Bayesian Nonparametric Network Models
Latent Space and Latent Attribute Approaches

James Robert Lloyd

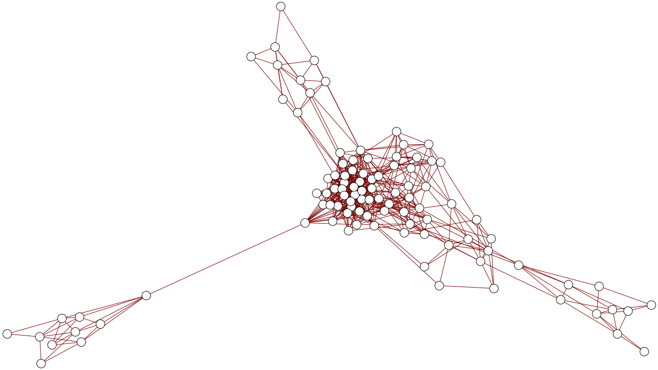
Machine Learning Group,
Department of Engineering,
University of Cambridge

June 2013

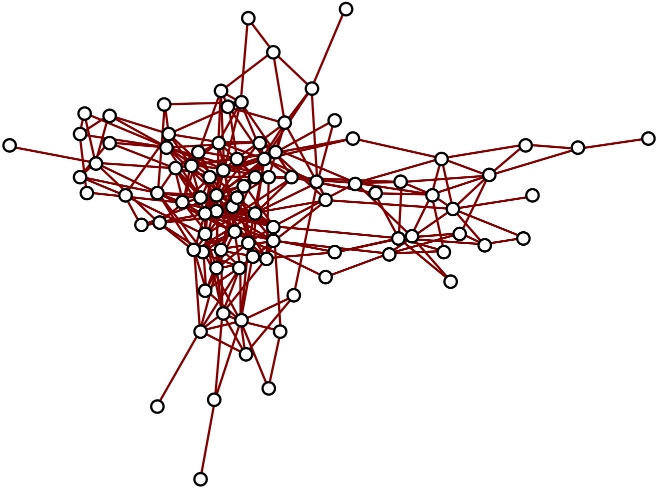
Collaborators

Daniel M. Roy (Cambridge)
Peter Orbanz (Columbia)
Zoubin Ghahramani (Cambridge)

e.g., Block / clique structure



e.g., Transitivity / latent space



PROBABILISTIC LATENT VARIABLE MODELS

- Each node is assigned a latent variable, say U_i
- Latent variable determines stochastic properties of each node i.e., existence of a link between nodes i and j, $(X_{ij} = 1)$, depends stochastically on U_i and U_j

Example : Latent class / block models [WW87]

U_i	$\sim_{\text{iid}}$	$\text{Multinomial}(K)$	-	Nodes assigned latent classes
Λ_{ij}	$\sim_{\text{iid}}$	Beta	-	Independent class interaction probabilities
W_{ij}	$:=$	$\Lambda_{U_i U_j}$	-	Node interaction probabilties depend on classes
X_{ij}	$\sim$	$\text{Bernoulli}(W_{ij})$	-	Bernoulli likelihood

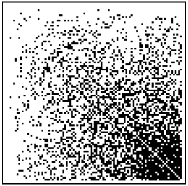
Example: Distance models [HRH02]

U_i	$\sim_{\text{iid}}$	$\mathcal{N}(0, I)$	-	Nodes assigned latent positions
d_{ij}	$:=$	$\lvert U_i - U_j \rvert$	-	Distances between latent positions
W_{ij}	$:=$	$\alpha - \beta d_{ij}$	-	‘Affinity’ of nodes decays with distance
X_{ij}	$\sim$	$\text{Bernoulli}(\sigma(W_{ij}))$	-	Bernoulli sigmoid likelihood

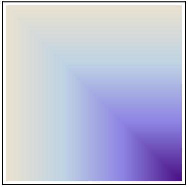
Results from probability theory…

- We assume the nodes of the network are exchangeable i.e., have no ordering
- We demonstrate a characterisation of all probability distributions for exchangeable networks
- We discuss the types of structures that can be found using two particular models

