Comparative Study for Inference of Hidden Classes in Stochastic Block Models

Pan Zhang
ESPCI, France

Florent Krzakala (*ESPCI, France*)
Jörg Reichardt (*University of Würzburg, German*)
Lenka Zdeborová (*CEA and CNRS, France*)

Block (community, clustering) detection

Spectral methods

- Adjacent matrix
- Random-walk matrix
- Modularity matrix

...

Model fitting

- Monte-Carlo
- Variational Mean-field
- Belief Propagation

...

Block (community, clustering) detection

Spectral methods

- Adjacent matrix
- Random-walk matrix
- Modularity matrix

...

Model fitting

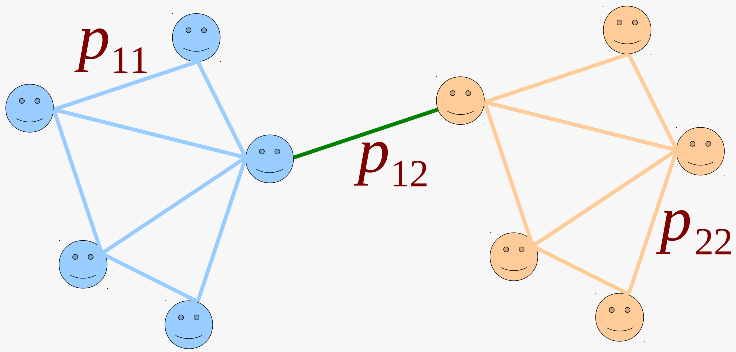
- Monte-Carlo
- Variational Mean-field
- Belief Propagation

...

Our result: Belief propagation shows exceptional performance in block detection and Inference of Stochastic Block Model in relational data.

Stochastic Block model

Stochastic Block model and its variations have been widely used in modeling real-world, social, biological and Internet networks.



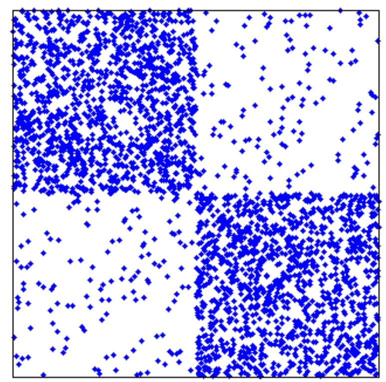
$$N=10$$
$$q=2$$
$$n_1=n_2=0.5$$
$$p_{11}=p_{22}=0.5$$
$$p_{12}=0.1$$

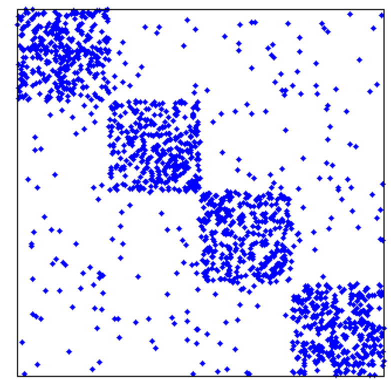
Generate a random network as follows:

- N nodes
- q groups
- n_a proportion of nodes in group a = 1,...,q
- p_{ab} probability that an edge present between nodes from group a and group b

Examples of adjacency matrices of generated graphs

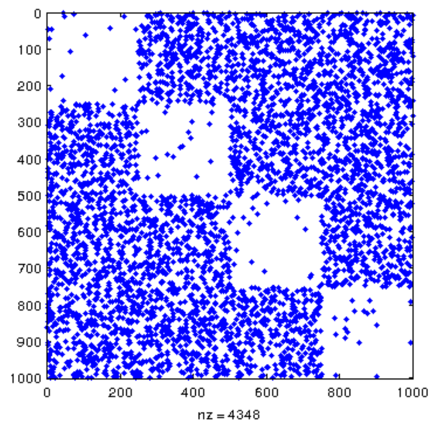
$q=2$
$n_a=n_b=0.5$
$p_{aa}=5.455/N$
$p_{ab}=0.5455/N$

