Department of Physics
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

Project Proposal:

What is the weakest point in the Social Networks with Community Structure and how to find it?

WANG Guan
wangguan@ust.hk

March 18th, 2011

Relating Topics: Network Flow, Community Detection, Social Networks Mining, Social Networks Properties.

Background

This part will mainly be the content of my presentation on April.12th as a background introduction.
To illustrate the importance of detecting the weakest point in the social networks, I start by presenting the following paper which talks about preventing the spreading of diseases in social networks with community detection. It introduces the evaluation of such weakest points, especially the concept of "Community Bridge", and some algorithm to detect it.
[1] M. Salathé and J. H. Jones, **Dynamics and Control of Diseases in Networks with Community Structure,** PLoS Comput.Biol 6 (2010).
http://www.ploscompbiol.org/article/info:doi%2F10.1371%2Fjournal.pcbi.1000736

Then, motivated by this paper presented in the course, I will try to propose a way for detecting overlapping communities by clustering on the linksm and thus look for the weakest point in the social networks as the nodes or links in the overlapping areas.
[2] Xufei Wang, Lei Tang, Huiji Gao, and Huan Liu. **Discovering Overlapping Groups in Social Media**. In Proceedings of The 10th IEEE International Conference on Data Mining (ICDM'10), 2010
http://www.public.asu.edu/~ltang9/papers/overlapping_groups_ICDM10.pdf

Finally I'll briefly introduce the following algorithm as a possible good way of illustrating some structures of communities in one network, which may help on showing the overlapping areas of the communities.
[3] Zhenggang Wang and K. Y. Szeto, **Structure profile of complex networks by a**

model of precipitation, Physica A: Statistical Mechanics and its Applications, Volume 389, Issue 11, 1 June 2010, Pages 2318-2324
http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TVG-4YC80S3-1&_user=2149949&_coverDate=06%2F01%2F2010&_rdoc=1&_fmt=high&_orig=gateway&_origin=gateway&_sort=d&_docanchor=&view=c&_searchStrId=1669788845&_rerunOrigin=google&_acct=C000056388&_version=1&_urlVersion=0&_userid=2149949&md5=5c51db570e3ae04a653056f84467228e&searchtype=a

Project Proposal

Motivated by the [1], the weakest point in social networks can be defined as:
The weakest nodes in social networks will be the ones that belong to the largest number of multiple communities, known as "Community Bridges".
So now the problem is how to find such heavily overlapping communities from the original data.
[2] actually gives a very good method of detecting communities in heavily overlapped social networks: **Community Detection by Clustering on the Links**. This is quite an interesting idea because the links describing the relationships between people will be more specific for different communities. For example the links between me and my undergraduate friends will be mostly in the community of Peking University while the links between me and my postgraduate friends will be mostly in the community of HKUST. Of course there will still be some overlapping since some of my friends from PKU also came to HKUST with me, but such overlapping will be much less than that of the community classification of nodes.
We can take advantage of the existing community detection algorithms for community detection on the links. The job we need to do is only to change the social networks to its **dual graph**. That is, **change the nodes into links and change the links into nodes**. This way we may use existing algorithms like [3] or others to find communities of links. This is the core part of this research project.
Then we change the dual graph back to its original social networks and check out the node that belongs to the largest number of communities will be the most influential, dangerous and thus the weakest point in the network.
The data we use should be networks with heavily overlapped communities. One way is to generate some artificial data ourselves (like in [1]) and check our method. Another and a more meaningful way is use real social networks data. This may be ambitious and I cannot guarantee good results. The dataset I find it proper now is the *email-Enron networks* and 5 *Collaboration networks* from the Stanford website http://snap.stanford.edu/data/index.html

About Me

Homepage: http://ihome.ust.hk/~wangguan
Facebook: http://www.facebook.com/crownguanwang
Twitter: @crownpku