…inspire directly modelling adjacency matrices with a latent variable model

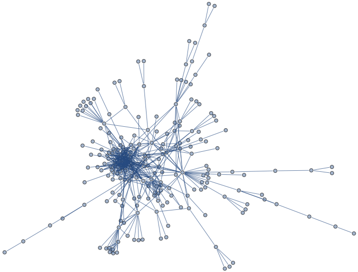


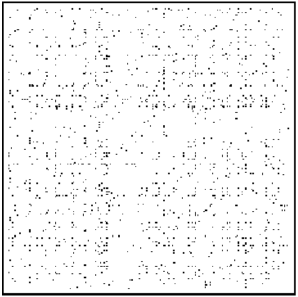
$\approx$

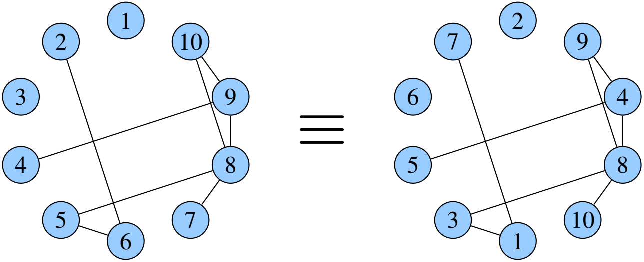


NETWORKS TYPICALLY REPRESENTED BY ARRAYS

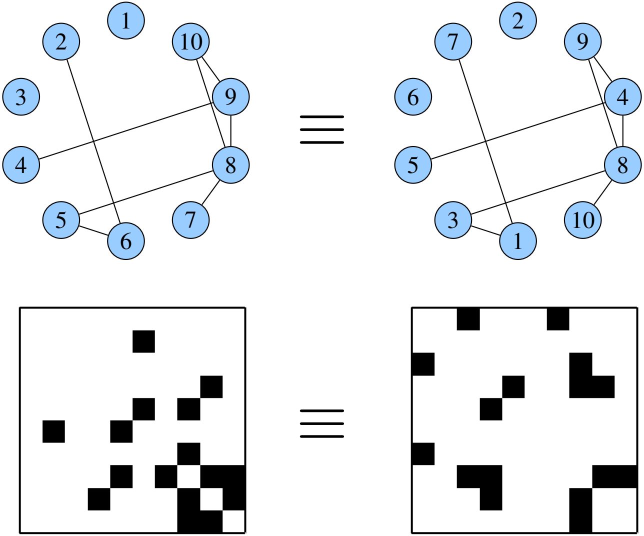
e.g., a protein interactome represented by its adjacency matrix






1
2
10
3
9
4
8
5
7
6
≡
2
7
9
6
4
5
8
3
10
1

Exchangeability for corresponding arrays



EXCHANGEABILITY CAN BE CHARACTERISED

Definition

An array $X = (X_{ij})_{i,j \in \mathbb{N}}$ is called a (jointly/weakly) *exchangeable array* if

$$(X_{ij}) \overset{\mathrm{d}}{=} (X_{\pi(i)\pi(j)}) \qquad \text{for every } \pi \in \mathbb{S}_\infty \ .$$

Theorem (Aldous [Ald81], Hoover [Hoo82])

A random 2-array (X_{ij}) is exchangeable if and only if there exists a random (measurable) function $F : [0,1]^3 \to \mathcal{X}$ such that

$$(X_{ij}) \overset{d}{=} (F(U_i, U_j, U_{ij})).$$

where $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i \leq j \in \mathbb{N}}$ are i.i.d. $\text{Uniform}[0,1]$ *random variables and $U_{ji} = U_{ij}$ for $j < i \in \mathbb{N}$.*

A simpler representation can be used

Call an array (X_{ij}) *simple* if it admits a representation

$$(X_{ij}) \stackrel{d}{=} (\Theta(U_i, U_j))$$

where $\Theta : [0,1]^2 \to \mathcal{X}$ is a random measurable function and $(U_i)_{i \in \mathbb{N}}$ are i.i.d. Uniform$[0,1]$ random variables.