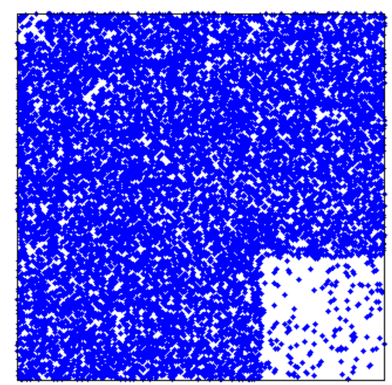




$q=4$
$n_a=1/4$
$p_{aa}=10/N$
$p_{ab}=0.5/N$

$q=4$
$n_a=1/4$
$p_{aa}=0.5/N$
$p_{ab}=10/N$





$q=2$
$n_1=2/3$
$n_2=1/3$
$p_{11}=120/N$
$p_{12}=114/N$
$p_{21}=114/N$
$p_{22}=12/N$

Inference and learning of the Stochastic Block Model

$$P\left(A_{ij},\{q_i\}\,|\,\underline{\{n_a\},\{p_{ab}\}}\right)=\prod_{i<j}\left[p_{q_i,q_j}^{A_{ij}}\left(1-p_{q_i,q_j}\right)^{(1-A_{ij})}\right]\prod_i n_{q_i}$$

Adjacency matrix
(Observables)

Parameters

Hidden assignment

Inference and learning of the Stochastic Block Model

$$P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})=\prod_{i<j}\left[p_{q_i,q_j}^{A_{ij}}(1-p_{q_i,q_j})^{(1-A_{ij})}\right]\prod_i n_{q_i}$$

Adjacency matrix
(Observables)

Parameters

Hidden assignment

Inference: find hidden assignment given parameters

$$P(\{q_i\}|A_{ij},\{n_a\},\{p_{ab}\})=\frac{P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})}{\sum_{\{q_i\}}P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})}$$

Inference and learning of the Stochastic Block Model

$$P(A_{ij},\{q_i\}|\underline{\{n_a\},\{p_{ab}\}})=\prod_{i<j}\left[p_{q_i,q_j}^{A_{ij}}(1-p_{q_i,q_j})^{(1-A_{ij})}\right]\prod_i n_{q_i}$$

Adjacency matrix
(Observables)

Parameters

Hidden assignment

Inference: find hidden assignment given parameters

$$P(\{q_i\}|A_{ij},\{n_a\},\{p_{ab}\})=\frac{P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})}{\sum_{\{q_i\}}P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})}$$

Learning: find parameters which maximize the log-likelihood

$$L(\{n_a\},\{p_{ab}\})=\log P(A|\{n_a\},\{p_{ab}\})=\log\sum_{\{q_i\}}P(A_{ij},\{q_i\}|\{n_a\},\{p_{ab}\})$$

Learning: Expectation-Maximization

At the maximum point of the log-likelihood:

$$\left\langle \sum_i \delta_{a,q_i} \right\rangle = n_a N$$

Number of nodes of group a

$$\left\langle \sum_{i<j} \delta_{a,q_i} \delta_{b,q_j} \right\rangle = p_{ab} n_a n_b N^2$$

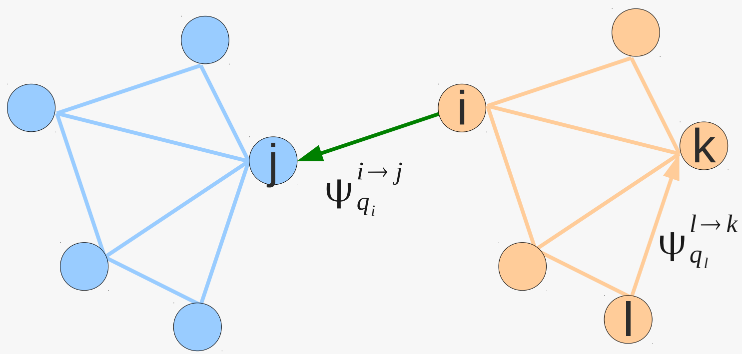
Number of edges between group a and group b

E-step

* Exact enumeration ?
* Monte Carlo (Gibbs Sampling) ?
* Variational mean-field approximation? Daudin et al 2008
* Belief Propagation ? Decelle et al 2011

Belief Propagation

- Statistical physics - Bethe-Peierls'35, cavity method Mezard, Parisi'01
- Information theory - LDPC codes R.G. Gallager'62
- Machine learning - Bayesian inference J. Pearl'82



