IV. Analyse de réseaux biologiques

Catherine Matias

CNRS - Laboratoire de Probabilités et Modèles Aléatoires, Paris
catherine.matias@math.cnrs.fr
`http://cmatias.perso.math.cnrs.fr/`

ENSAE - 2014/2015

Sommaire

Outline

Some statistical networks models

- The 'most famous' ones
- Exponential random graphs
- (Overlapping) Stochastic block models
- Latent space models

Analyzing networks: (probabilistic) node clustering

Applications to biological networks

Notation

- $\mathcal{G} = (V, E)$ is a graph with nodes set $V = \{1, \dots, n\}$ and edges set E,
- For any $i, j \in V$, random variable $X_{ij} = 1\{(i, j) \in E\}$ is the edge indicator (binary graph). Sometimes, we will consider weighted graphs and then $X_{ij} \in \mathbb{R}$ is a weight on edge $(i, j) \in E$.
- $X = (X_{ij})_{1 \le i,j \le n}$ is the adjacency matrix of the graph.
- Graphs may be undirected ($X_{ij} = X_{ji}, 1 \le i \le j \le n$) or directed ($X_{ij} \neq X_{ji}, 1 \le i, j \le n$). They may admit self-loops (random variables X_{ii}) or not (then set $X_{ii} = 0$).
- For undirected graphs, $D_i = \sum_{j \neq i} X_{ij}$ is the degree of node i.
- For directed graphs, $D_{\rightarrow, i} = \sum_{j \neq i} X_{ji}$ is the incoming degree of node i (resp. $D_{\leftarrow, i} = \sum_{j \neq i} X_{ij}$ outcoming degree).

Erdős Rényi random graph

Erdős Rényi model

Undirected graph with no self-loops, where $\{X_{ij}\}_{1 \le i < j \le n}$ are i.i.d. with distribution $\mathcal{B}(p)$.

Characteristics

- Formulated by Erdős and Rényi in the late 50's,
- Huge literature, describing phase transitions behaviors as $n \to \infty$ and $p \to 0$ (existence of a giant component).
- Many links with branching processes.

R. Durrett.
Random Graph Dynamics. Cambridge University Press, 2006.

Drawbacks

- Independence and identical distribution hypothesis both are not realistic.
- The degree distribution is $\mathcal{B}in(n, p) \approx \mathcal{P}(\lambda)$ where $\lambda = np$ and thus does not follow a power law.

Degree distribution (power law, fixed degree . . .)

The power-law phenomenon (or scale free distribution)

- During the 00's, many authors focused on the degree distribution of observed networks and claimed it *always* follows a power law

  $\mathbb{P}(D_i = d) = cd^{-\alpha}$, α being the exponent of the power law.
- Some (few) nodes have a very large degree: hubs.
- They started describing networks distributions by specifying the distribution of $\{D_i\}_{i \in V}$.

Degree distribution (power law, fixed degree …)

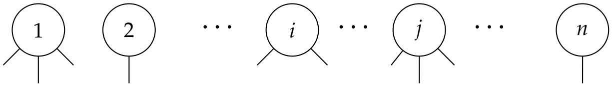
Fixed degree distribution

- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph,
- The *null* model is obtained by sampling in the set of graphs with the same degree distribution $\rightsquigarrow$ rewiring algorithm.

Degree distribution (power law, fixed degree . . .)

Fixed degree distribution

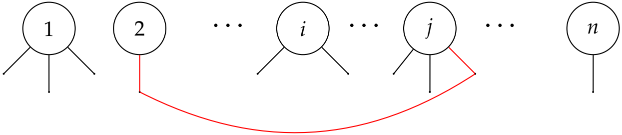
- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph,
- The *null* model is obtained by sampling in the set of graphs with the same degree distribution $\rightsquigarrow$ rewiring algorithm.



Degree distribution (power law, fixed degree …)

Fixed degree distribution

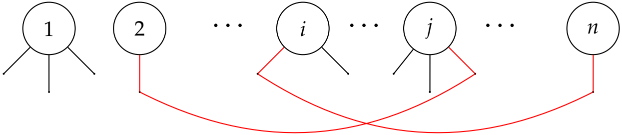
- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph,
- The *null* model is obtained by sampling in the set of graphs with the same degree distribution $\rightsquigarrow$ rewiring algorithm.



