**Project1-** *Modeling network formation via correlated community structure*

**Abstract**

It is widely known that individuals in a network form declustered communities, where each individual can belong to several groups. Although another important aspect of these separate/shared communities should be encoded via their correlation, as some may induce friendships while others inhibit such connections. In this paper we account for such co-occurrence via introducing correlations among the community memberships of individuals via logistic normal prior. We argue that accounting for such correlations both complies with the assortative mixing theory and performs better compared to the widely known mixed-membership stochastic block model(Airoldi 2008). We furthermore use stochastic variational methods to offer both fast and efficient inference.

Outline for the first paper (Keeping in mind marketing implications)

Literature on

* Theory
  + How connections form
    - Homophily((dis)assortative mixing)
    - Preferential attachment
    - From network perspective
  + Communities
    - Network communities
    - Overlapping versus clustered
    - Multilayered networks/roles versus communities
  + Directed versus undirected
  + Why matters: dim reduction, insights into opinion leaders
  + Marketing applications
* Methodology
  + Model based vs algorithmic
  + Latent variable models
  + MMSB, latent space model, AGM…
  + Correlated topic models and priors
  + Challenges in inference and estimation
    - general scalability issues
    - pros and cons MCMC
    - pros and cons tensor factorization
    - pros and cons variational inference/non-conjugacy
    - pros and cons SVI

Question

* Dim reduction of a large scale observed network and *insights(better not be totally exploratory)*

Model

* Correlational and directional LNMMSB, departure from (a)-MMSB
* Acknowledging flaws, and how to overcome them
* Choice of priors and implications
* Inference
  + VI setting
  + SVI setting

Data

* Stats
* Simulation and visualization
  + Small scale
  + Large scale
* *(ABM)*
* Real data and visualization

Inference

* Inference derivations
* Estimation

Results

* Visualization
* Performance
  + Simulations (NMI, predictive likelihood, perplexity…)
    - *(comparison with other methods)*
  + Real data (predictive likelihood, perplexity…)
* Interpretation for real data
  + How to combine with current measure, centrality, clustering, betweenness…

Implications

* Insights
* Extensions and relation to next steps