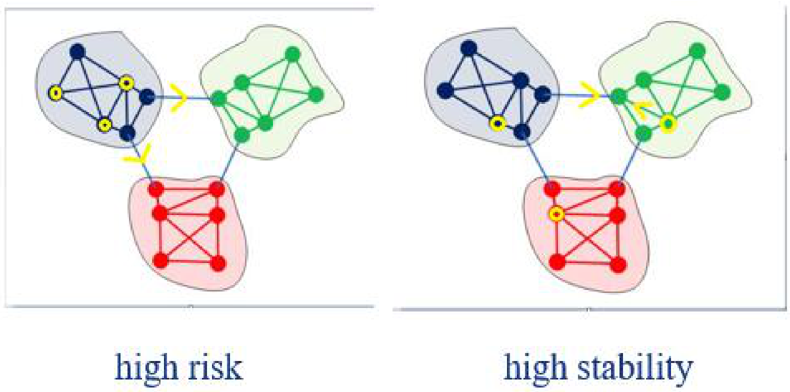


Fig.8 Why Distributing Seeds into Different Communities is Reasonable

2. More intra-community links means more chances to get activated.
Fig.9 shows that in communities, the relationship of nodes is much closer because of more edges and high edge weight.

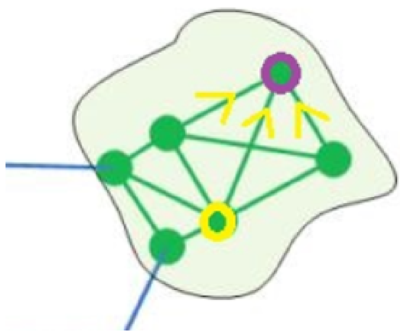


Fig.9 More Chances for a Node to Be Activated in Community Structure

3. Smaller targeting scale by community structure means less calculation complexity.
Let N represents the number of nodes in the whole network. Take fig.10 as an example, If we calculate the original complexity as $O(N^k)$, then the after detection complexity is $O(3 \cdot O((\frac{N}{3})^k))$, which means the complexity is reduced by $3^{1-k} < 1(k > 1)$.

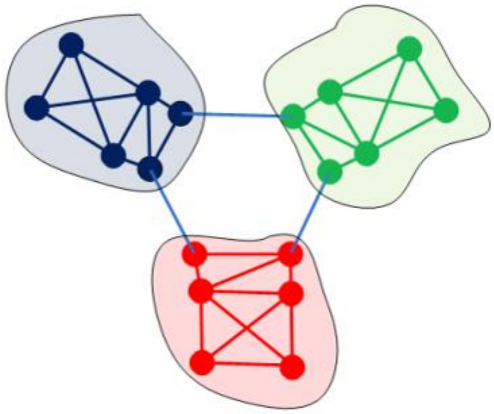


Fig.10 Reduce of complexity by community detection

4.3 Evaluation Index of the Method

Generally, there are two main evaluation indexes to evaluate the diffusion maximization problem in our project.

The complexity of targeting A

$complexity$ = O(fast community detection) + O(quick targeting in scale of communities)

The influence extent

$$\sigma(A)$$

5 Summary

In this project, we come up with the model of diffusion maximization problem and an original method of community detection to solve the problem in mobile geo-social network. Below are two main parts of my work that distinguish myself from other researchers.

A diffusion model in mobile network which considers both social tie strength and geo-transmission impedance.

Using community detection as a reasonable method to solve the NP-hard influence maximization problem.

6 Future Work

1. Algorithms
2. Experiments
3. Other circumstances (e.g. overlap of communities)
4. Other diffusion problem (e.g. influence maximization with time limit/time minimization)

7 Reference

1. Nam P. Nguyen, Thang N. Dinh, Ying Xuan, My T. Thai ”Adaptive Algorithms for Detecting Community Structure in Dynamic Social Networks” in INFOCOM
2. J.-P. Onnela, J. Sarama ki, J. Hyvo nen, G. Szabo, D. Lazer, K. Kaski, J. Kerte ’sz, and A.-L. Baraba ’si ”Structure and tie strengths in mobile communication networks” in PNAS
3. David Kempe, Jon Kleinberg, Ev ’ a Tardos ”Maximizing the Spread of Influence through a Social Network” in ACM
4. M. E. J. Newman ”Modularity and community structure in network” in PHYSICS
5. Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, Albert-Lszl Barabsi ”Human Mobility, Social Ties, and Link Prediction” in ACM