Degree distribution (power law, fixed degree ...)

Fixed degree distribution

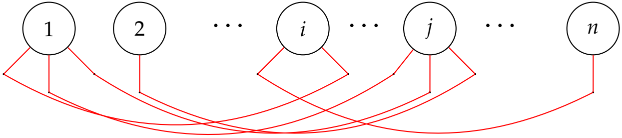
- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph,
- The *null* model is obtained by sampling in the set of graphs with the same degree distribution $\rightsquigarrow$ rewiring algorithm.



Degree distribution (power law, fixed degree ...)

Fixed degree distribution

- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph,
- The *null* model is obtained by sampling in the set of graphs with the same degree distribution $\rightsquigarrow$ rewiring algorithm.



Note that sampling in this model is expensive. Alternative?

Degree distribution (power law, fixed degree ...)

Mean degree distribution

- Let $(d_1, \ldots, d_n)$ be the degrees of an observed graph and $d_+ = \sum_i d_i$.
- Let $\{X_{ij}\}_{1 \leq i < j \leq n}$ be independent with $X_{ij} \sim \mathcal{B}(p_{ij})$ and $p_{ij} = \frac{d_i d_j}{C}$, where C is a normalizing cst s.t. $p_{ij} \in (0, 1)$. For instance $C = \max_{i \neq j} d_i d_j$.
- Contrarily to 'fixed-degree' model, we do not have $D_i = d_i$.
- Instead, $\mathbb{E}(D_i) = d_i \frac{(d_+ - d_i)}{C}$. Ideally, d_i is not too large and $C \simeq d_+$, then $\mathbb{E}(D_i) \simeq d_i$.
- If the d_i's are not too large with respect to n, then one can take $C = C_0 := \sum_i d_i \frac{(d_+ - d_i)}{d_+}$. Then, one gets exactly $\frac{1}{n} \sum_i \mathbb{E}(D_i) = \frac{d_+}{n}$.

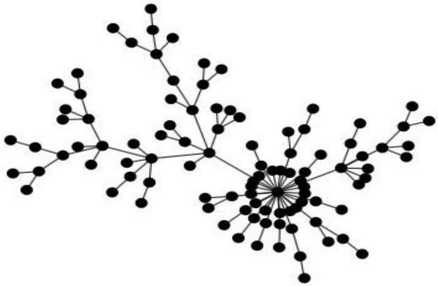
Degree distribution (power law, fixed degree ...)

Advantages and drawbacks of degree distributions

- Mean degree distribution induces independent but non i.d. edges $X_{ij} \sim \mathcal{B}(p_{ij})$. Too many parameters to be fitted to data ! Mean degree fixed them to $p_{ij} \propto d_i d_j$.
- Degree distribution alone does not capture all the information encoded in the graph.

Preferential attachment (dynamic) I

- Start with a small initial graph $\mathcal{G}_0 = (V_0, E_0)$,
- at time t, add a new node i_t. For each previous node $j \in V_0 \cup \{i_1, \ldots, i_{t-1}\}$, draw edge (i_t, j) with prob. $d_{j,t}/d_{+,t}$, where $d_{j,t}$ is the degree of j at time t.



R. Albert & A.L. Barabási.
Statistical mechanics of complex networks, Reviews of modern physics, 2002.

Preferential attachment (dynamic) II

Advantages and drawbacks

- Generative model,
- Explains the power law distribution,
- Pbm of parameter choice $(V_0, E_0, t, \ldots)$.

Probabilistic models

Here, we are going to focus on (static) 'statistical' models,

- Exponential random graph model (ERGM).
- Stochastic block model (SBM) or MixNet.
- Overlapping stochastic block models (OSBM) or mixed membership SBM.
- Latent space models.

Some recent reviews

[Matias & Robin 14] C. Matias and S. Robin.
Modeling heterogeneity in random graphs: a selective review, http://hal.archives-ouvertes.fr/hal-00948421 , 2014.