Let $\mathcal{L}(Y)$ be the law (distribution) of a random variable Y and define $\chi_m X := (X_{ij};\ i,j \leq m)$.

Theorem (Kallenberg [Kal99])

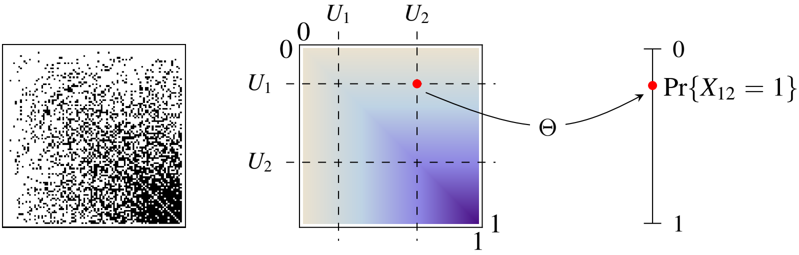
Let X be an exchangeable array in a Borel space $\mathcal{X}$. Then there exist some simple exchangeable arrays $X_1, X_2, \ldots$ such that $\mathcal{L}(\chi_m X_n)$ and $\mathcal{L}(\chi_m X)$ are mutually absolutely continuous for all $m, n \in \mathbb{N}$ and the associated Radon–Nikodym derivatives converge uniformly to 1 as $n \to \infty$ for fixed m.

Directly modelling adjacency matrices

Representation results provide a generic modelling recipe

Θ	-	Adjacency matrix approximated by function on unit square
U_i	-	Each node associated with a latent variable in $[0, 1]$
$W_{ij} := \Theta(U_i, U_j)$	-	Evaluation of approximate adjacency matrix
$X_{ij} \sim \text{Bernoulli}(W_{ij})$	-	Bernoulli likelihood (can be shown to be general)

Θ can be pictured as blurred adjacency matrix



EXAMPLES THAT FIT THIS PATTERN

Note

- U_i not restricted to be Uniform$[0, 1]$ - used in theorems as canonical distribution
- W_{ij} often specified directly, but function Θ can often be characterised

Latent class / block models

$$
\begin{array}{lcllll}
U_i & \sim_{\text{iid}} & \text{Multinomial}(K) & - & \text{Nodes assigned latent classes} \\
\Lambda_{ij} & \sim_{\text{iid}} & \text{Beta} & - & \text{Independent class interaction probabilities} \\
W_{ij} & := & \Lambda_{U_i U_j} & - & \text{Node interaction probabilties depend on classes}
\end{array}
$$

Distance models

$$
\begin{array}{lcllll}
U_i & \sim_{\text{iid}} & \mathcal{N}(0, I) & - & \text{Nodes assigned latent positions} \\
d_{ij} & := & |U_i - U_j| & - & \text{Distances between latent positions} \\
W_{ij} & := & \sigma(\alpha - \beta d_{ij}) & - & \text{Probability of interaction decays with distance}
\end{array}
$$

