Community detection and missing-link prediction on bi-partite networks by non-backtracking random walk matrix

With Cris Moore and other collaborators, we have studied community detection on networks using non-backtracking random walk matrix [PNAS 110, 52(2013)], which works all the way down to the detectability transition on networks generated by stochastic block model and also gives good performance on real-world networks.

Lots of real-world networks are bi-partite, which means there are two types of nodes and edges in the network always come from one type of node to the other type of node. For example, user-movie networks have edges connecting movies and users.
It would be interesting to extend our study on non-backtracking random walk matrix from uni-partite graphs to bi-partite graphs. The advantages of using non-backtracking matrix over other standard matrices e.g. random walk matrix and Laplacians are: 1.) non-backtracking matrix is less influenced by high-degree nodes and local structures in the networks. 2.) non-backtracking matrix gives a way to select number of groups in networks. So we expect non-backtracking matrix can also tell us how many groups in each type of nodes in bi-partite networks.

I already have some experience of using non-backtracking matrix on bi-partite graph, not on community detection task, but on finding solutions of some mathematical problems. And it would be not so hard for a REU to adapt the code to the community detection task on some real-world networks.

From this project REUs can learn some theory and knowledge in the field of spectral clustering, community detection and networks, and practice scientific programming using MATLAB.

Pan Zhang
03/31/2014