[Goldenberg *et al.* 10] A. Goldenberg, A.X. Zheng, S.E. Fienberg and E.M. Airoldi.
A Survey of Statistical Network Models, Found. Trends Mach. Learn., 2010.

Exponential random graphs I

Notation

- $X = (X_{ij})_{1 \le i,j \le n}$ the (binary) adjacency matrix,
- $S(X)$ a known vector of graph statistics on X
- θ a vector of unknown parameters

$$\mathbb{P}_\theta(X = x) = \frac{1}{c(\theta)} \exp(\theta^\intercal S(x)), \quad c(\theta) = \sum_{graphs\, y} \exp(\theta^\intercal S(y)).$$

Statistics

- $S(X)$ is a vector of sufficient statistics. It may contain number of edges, triangles, k-stars, … and also covariates.
- Note that $c(\theta)$ is not computable.
- Example: If $S(x) = (x_{ij})_{1 \le i,j \le n}$ then $\mathbb{P}_\theta(X = x) \propto \exp(\sum_{i,j} \theta_{ij} x_{ij})$, *i.e.* X_{ij} are independent non i.d. $X_{ij} \sim \mathcal{B}(p_{ij})$ with $p_{ij} = \exp(\theta_{ij})/(1 + \exp(\theta_{ij}))$.

Exponential random graphs II

More examples

- Imposing the constraint $\theta_{ij} = \theta$, one recovers Erdős Rényi model: $\mathbb{P}_\theta(X = x) \propto \exp(\theta S_1(x))$, where $S_1(x) = \sum_{i,j} x_{ij}$, the total number of edges is a sufficient stat. and $\hat{p} = \frac{S_1(X)}{n(n-1)/2}$.
- If $S(x) = (S_1(x), S_2(x))$ with $S_2(x) = \sum_{i,j,k} X_{ij}X_{ik}$ then the variables X_{ij} are non independent.
- Markov random graph: Let $S_k(x)$ be the number of k-stars and $T(x) = \sum_{i,j,k} x_{ij}x_{jk}x_{ki}$ the number of triangles. For $S = (S_1, \dots, S_{n-1}, T)$ we get $\mathbb{P}_\theta(X = x) \propto \exp(\sum_{k=1}^{n-1} \theta_k S_k(x) + \theta_n T(x))$

O. Frank & D. Strauss
Markov Graphs, JASA, 1986.

- In practice, use only $S = (S_1, \dots, S_k, T)$ for $k << n - 1$.

Exponential random graphs III

Issues on parameter estimation

- Maximum likelihood estimation is difficult
- Maximum pseudo-likelihood estimators may be used [1]. Quality of approximation ?
- MCMC approaches [Hunter *et al.* 11]: may be slow to converge.
- Very different values of θ can give rise to essentially the same distribution.
- [CD11] established a 'degeneracy' of these models, which are 'ill-posed'.

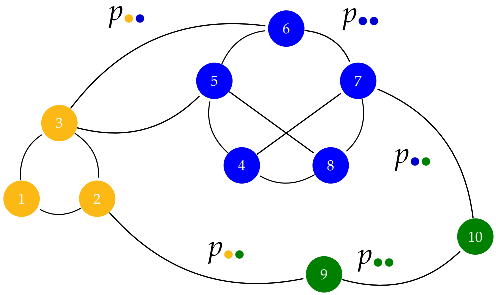
[CD11] S. Chatterjee and P. Diaconis
Estimating and understanding exponential random graph models, arXiv:1102.2650, 2011.

[Hunter *et al.* 11] D. R. Hunter, S. M. Goodreau and M. S. Handcock
ergm.userterms: A Template Package for Extending statnet.
Journal of Statistical Software, 52(2), 2013.

Stochastic block models: some motivations

- Previous models do not provide a clustering of the nodes,
- Erdős Rényi model is too homogeneous: introduce heterogeneity by using groups (cheaper than having a parameter p_{ij} for each r.v. X_{ij}).
- Groups could be put on edges, but does not take advantage of the graph structure. Rather put the groups on the nodes.

