A study of stochastic mixed membership models for link prediction in social networks

—

Dulac Adrien

Christine Largeron – Eric Gaussier

LIG / AMA

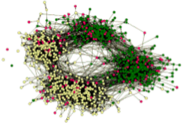
Juillet 2017

Table of contents

Summary

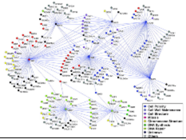
Complex Networks

Friendship Network

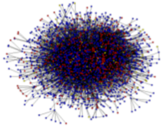


Business ties in US biotech-industry

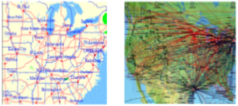
Genetic interaction network



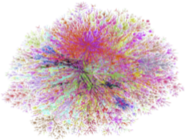
Protein-Protein Interaction Networks



Transportation Networks



Internet



Social Networks Properties

Important properties observed in social networks

Property	Measure
Small World	Diameter (Millgram 1978, Leskovec 2008)
Community	Modularity, Clustering Coefficient (Newman 2006)
Homophily	Homophily Coefficient (McPherson 2001)
Preferential Attachment / Burstiness	Degree Distribution (Barabási 1999)
Sparsity	Density (Haldous, Hoover 1981)

Statement

We focus on two models in the class of Mixed-Membership Models:

- IMMSB (infinite Mixed-Membership Stochastic Blockmodel)
- ILFM (infinite Latent Feature Model)

Do those models comply with preferential attachment?

homophily?

What is the impact on predictions?

Statement

We focus on two models in the class of Mixed-Membership Models:

- IMMSB (infinite Mixed-Membership Stochastic Blockmodel)
- ILFM (infinite Latent Feature Model)

Do those models comply with preferential attachment?

homophily?

What is the impact on predictions?

Claims

	Preferential Attachment		
	global	local	homophily
ILFM	No	No	depends on similarity
IMMSB	No	Yes	depends on similarity

Summary

Probabilistic Model

The Joint Distribution

$P(Y, \Theta \mid \sigma)$: How observations of a natural phenomena arise

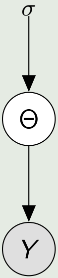
Likelihood

$P(Y \mid \Theta)$: How observations depend on the latent variables

Prior

$P(\Theta \mid \sigma)$: Describe the latent variables of the model

Graphical Model



Probabilistic Model

The Joint Distribution

$P(Y, \Theta \mid \sigma)$: How observations of a natural phenomena arise

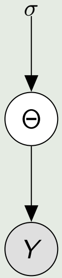
Likelihood

$P(Y \mid \Theta)$: How observations depend on the latent variables

Prior

$P(\Theta \mid \sigma)$: Describe the latent variables of the model

Graphical Model



The generative process

$$\Theta \sim P(\Theta \mid \sigma)$$
$$y \sim P(y \mid \Theta)$$

Inference / Optimisation

Posterior Distribution: $P(\Theta \mid Y) = \frac{P(Y|\Theta)P(\Theta)}{P(Y)}$

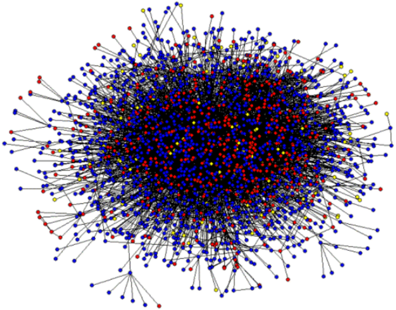
Social Network to Graph

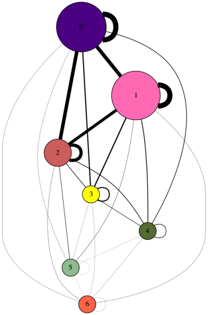
A graph : $G = (V, E)$

- V: set of vertices, drawn as nodes,
- E: set of edges, drawn as lines connecting vertices,

containing N nodes : $|V| = N$

Define a adjacency matrix : $Y = (y_{ij})_{i,j \in V^2} \left| \begin{array}{ll} y_{ij}{=}1 & \text{if } (i,j){\in}E \\ y_{ij}{=}0 & \text{otherwise} \end{array} \right.$





Random Graph model

Given a Graph $G = (V, E)$, and F and Φ such that:

$$Y \sim P(Y|\ F, \Phi) = h(F \Phi F^T)$$

F : a latent feature matrix (describe the membership of nodes)

Φ : a latent weight matrix (describe the correlation between classes)

h : a mapping functions to a probability space.