☆ Exact on trees

☆ Good approximation in many (random) systems

Belief Propagation for Stochastic Block Model

$$\psi_{q_i}^{i\to j} = \frac{1}{Z^{i\to j}} n_{q_i} \prod_{k\neq i,j} \sum_{q_k} p_{q_i q_k}^{A_{ik}} (1-p_{q_i q_k})^{1-A_{ik}} \psi_{q_i}^{k\to i} \qquad O(q^2 N^2)$$

$$\psi_{q_i}^{i} = \frac{1}{Z^{i}} n_{q_i} \prod_{k\neq i} \sum_{q_k} p_{q_i q_k}^{A_{ik}} (1-p_{q_i q_k})^{1-A_{ik}} \psi_{q_i}^{k\to i}$$

Marginals

Decelle et al 2011

BP on sparse graphs

$$\psi_{q_i}^{i\to j} = \frac{1}{Z^{i\to j}} n_{q_i} e^{-h_{q_i}} \prod_{k\in i\setminus j} \sum_{q_k} p_{q_i q_k}^{A_{ik}} (1-p_{q_i q_k})^{1-A_{ik}} \psi_{q_i}^{k\to i}$$

$O(q^2 Nc)$

$$h_{q_i} = \frac{1}{N} \sum_k \sum_{q_k} p_{q_i q_k} \psi_{q_k}^k$$

k belongs to neighbors
of I, except j

EM learning with BP

$$n_a = \frac{1}{N} \sum_i \psi_a^i$$

$$p_{ab} = \frac{1}{N^2} \frac{1}{n_a n_b} \frac{\sum_{(a,b)\in E} p_{ab} (\psi_a^{i\to j} \psi_b \; j \to i) + (\psi_a^{j \to i} \psi_b \; i \to j)}{Z^{ij}}$$

Stationary conditions for the Bethe free energy

EM is sensitive to the initial parameters!

Belief Propagation

$$\psi_{q_i}^{i\to j} = \frac{1}{Z^{i\to j}} n_{q_i} \prod_{k\neq i,j} \sum_{q_k} p_{q_i q_k}^{A_{ik}} (1 - p_{q_i q_k})^{1-A_{ik}} \psi_{q_i}^{k\to i}$$

$$\Big\downarrow \quad \psi_{q_i}^{i\to j} \to 1$$

Variational mean-field

$$\psi_{q_i} = \frac{1}{Z^{i\to j}} n_{q_i} \prod_{k\neq i} \prod_{q_k} \left[p_{q_i q_k}^{A_{ik}} (1 - p_{q_i q_k})^{1-A_{ik}} \right]^{\psi_{q_i}^{k}}$$

Decelle et al 2011

BP

vs

Mean-Field

Performance measure (Original assignment known)

Overlap of reconstructed group assignment with original one

$$Q_{\mathrm{true}} = \frac{1}{N} max_{\pi} \sum_{i=1}^{N} \delta_{q_i, \pi(t_i)}$$

$$t_i = \text{original assignment}$$

$$q_i = argmax_{q_i} \psi_{q_i}$$

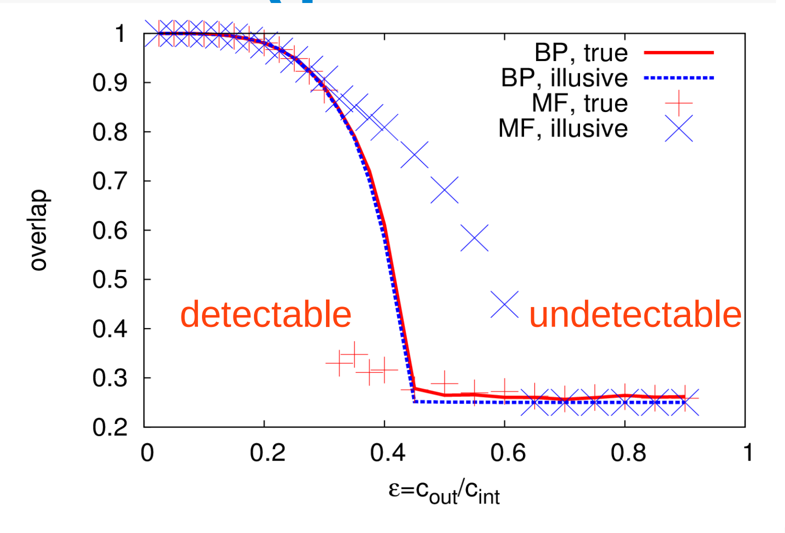
Performance measure
(Original assignment not known)