Stochastic block model (binary graphs)

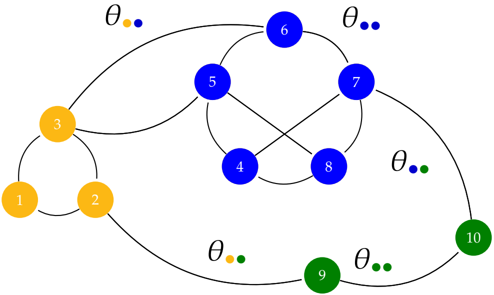


- $n = 10, Z_{5\bullet} = 1$
- $X_{12} = 1, X_{15} = 0$

Binary case

- Q groups (=colors $\bullet\bullet\bullet$).
- $\{Z_i\}_{1 \le i \le n}$ i.i.d. vectors $Z_i = (Z_{i1}, \dots, Z_{iQ}) \sim \mathcal{M}(1, \boldsymbol{\pi})$, where $\boldsymbol{\pi} = (\pi_1, \dots, \pi_Q)$ group proportions. Z_i is not observed,
- Observations: edges indicator X_{ij} , $1 \le i < j \le n$,
- Conditional on the $\{Z_i\}$'s, the random variables X_{ij} are independent $\mathcal{B}(p_{Z_i Z_j})$.

Stochastic block model (weighted graphs)



$n = 10, Z_{5\bullet} = 1$

$X_{12} \in \mathbb{R}, X_{15} = 0$

Weighted case

- Observations: weights X_{ij} , where $X_{ij} = 0$ or $X_{ij} \in \mathbb{R}^s \setminus \{0\}$,
- Conditional on the $\{Z_i\}$'s, the random variables X_{ij} are independent with distribution

$$\mu_{Z_iZ_j}(\cdot) = p_{Z_iZ_j} f(\cdot, \theta_{Z_iZ_j}) + (1 - p_{Z_iZ_j})\delta_0(\cdot)$$

(Assumption: f has continuous cdf at zero).

SBM properties

Results

- Identifiability of parameters [AMR09, AMR11].
- Parameter estimation / node clustering procedures:
  computation of the likelihood is not feasible (sum over Q^n terms),
  exact EM approach is not possible,
  instead, variational EM or variants.
  In some cases, other specific methods may be developed (ex: [AM12])
- Model selection: ICL criteria.

[AMR09] E.S. Allman, C. Matias and J.A. Rhodes.
Identifiability of parameters in latent structure models with many observed variables, Ann. Statist., 2009.

[AMR11] E.S. Allman, C. Matias and J.A. Rhodes.
Parameter identifiability in a class of random graph mixture models, JSPI, 2011.

[AM12] C. Ambroise and C. Matias.
New consistent and asymptotically normal estimators for random graph mixture models, JRSSB, 2012.

Variational EM algorithm in SBM

Let $\ell_n^c(\theta) := \log \mathbb{P}_\theta(Z_{1:n}, \{X_{ij}\}_{ij})$ be the complete log-likelihood of the model.

Why EM is not possible

- EM algorithm computes $Q(\theta, \theta') := \mathbb{E}_{\theta'}(\ell_n^c(\theta) | \{X_{ij}\}_{ij})$,
- Requires the knowledge of the distribution of $Z_{1:n}$ conditional on $\{X_{ij}\}_{ij}$
- In many setups (mixtures, HMM), this distribution factorizes: $\mathbb{P}(Z_{1:n} | \{X_{ij}\}_{ij}) = \prod_{k=1}^n \mathbb{P}(Z_k | \{X_{ij}\}_{ij})$
- This is not the case in SBM. Because of the structure of the DAG

$$\begin{tikzcd}
\cdots & & Z_i \arrow[dl] \arrow[dr] & & Z_j \arrow[dl] \arrow[dr] & & Z_k \arrow[dl] \arrow[dlllll] & \cdots \\
& X_{ik} & & X_{ij} & & X_{jk} & &
\end{tikzcd}$$