Random Graph model

Given a Graph $G = (V, E)$, and F and Φ such that:

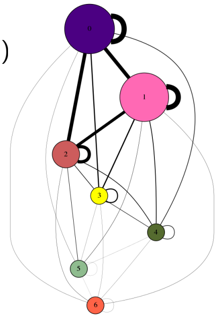
$$Y \sim P(Y|\ F, \Phi) = h(F\Phi F^T)$$

F : a latent feature matrix (describe the membership of nodes)

Φ : a latent weight matrix (describe the correlation between classes)

h : a mapping functions to a probability space.

$$\underset{N \times N}{Y} \sim h\big(\underset{N \times K}{F} \ \underset{K \times K}{\Phi} \ F^T\big)$$



Likelihood

$$P(y_{ij} = 1|F, \Phi) = f_i \Phi f_j^T$$

What prior over F and Φ ?

Stochastic Blockmodel

The Stochastic Blockmodel[1]

- We have K blocks and,
- Any node i belongs to one block c:
- Each weights $(\phi_{kl})_{k,l \in (1,..,K)^2}$, encode the strength between two blocks.

One Example:

Block membership for node 4 and 7 (K=3 and N=9)

$f_4 = \boxed{0 \mid 1 \mid 0}$ $\quad f_7 = \boxed{0 \mid 0 \mid 1}$ $\quad \phi_{12} = 0.3$

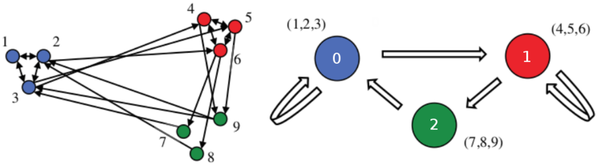


Figure: Stochastic Blockmodel

[1] Anna Goldenberg et al. "A survey of statistical network models". In: *Foundations and Trends® in Machine Learning* 2.2 (2010), pp. 129–233.

Mixed Membership Models

Relax the single membership assumption (soft clustering).

Binary features

$$f_i = \begin{array}{|c|c|c|c|c|} \hline 1 & 0 & 1 & 1 & 0 \\ \hline \end{array}$$

Also relax the fixed number of features K with an Indian Buffet Process (IBP).

or

Proportion features

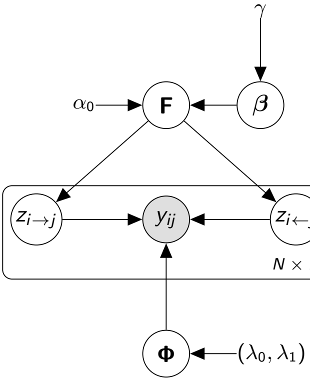
$$f_i = \begin{array}{|c|c|c|c|c|} \hline .5 & .2 & .1 & .1 & .1 \\ \hline \end{array} \qquad \sum f_i = 1$$

Also relax the fixed number of features K with a (Hierarchical) Dirichlet Process (DP).

IMMSB

Infinite Mixed Membership Stochastic Model (IMMSB)[2]

$$
\begin{aligned}
&\boldsymbol{\beta} \sim \mathrm{GEM}(\gamma) \\
&\text{For each } i \in \{1, .., N\} \\
&\qquad \boldsymbol{f}_i \sim \mathrm{DP}(\alpha_0, \boldsymbol{\beta}) \\
&\text{For each } (m, n) \in \{1, .., K\}^2 \\
&\qquad \phi_{mn} \sim \mathrm{Beta}(\lambda_0, \lambda_1) \\
&\text{For each } (i, j) \in V^2 \\
&\qquad z_{i\rightarrow j} \sim \mathrm{Mult}(\boldsymbol{f}_i) \\
&\qquad z_{i\leftarrow j} \sim \mathrm{Mult}(\boldsymbol{f}_j) \\
&\qquad y_{ij} \sim \mathrm{Bern}(z_{i\rightarrow j} \Phi z_{i\leftarrow j}^T)
\end{aligned}
$$

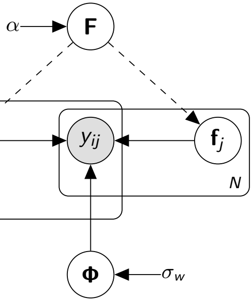


[2] Phaedon-Stelios Koutsourelakis and Tina Eliassi-Rad. "Finding Mixed-Memberships in Social Networks.". In: *AAAI Spring Symposium: Social Information Processing*. 2008, pp. 48–53.

