Network Structure and Influence

A. Yazdiha[1] B. Donkers[2]

[1]Erasmus School of Economics
[2]Erasmus School of Economics

Department Lunch Meeting, 2018

Outline

1. Motivation
   - A Prologue to Opinion Leadership
   - Our View of the Problem
   - Findings from Email data
2. 1st Project
   - Community Detection
   - Some Preliminary Results
3. 2nd Project
   - Resolution
   - Current Status
4. 3rd Project
   - Dynamic Evolution of Behavior and Structure

Outline

Networks and Influentials

Frame subtitles are optional. Use upper- or lowercase letters.

- This is a very well known problem in networks and marketing in general
- Who is an influential?
- Abundant literature on identifying influence and detecting opinion leaders(Papers)
- Among these the role of sociometric data has been quite common

Networks and Influentials

Frame subtitles are optional. Use upper- or lowercase letters.

- This is a very well known problem in networks and marketing in general
- Who is an influential?
- Abundant literature on identifying influence and detecting opinion leaders(Papers)
- Among these the role of sociometric data has been quite common

Networks and Influentials

Frame subtitles are optional. Use upper- or lowercase letters.

- This is a very well known problem in networks and marketing in general
- Who is an influential?
- Abundant literature on identifying influence and detecting opinion leaders(Papers)
- Among these the role of sociometric data has been quite common

Networks and Influentials

Frame subtitles are optional. Use upper- or lowercase letters.

- This is a very well known problem in networks and marketing in general
- Who is an influential?
- Abundant literature on identifying influence and detecting opinion leaders(Papers)
- Among these the role of sociometric data has been quite common

Previous Research

- A lot of attention is given to testing influence w.r.t individual positions at network level
  - i.e. Network centrality:degree, betweenness, prestige, etc
- Take co-authorship network; Who is the most influential?
- Almost all of the previous works deals with the global view of the network
  - hubs(),degrees(), betweenness(), ...

Previous Research

- A lot of attention is given to testing influence w.r.t individual positions at network level
  - i.e. Network centrality:degree, betweenness, prestige, etc
- Take co-authorship network; Who is the most influential?
- Almost all of the previous works deals with the global view of the network
  - hubs(),degrees(), betweenness(), ...

Previous Research

- A lot of attention is given to testing influence w.r.t individual positions at network level
  - i.e. Network centrality:degree, betweenness, prestige, etc
- Take co-authorship network; Who is the most influential?
- Almost all of the previous works deals with the global view of the network
  - hubs(),degrees(), betweenness(), ...

Previous Research

- With this view some conclusion can contradict
- Influentials have high degree(paper)
- Degree is irrelevant, betweenness matters(paper)
- For a comparison see (paper)
- This is mainly due to disregard of the story behind network structure

Outline

About the Network Structure

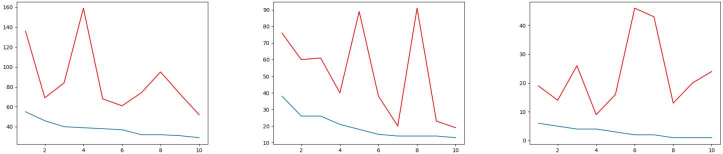
- Homophily: birds of a feather flock together
- Result: Networks end up with communities, where connections are:
  - Denser within vs. across
- As a matter of fact, networks are a multiplex of smaller sub-networks
- Communities as a signal for domains suggest a finer grained opinion leadership

Application of Community Detection in SNs

- Gain insights at the community level about individual positions rather than at the global network level, (Project 1)
- Allows for a good control when identifying influence by disentangling selection from causality, (Project 2)
- Enables dynamic monitoring of network and behavior evolution in a time-stamped data, (Project 3)

Outline

Knowing the Ground Truth Communities(SNAP - Leskovec & Krevl, 2014)



(a) (b) (c)

Figure: Degree centrality for the ground truth communities(blue line), and degree centrality for the same nodes at the network level(red line)

Outline

1. Motivation
   - A Prologue to Opinion Leadership
   - Our View of the Problem
   - Findings from Email data
2. 1st Project
   - Community Detection
   - Some Preliminary Results
3. 2nd Project
   - Resolution
   - Current Status
4. 3rd Project
   - Dynamic Evolution of Behavior and Structure

Big Picture

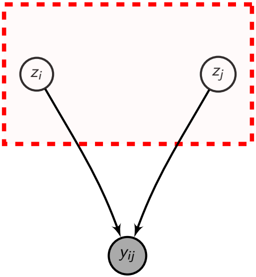


Figure: DGP for Networks

Generative Story of the Network Structure

- Individuals belong to a set of communities to different extents
    - membership probability vector
- Individuals take up different roles(activate communities) depending on whom they interact with
    - interaction based community indicator
- Individuals form a link to each other based on the chance of communication between announced communities
- This is mainly a mixed membership stochstic blockmodel, that allows for overlapping communities.