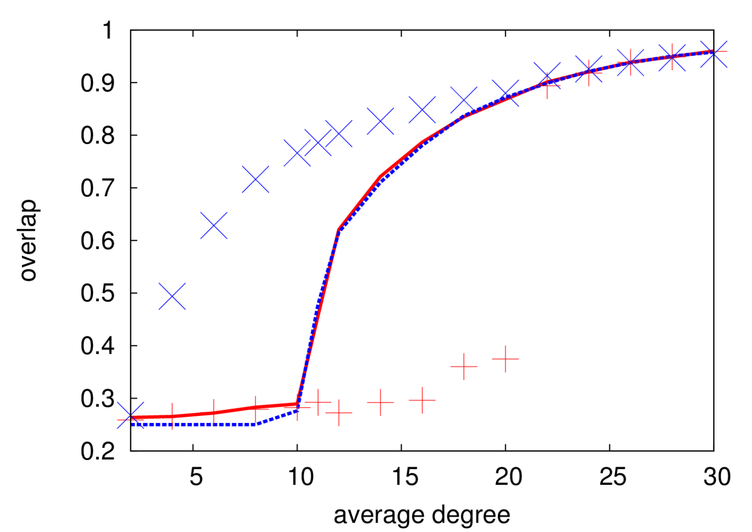
$$Q_{\text{illusive}} = \frac{1}{N} \sum_i max_{q_i} \psi_{q_i}$$

Sanity check:

$$Q_{\text{illusive}} = Q_{\text{true}}$$

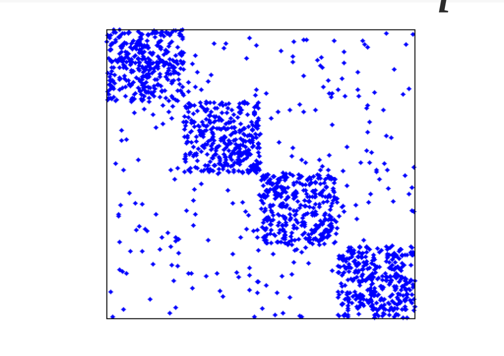
Overlap comparison (parameters known)