ILFM

Infinite Latent Feature Model (ILFM)[3]

$$F \sim \mathrm{IBP}(\alpha)$$
$$\text{For each } (m,n) \in \{1,..,K\}^2$$
$$\phi_{mn} \sim \mathrm{Gaussian}(0, \sigma_w)$$
$$\text{For each } (i,j) \in V^2$$
$$y_{ij} \sim \mathrm{Bern}(\boldsymbol{f}_i \Phi \boldsymbol{f}_j^T)$$



[3] Kurt Miller, Michael I Jordan, and Thomas L Griffiths. "Nonparametric latent feature models for link prediction". In: *Advances in Neural Information Processing Systems*. 2009, pp. 1276–1284.

Summary

Preferential attachment

Definition (Degrees)

Let $\mathcal{M}_e = \{F, \Phi\}$ be a probabilistic link prediction model such that

$$Y \sim P(Y|\mathcal{M}_e)$$

Let $\mathcal{C}_k$ be the set of nodes belonging to the class k.

- Global degree : $d_i = \sum_{j \neq i} y_{ij}$
- Local degree : $d_{ik} = \sum_{j \neq i} y_{ij} \mathbb{1}(i, j \in \mathcal{C}_k)$
- Expected Local degree : $D_{ik} = \sum_{j \neq i} P(y_{ij} = 1, i, j \in \mathcal{C}_k | \ \mathcal{M}_e)$

Preferential attachment (cont.)

The rich get richer...

Definition (Global Preferential attachment)

we say that $\mathcal{M}_e$ satisfies the global preferential attachment iff, for any node i,

$$P(d_i \geq n+1 \mid d_i \geq n, \mathcal{M}_e) \text{ increases with } n \in \{0, ..., N-2\}$$

Definition (Local Preferential attachment)

we say that $\mathcal{M}_e$ satisfies the local preferential attachment iff, for any node i,

$$P(D_{i,k} \geq x + \epsilon \mid D_{i,k} \geq x, \mathcal{M}_e) \text{ increases with } x \in [0, N[$$

Homophily

Birds of a feather flock together...

Definition (Homophily)

Let s be a similarity measure between nodes. We say that $\mathcal{M}_e$ is homophilic under the similarity s iff, $\forall (i,j,i',j') \in V^4$:

$$s(i,j) > s(i',j') \implies P(y_{ij} = 1 \mid \mathcal{M}_e) > P(y_{i'j'} = 1 \mid \mathcal{M}_e)$$

Similarities

Natural similariry : $s_n = f_i \Phi f_j \top$

Latent similarity : $s_l = f_i f_j^\top$

Summary

Theoretical Results

Summary of our theoretical results.

Theoretical results

	Preferential Attachment		
	global	local	homophily
ILFM	No	No	depends on similarity
IMMSB	No	Yes	depends on similarity

Empirical Results (Datasets)

Table: Characteristics of artificial and real networks.

Networks	nodes	edges	density
Network1	1000	3507	0.007
Network2	1000	31000	0.062
Blogs	1490	20512	0.009
Manufacturing	167	5950	0.215

Table: Preferential attachment measures for training datasets and networks generated with fitted models.

Training Datasets	Global		Local	
	p-value	α	*p*-value	α
Network1	1	2.4	1.0 ± 0.0	1.8 ± 0.03
Network2	0	1.3	0.0 ± 0.0	1.2 ± 0.01
Blogs	1	1.5	1.0 ± 0.0	1.4 ± 0.03
Manufacturing	0	1.4	0.4 ± 0.3	1.3 ± 0.05

Empirical Results (Preferential attachment I)