Many other models fit this pattern

	W_{ij}	κ	$U_i \sim$.
Random function model	$\phi(U_i)'\Lambda$	$\kappa_{U\times U}$	Gaussian
SMGB, InfTucker	$\phi(U_i)'\Lambda\phi(U_j)$	$\kappa_U \otimes \kappa_U$	Laplace
GPLVM	$\phi(U_i)'\Lambda$	$\kappa_U \otimes \delta$	Gaussian
Eigenmodel	$U_i'\Lambda U_j$	$L_U \otimes L_U$	Gaussian
Linear relational GP	$U_i'\Lambda U_j$	$L_U \otimes L_U$	Gaussian
PCA, PMF	$U_i'\Lambda$	$L_U \otimes \delta$	Gaussian
Latent distance	$-\lvert U_i - U_j\rvert$	0	Gaussian
Mondrian process based	Decision tree	*	Uniform
Latent class	$\Lambda_{U_iU_j}$	$\delta_{U\times U}$	Multinomial
IRM, IHRM	$\Lambda_{U_iU_j}$	$\delta_{U\times U}$	CRP
BMF, LFRM	$U_i'\Lambda U_j$	$L_U \otimes L_U$	IBP
ILA	$\sum_d \mathbb{I}_{U_{id}}\mathbb{I}_{U_{jd}}\Lambda^{(d)}_{U_{id}U_{jd}}$	*	CRP + IBP

Notes

κ is the kernel in the often equivalent Gaussian process representation; ϕ is the corresponding feature map. L is a linear kernel, δ is the Kronecker delta function, $\otimes$ is a tensor / Kronecker product. Λ is a matrix. $\mathbb{I}$ is an indicator function.

EXAMPLE: RANDOM FUNCTION MODEL (RFM)

Directly model smoothed adjacency matrix [LOGR12]

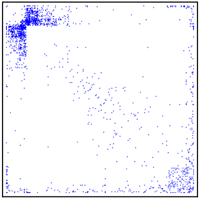
U_i	$\sim_{\text{iid}}$	$\mathcal{N}(0, I)$	-	Nodes embedded in latent space
Θ	$\sim$	$\mathcal{GP}(0, \kappa)$	-	Adjacency matrix modelled by Gaussian process
W_{ij}	$:=$	$\Theta(U_i, U_j)$	-	Evaluation of smoothed adjacency matrix
X_{ij}	$\sim$	$\text{Bernoulli}(\sigma(W_{ij}))$	-	Bernoulli sigmoid likelihood

Observations

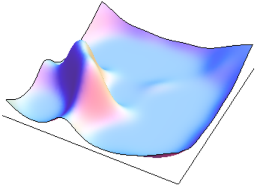
- Gaussian processes can approximate any measurable function, so this model can approximate any exchangeable distribution for networks
- Model will favour functions Θ that are smooth
- Smoothness can be seen when reordering adjacency matrices using U_i learnt from data, resulting in visually smooth adjacency matrix

RFM FINDS SMOOTH STRUCTURES

A protein interactome

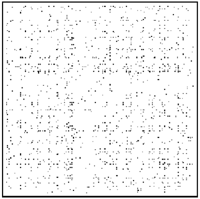


Adjacency matrix sorted by MAP embedding

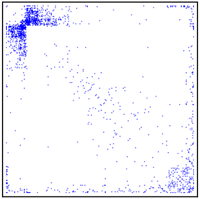


MAP Θ

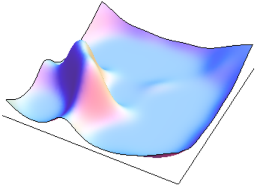
RFM FINDS SMOOTH STRUCTURES



Unsorted adjacency matrix

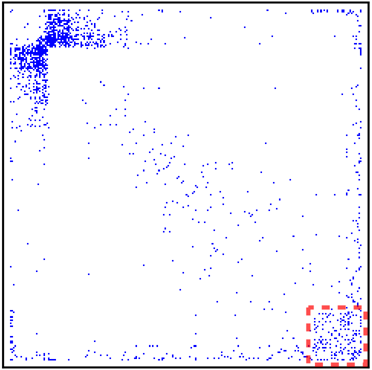


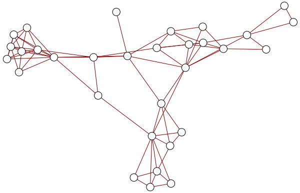
Adjacency matrix sorted by MAP embedding



