Dear Editors

The motivation for this paper is to apply Bayesian structure learning using Model Averaging in large-scale networks. Currently, Bayesian model averaging algorithm is applicable to networks with only tens of variables, restrained by its super-exponential complexity. We present a novel framework, called Lama (Large-Scale Model Averaging), making it possible to handle networks with very large scale by divide-and-conquer.

In comparison with other four state-of-art large-scale network structure learning algorithms such as ARACNE, PC, Greedy Search and MMHC, Lama shows comparable results in five common benchmark datasets, evaluated by precision, recall and f-score. What’s more, Lama makes it possible to learn large-scale Bayesian structure by Model Averaging which used to be intractable. Also, the hierarchical information of overlapping communities serves as the byproduct, which could be used to mine meaningful clusters in biological networks, such as protein-protein-interaction network or gene regulatory network, as well as in social network.

There are two papers most closely related to the contributions of my paper. First paper is about how to partition the graph into hierarchically-organized overlapping communities, by Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, named “Link communities reveal multiscale complexity in networks”. The second paper is about how to reconstruct Bayesian structures by model averaging and MCMC, by N. Friedman and D. Koller, named “Being bayesian about network structure: A bayesian approach to structure discovery in bayesian networks”. The former paper provides me the tools to decompose the large-scale network, and the latter paper provides me the essence of intra-community structure learning algorithm.

I haven’t submitted the paper before. Also, I haven’t published parts of this paper before.

If the paper is accepted, it will be presented at the ECMLPKDD 2013 conference by one of the authors.

Respectfully yours,

Yang Lu