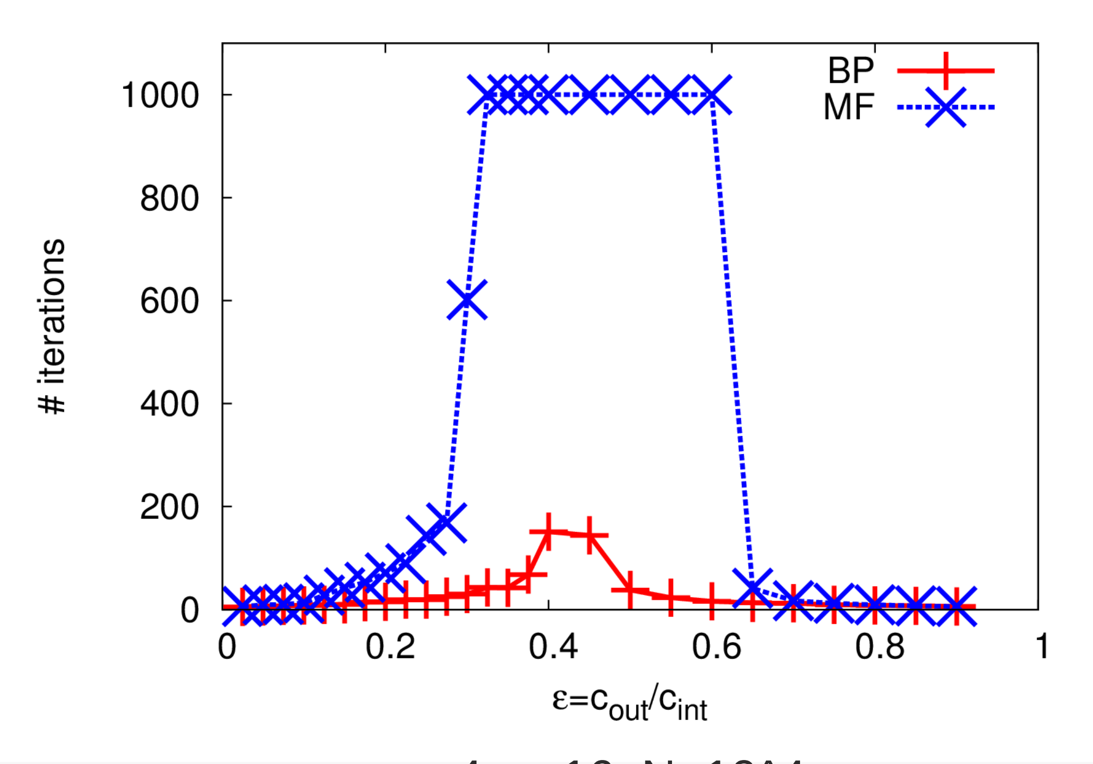
$q = 4, c = 16, N = 10^4$

$q = 4, \epsilon = 0.35, N = 10^4$



BP > MF when 0<Q<1

Convergence time comparison



q=4, c=16, N=10^4

BP converges faster than MF

Spectral Partitioning methods

Methods based on the eigenvectors of the adjacency matrix A_{ij} provide one of the most flexible approaches of graph clustering problems applied in the practice.
Advantage: Do not need any assumption of n_a, p_{ab}

Modularity SP Newman' 06
Eigenvector corresponds to largest eigenvelue of Modularity matrix

$$B_{ij} = A_{ij} - \frac{k_i k_j}{2\mathrm{M}}$$

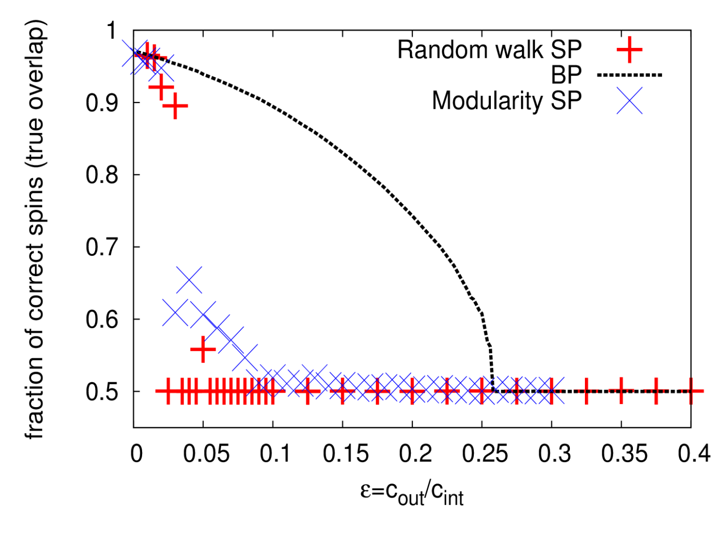
k_i: degree of i
2M: total degree

Random Walk SP Lafon' 06
Eigenvector corresponds to second largest eigenvelue of random-walk matrix

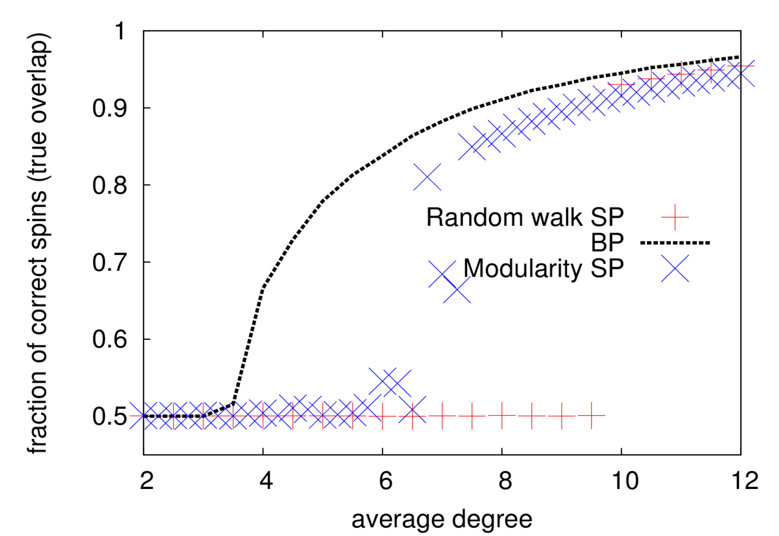
$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{A}$$

k_i: diagonal matrix with $D_{ii} = k_i$

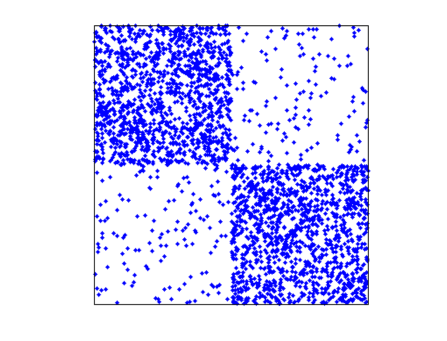
BP vs SP (parameters known for BP)