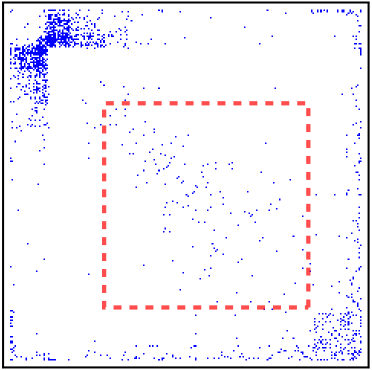
MAP Θ

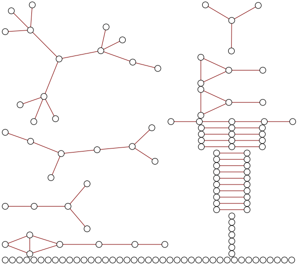
RFM: Block structure



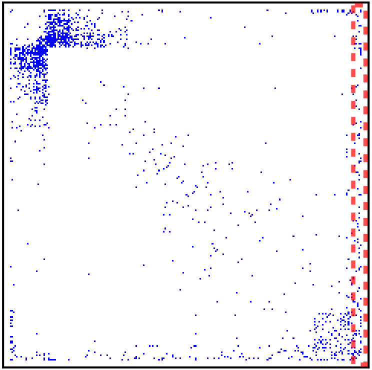


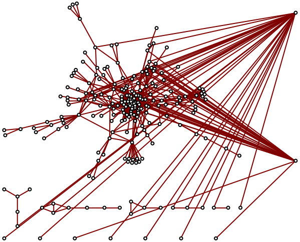
RFM: Sparse with some transitivity



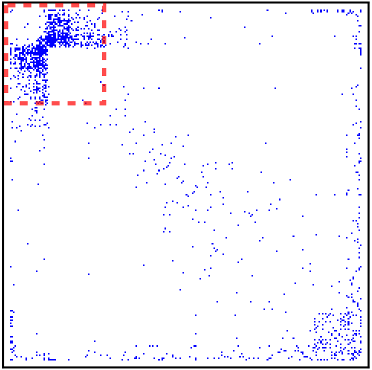


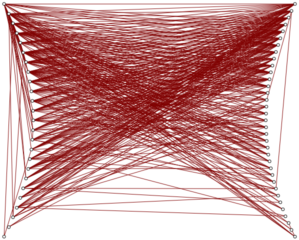
RFM: Hub nodes



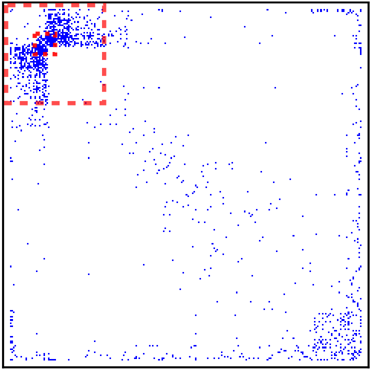


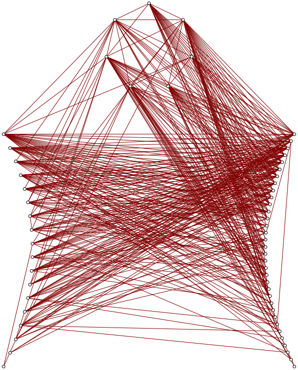
RFM: Ordered and almost bipartite



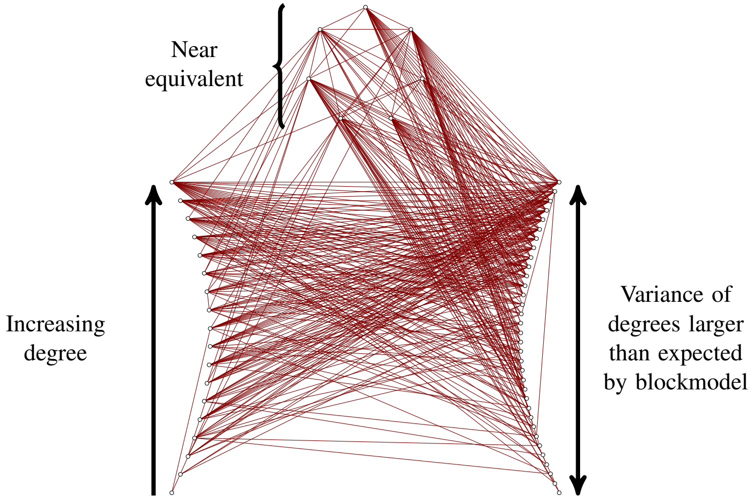


RFM: Ordered and almost bipartite





RFM: Ordered and almost bipartite



Good predictions with low dimension variables

RFM performs well on prediction tasks...