Variational EM algorithm in SBM

Principle of the variational EM

- Idea: Replace $\mathbb{P}(Z_{1:n}|\{X_{ij}\}_{ij})$ by its best approximation among the factorized distributions $q(Z_{1:n}) := \prod_{k=1}^{n} q_k(Z_k)$.
- More rigorously, for any distribution q on $\{1,\dots,Q\}^n$, let $\mathcal{L}(q,\theta) = \sum_{z_{1:n}} q(z_{1:n}) \log \frac{\mathbb{P}_\theta(z_{1:n},\{X_{ij}\}_{ij})}{q(z_{1:n})}$. Then we have

$$\log \mathbb{P}_\theta(\{X_{ij}\}_{ij}) = \mathcal{L}(q,\theta) + KL(q(\cdot)||\mathbb{P}_\theta(Z_{1:n} = \cdot|\{X_{ij}\}_{ij})) \geq \mathcal{L}(q,\theta).$$

- Minimizing KL w.r.t. $q \leftrightarrow$ Maximizing the lower bound $\mathcal{L}(q,\theta)$ w.r.t. q.

Algorithm description

- Initialize the parameter θ^0,
- Iterate:
  - E-step: θ is fixed, maximize $\mathcal{L}(q,\theta)$ w.r.t. q,
  - M-step: q is fixed, maximize $\mathcal{L}(q,\theta)$ w.r.t. θ.

Variational EM algorithm in SBM

References

[DPR08] J-J. Daudin, F. Picard and S. Robin.
A mixture model for random graphs, Statist. Comput., 2008.

[PMDCR09] F. Picard, V. Miele, J-J. Daudin, L. Cottret and S. Robin.
Deciphering the connectivity structure of biological networks using MixNet, Bioinformatics, 2009.

Variants

- Variational **Bayes**

  [LBA12] P. Latouche, E. Birmelé and C. Ambroise.
  Variational Bayesian Inference and Complexity Control for Stochastic Block Models, Statistical Modelling, 2012.

- **Online** variational EM

  H. Zanghi, C. Ambroise and V. Miele.
  Fast online graph clustering via Erdős Rényi mixture, Pattern Recognition, 2008.

Model selection criteria in SBM ([DPR08, LBA12])

- BIC can not be computed as the maximum likelihood is still unknown
- Replace the likelihood by another (close) quantity

Integrated classification likelihood (ICL)

When convergence of variational EM is attained (step K), fix $\hat{\theta} := \theta^K$ and let $\hat{Z}_i = (\hat{Z}_{i1}, \dots, \hat{Z}_{iQ}) := (q_i^K(1), \dots, q_i^K(Q))$ be the estimated posterior distribution of node i. Then define

$$ICL(Q) := \log \mathbb{P}_{\hat{\theta}}(\hat{Z}_{1:n}, \{X_{ij}\}_{ij}) - \frac{N(Q)}{2} \log n,$$

where $N(Q)$ is the number of parameters of SBM with Q groups. Then

$$\hat{Q} := \mathrm{Argmin}_Q ICL(Q).$$

C. Biernacki, G. Celeux and G. Govaert
Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000.

Other properties

- Behavior of the nodes posterior dist. / Quality of variational approx. ?
  → the groups posterior distribution converges to a Dirac mass at the true groups values
- Consistency of the MLE ?
  → the MLE of the parameter converges to the true parameter value.

A. Celisse, J.-J. Daudin and L. Pierre
Consistency of maximum-likelihood and variational estimators in the Stochastic Block Model, Elec. J. of Statistics, 2012.

Mariadassou, M. and Matias, C. Convergence of the groups posterior distribution in latent or stochastic block models, Bernoulli, to appear 2014.

Overlapping SBM / Mixed membership SBM

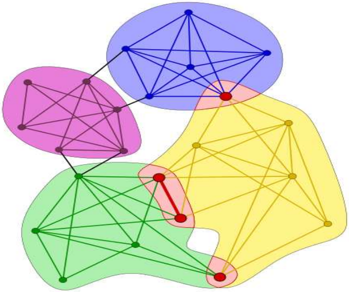


Figure : Overlapping mixture model. Source: Palla *et al.*, Nature, 2005.