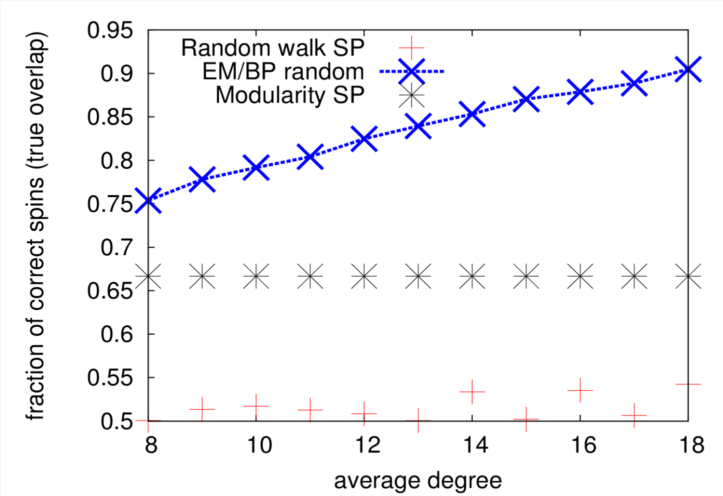
$q=2, c=3, N=10^5$

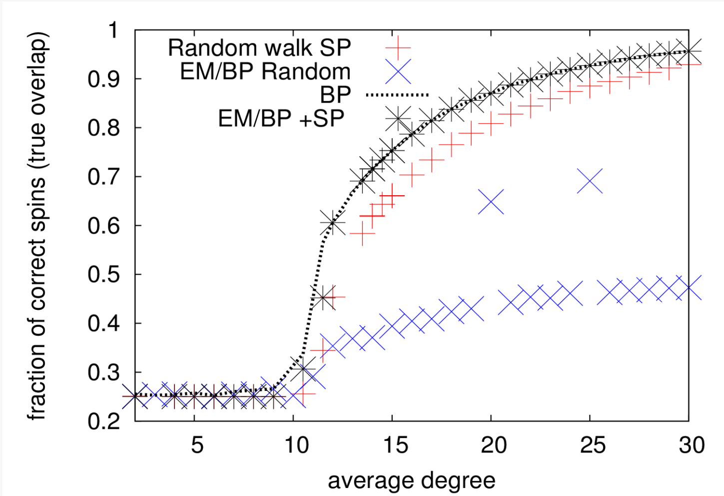


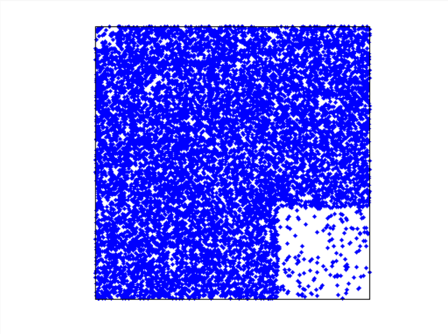
$q=2, \epsilon=0.3, N=10^5$

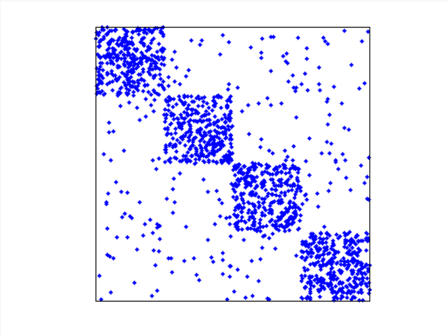


BP vs SP (parameters not known)









$p_1=2/3, p_2=1/3, N=10^5$

$q=4, \epsilon=0.35, N=10^4$

Conclusion

- BP is more accurate than variational mean-field approximation and converges faster in E-step of EM learning.

- EM/BP performs better than spectral methods in relational data with good initial parameters.

- SP can be used as a starting point to EM/BP.

Thanks!