Generative Algorithm

$\forall k \in [1, .., K]$

DRAW THE DIAGONAL ELEMENTS OF THE BLOCK MATRIX B VIA $\beta_{k,k} \sim Beta(\eta_0, \eta_1)$

$\forall i \in \mathcal{N}$

DRAW THE MEAN OF THE LOGIT MIXED MEMBERSHIP VECTOR THROUGH $\boldsymbol{\mu} \sim Normal(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$

DRAW THE PRECISION OF THE LOGIT MIXED MEMBERSHIP VECTOR THROUGH $\boldsymbol{\Lambda} \sim Wishart(\ell_0, \boldsymbol{L}_0)$

DRAW A K-DIMENSIONAL VECTOR, $\boldsymbol{\theta}_i^* \sim Normal(\boldsymbol{\mu}, \boldsymbol{\Lambda})$

BUILD THE MEMBERSHIP VECTOR ON A SIMPLEX VIA LOGISTIC FUNCTION, $\boldsymbol{\theta}_{i,k} = \dfrac{exp(\boldsymbol{\theta}^*_{i,k})}{\sum_l exp(\boldsymbol{\theta}^*_{i,l})}$

$\forall (i,j) \in \mathcal{E}$

DRAW ONE-HOT MEMBERSHIP INDICATOR VECTOR FOR i WHEN CONTACTING j, $\boldsymbol{z}_{i\rightarrow j} \sim Categorical(\boldsymbol{\theta}_i)$

DRAW ONE-HOT MEMBERSHIP INDICATOR VECTOR FOR j WHEN CONTACTED BY I, $\boldsymbol{z}_{i\leftarrow j} \sim Categorical(\boldsymbol{\theta}_j)$

SAMPLE A LINK BETWEEN $i \longrightarrow j$ WITH PROBABILITY $\boldsymbol{z}_{i\rightarrow j}\boldsymbol{B}\boldsymbol{z}_{i\leftarrow j}$, $\boldsymbol{Y}(i,j) \sim Bern(\boldsymbol{z}_{i\rightarrow j}\boldsymbol{B}\boldsymbol{z}_{i\leftarrow j})$

Erasmus School of Economics

MMSB model Graphical Model

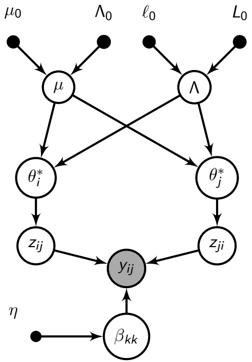


Figure: MMSB model

Inference Engine

- For all models we use Stochastic Variational Inference to get back paramaters
- Variational Inference(VI)
  - Approximating posterior through an easier tractable proposal distribution by minimizing the KL divergence between true posterior and proposal

- Stochastic Variational Inference(SVI)
  - Using Minibatch sampling to have a noisy estimate of the gradient with decreasing step size to optimize the variational objective
  - allows handling large datasets

- Speed up on 4 levels:
  - Variational engine
  - Stochastic minibatch sampling
  - Set informative update of parameters
  - Parallel implementation of batch updates(ongoing)

Outline

Accuracy(When ground truth is available)

- For simulated data, we assess the performance by using the Normalized Mutual Information(NMI)
- NMI scores are lied between 0, and 1.
- Given the ground truths communities, we can check the overlap in the clustering
- Higher NMI score signals better community detection
- For various simulated data, we achieve NMI score of 0.8 and above

Accuracy(When ground truth is unavailable)

- We use the accuracy measures on the holdout set
- We use the training set to train our model and learn the parameters
- We use perplexity (exponential of the predicted weighted average of the log likelihood of holdout sample)