Figure: Synthetic Datasets

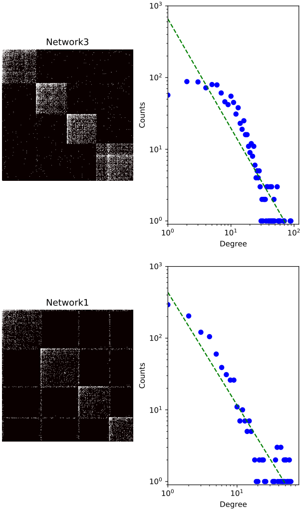
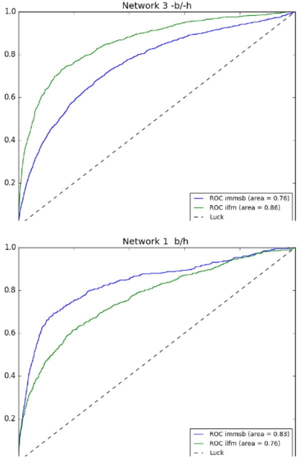
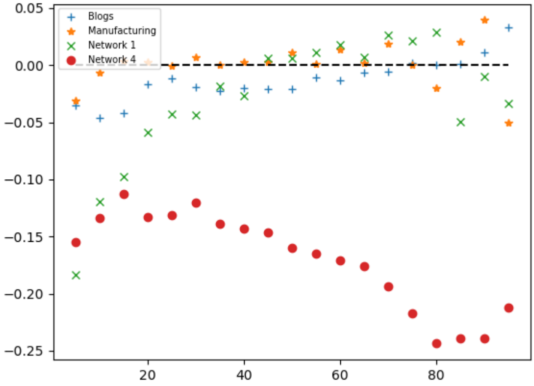


Figure: Prediction Performance



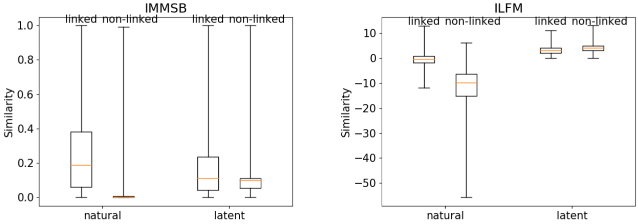
Empirical Results (Preferential attachment II)

Figure: Comparison performance of IMMSB and ILFM based on the size of the training set.



Empirical Results (Homophily)

Figure: Natural and latent similarities aggregated over all datasets and computed on linked and non-linked pairs of nodes for IMMSB (top) and ILFM (bottom).



Conclusion

Wrap up

- Formal Definition of social networks properties [Homophily and Preferential Attachments].
- Study of compliance of Mixed Membership Models with the properties.
- Empirical implications on prediction performances.

Future work

- Extend to weighted networks (and scalable inference).
- Extend to temporal networks.

Bottom

Thank You

Limitation and future work

Representation Theorem for exchangeable graph[4].

Aldous-Hoover Theorem

An array $(Y_{ij})_{ij \in \mathbb{N}}$ is jointly exchangeable if and only if :

$$Y_{ij} \sim F(U_i, U_j, U_{ij}) \tag{1}$$

- With $F : [0,1]^3 \rightarrow [0,1]$ a random measure,
- $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij \in \mathbb{N}})$ i.i.d Uniform[0,1] random variables.

A corollary of this theorem is that random exchangeable graph are either **empty or dense**.

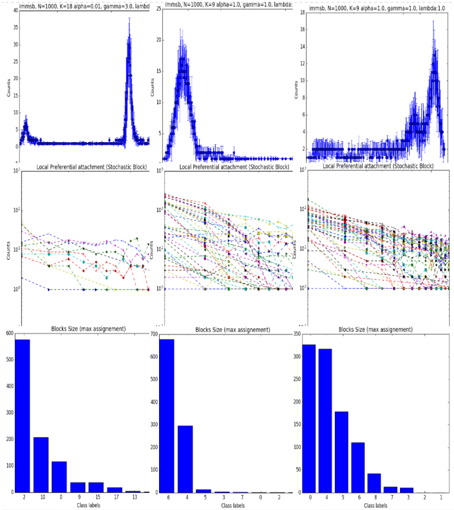
→ What network structure and invariance to comply with the Global preferential attachment ?

[4] Peter Orbanz and Daniel M Roy. "Bayesian models of graphs, arrays and other exchangeable random structures". In: *IEEE transactions on pattern analysis and machine intelligence* 37.2 (2015), pp. 437–461.

Characterisation of Burstiness (cont .)

Generative Process for IMMSB.

Figure: IMMSB Simulation.



Characterisation of Burstiness (cont .)

Is there a proof ?

$$P(d_i = n|\ \alpha) \quad ?$$

Problem

- d_i random variable ?
- $P(\boldsymbol{Y}_i \mid \alpha, \lambda) = \int_F \int_\Phi p(f_i|\alpha) p(\Phi|\lambda) \prod_{j \neq i} p(f_j|\alpha) p(y_{ij} \mid f_i, f_j, \Phi)\ dF d\Phi$

Intractable...