5 fold cross validation AUC results

Data set	High school			NIPS			Protein		
Latent dim.	1	2	3	1	2	3	1	2	3
PMF	0.747	0.792	0.792	0.729	0.789	0.820	0.787	0.810	0.841
Eigenmodel	0.742	0.806	0.806	0.789	0.818	0.845	0.805	0.866	0.882
GPLVM	0.744	0.775	0.782	0.888	0.876	0.883	0.877	0.883	0.873
RFM	**0.815**	**0.827**	**0.820**	**0.907**	**0.914**	**0.919**	**0.903**	**0.910**	**0.912**

...even with low dimensional latent space

- Performance of RFM with one dimensional latent space outperformed all benchmarks with up to three dimensions
- Benchmarks include Hoff's eigenmodel [Hof08] which empirically outperforms block models and latent distance models
- High predictive performance with low dimensional latent space may lead to interpretability

An alternative interpretation

- RFM assumes simple priors on latent variables, but allows Θ to be any function
- Alternative modelling paradigm is to use highly structured priors for the latent variables and a simple Θ
- The Inifinite Latent Attribute model assumes a multiple clustering prior for the latent variables and a linear Θ

Motivation for latent attribute model

- Imagine a social network in a collegiate university. Friendships may arise based on attributes / features each person has and their values

- e.g., A person may be a member of a college
  - This is then partitioned by the different colleges e.g., King's, Trinity etc.

- e.g., A person may play a sport
  - This is then partitioned by the different sports e.g., Tennis, Hockey etc.

- This type of structure can be succinctly expressed by multiple overlapping clusterings

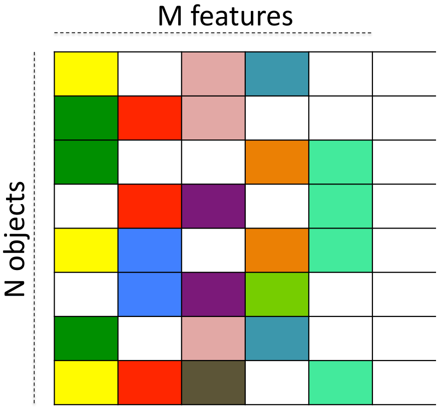
A multiple clustering model using highly structured latent variables [PKG12]

- Each node assumed to posses some collection of attributes / features
  - Specified by IBP prior
- Within each feature nodes are assumed to belong to a cluster
  - Specified by CRP prior

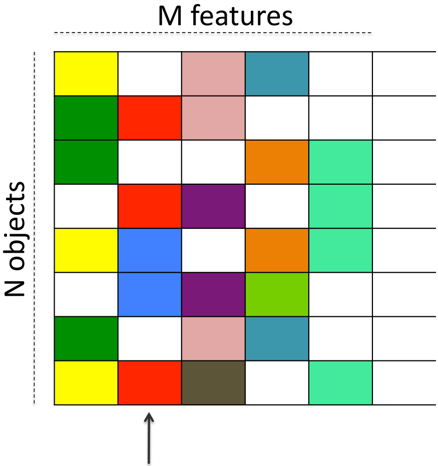
Generative Model

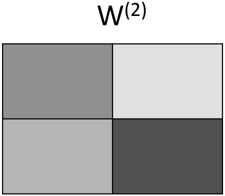
$$
\begin{aligned}
\mathbf{Z}|\alpha &\sim \text{IBP}(\alpha) \\
\mathbf{c}^{(m)}|\gamma &\sim \text{CRP}(\gamma), \text{where } m \in \{1, \ldots M\} \\
\lambda_{kk'}^{(m)}|\sigma_w &\sim N(0, \sigma^2), \text{where } k, k' \in \{1, \ldots, K^{(m)}\} \\
W_{ij} &= \sum_m z_{im} z_{jm} w_{c_i^m c_j^m}^{(m)} + s.
\end{aligned}
$$