Motivation

- On top of communities, behavioral data can help us with better labels
- We can redefine opinion leadership as a function of community roles
- "Latent" Homophily if not accounted for, still can be a problem in identifying social influence

Big Picture

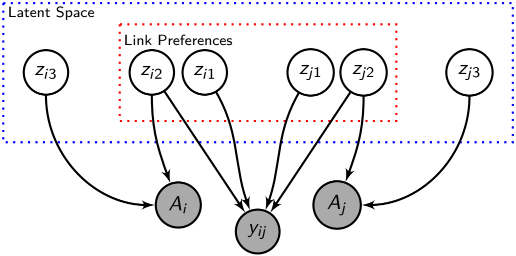


Figure: Graphical representation of network structure and behavior

When there is Social Influence

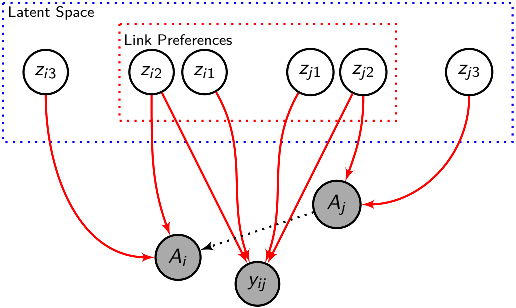


Figure: Graphical representation of network structure and behavior with influence-red signifies the open confounding path

Outline

Our Resolution

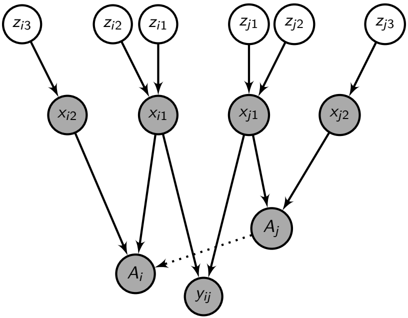


Figure: Solution to the confounding path

What we do in practice

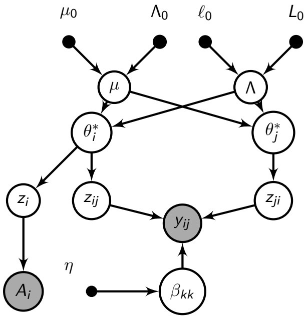


Figure: Modified graph for identification

Outline

Stage where we are

- We are still on the theory
- But have implementation recipe(variational derivations)

Outline

Motivation

- Static representation of networks can be naive for many networks
- ties form and severe
- As a result of endogenous network formation,
  - new behaviors may emerge
  - some behaviors may be discontinued

Big Picture

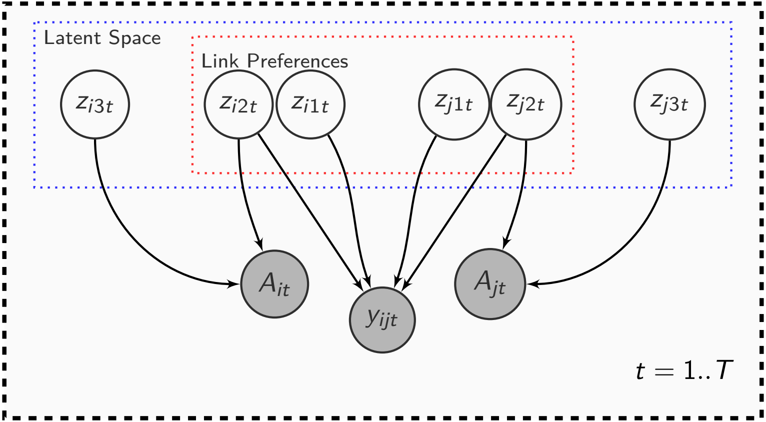


Figure: Dynamic evolution of network and behavior

Data that we use for now

- Simulated data:
  - small(100-250 nodes)
  - relatively larger networks(1000-10000 nodes).
- Real data: co-authorship data from the Marketing Science database submission
  - network data of authors in marketing from 1973 to 2009 are available on a yearly basis.
  - more than 25000 authors and approx. 1500 estimated communities(Ongoing)

Some speed up improvement

- Infromative sets for updating community specific parameters(done)
- Parallel computing(ongoing)
- Adjusting learning rate and mini-batch size(ongoing)

Questions

- Positioning
- Job Market
- Which aspect to be avoided or further followed

Questions

- My Github: https://github.com/56592aya/LNMMSB