The Infinite Latent Attribute Model



THE INFINITE LATENT ATTRIBUTE MODEL





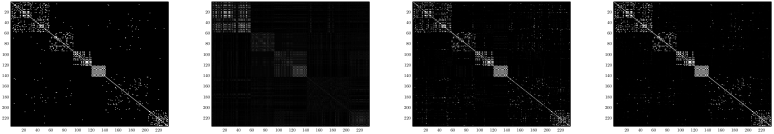
ILA CAN PRODUCE VERY ACCURATE PREDICTIONS

NIPS coauthorship network prediction

Cross validation on NIPS 1-17 coauthorship dataset (Globerson et al., 2007). 234 most connected authors, 10 repeats, holding out 20% of the data. ILA 500 iterations, IRM and LFRM 1000 iterations.

	IRM	LFRM	ILA ($M = \infty$)
Test error (0-1 loss)	0.0440 ± 0.0014	0.0228 ± 0.0041	$\mathbf{0.0106 \pm 0.0007}$
Test log likelihood	-0.0859 ± 0.0043	-0.0547 ± 0.0079	$\mathbf{-0.0318 \pm 0.0094}$
AUC	0.9565 ± 0.0037	0.9631 ± 0.0150	$\mathbf{0.9910 \pm 0.0056}$

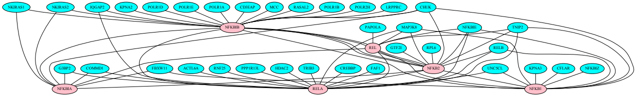
Visualisation of link prediction



True links IRM LFRM ILA

The lighter the entry, the more confident the model is that the corresponding authors would collaborate.

ILA finds disconnected group in protein interactome…



…corresponding to a feature with two sub-clusters

- ILA has identified similar structure to RFM but automatically identifies it as a separate sub-clustering
- Ongoing work to better intepret this model and find biologically interesting / relevant structures

Summary

- Latent variable models are a generic way to probabilistically model networks
  - Assuming exchangeability, networks can be modelled by a random function on the unit square
  - Framework encompasses many standard models of networks
- Introduced the RFM which directly instantiates the random function representation
  - Inference in RFM reveals block and latent space structure, as well as hub nodes and other structures
  - Good predictive performance even with low dimensional node latent variables
- Also briefly discussed the multiple clustering model ILA
  - Automatically reveals structures similar to those found by the RFM
  - ILA also has excellent predictive performance

References I

[Ald81] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[Hof08] Peter D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 657–664, 2008.

[Hoo82] D N Hoover. Row-column exchangeability and a generalized model for probability. In *Exchangeability in Probability and Statistics*, pages 281–291, 1982.

[HRH02] Peter D. Hoff, Adrian E Raftery, and Mark S Handcock. Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098, December 2002.

[Kal99] Olav Kallenberg. Multivariate sampling and the estimation problem for exchangeable arrays. *Journal of Theoretical Probability*, 12(3):859–883, 1999.

[LOGR12] James Robert Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M. Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, 2012.

[PKG12] Konstantina Palla, David A. Knowles, and Zoubin Ghahramani. An infinite latent attribute model for network data. In *Proceedings of the 29th International Conference on Machine Learning*, ICML 2012. Edinburgh, Scotland, GB, July 2012.

[WW87] Yuchung J. Wang and George Y. Wong. Stochastic Blockmodels for Directed Graphs. *Journal of the American Statistical Association*, 82(397):8–19, 1987.