Nodes may belong to many classes.

[Airoldi *et al.* 08] E.M. Airoldi, D.M. Blei, S.E. Fienberg and E.P. Xing.
Mixed Membership Stochastic Blockmodels, J. Mach. Learn. Res., 2008.

[Latouche *et al.* 11a] P. Latouche, E. Birmelé and C. Ambroise.
Overlapping Stochastic Block Models With Application to the French Political Blogosphere, Annals of Applied Statistics, 2011.

OSBM [Latouche *et al.* 11a]

Model

- $Z_i = (Z_{i1}, \dots, Z_{iQ}) \sim \prod_{q=1}^{Q} \mathcal{B}(\pi_q)$
- $X_{ij} | Z_i, Z_j \sim \mathcal{B}(g(p_{Z_i Z_j}))$ where $g(x) = (1 + e^{-x})^{-1}$ (logistic function) and

$$p_{Z_i Z_j} = Z_i^\intercal W Z_j + Z_i^\intercal U + V^\intercal Z_j + \omega$$

  W is a $Q \times Q$ real matrix while U and V are Q-dimensional real vectors and ω real number.

Results [Latouche *et al.* 11a]

- Parameter's identifiability
- Variational Bayes approach + variational logistic Bayes
- Model selection criterion

Issues

- Quality of (double) variational approximation ?

Latent space models [Handcock *et al.* 07]

Model

- Z_i i.i.d. vectors in a *latent space* $\mathbb{R}^d$.
- Conditional on $\{Z_i\}$, the $\{X_{ij}\}$ are independent Bernoulli

  $$\text{log-odds}(X_{ij} = 1 | Z_i, Z_j, U_{ij}, \theta) = \theta_0 + \theta_1^\intercal U_{ij} - \|Z_i - Z_j\|,$$

  where $\text{log-odds}(A) = \log \mathbb{P}(A)/(1 - \mathbb{P}(A))$; $\{U_{ij}\}$ set of covariate vectors and θ parameters vector.
- This may be extended to weighted networks

Latent space models [Handcock *et al.* 07]

Results [Handcock *et al.* 07]

- Two-stage maximum likelihood or MCMC procedures are used to infer the model's parameters
- Assuming Z_i sampled from mixture of multivariate normal, one may obtain a clustering of the nodes.

Issues

- No model selection procedure to infer the 'effective' dimension d of latent space and the number of groups.

[Handcock *et al.* 07] M.S. Handcock, A.E. Raftery and J.M. Tantrum
Model-based clustering for social networks, J. R. Statist. Soc. A., 2007.

Outline

Some statistical networks models

- The 'most famous' ones
- Exponential random graphs
- (Overlapping) Stochastic block models
- Latent space models

Analyzing networks: (probabilistic) node clustering

Applications to biological networks

Clustering the nodes of a network

Probabilistic approach

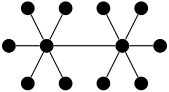
- Using either mixture or overlapping mixture models, one may recover nodes groups.
- These groups reflect a common 'connectivity behaviour'.

Non probabilistic approach = community detection

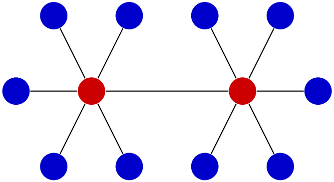
- Many clustering methods try to group the nodes that belong to the same clique.
- Here the nodes in the same groups tend to be connected with each other.

Major difference between probabilistic/non probabilistic approach

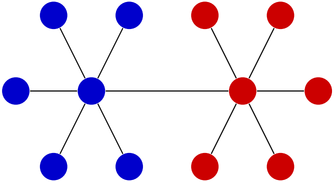
Observation of



may lead to either



MixNet model



Clustering based on cliques

Remaining challenges

- Dynamic clustering of networks

Outline

Some statistical networks models
- The 'most famous' ones
- Exponential random graphs
- (Overlapping) Stochastic block models
- Latent space models

Analyzing networks: (probabilistic) node clustering

Applications to biological networks

Transcription regulatory network (TRN) of *E. coli* [PMDCR09]

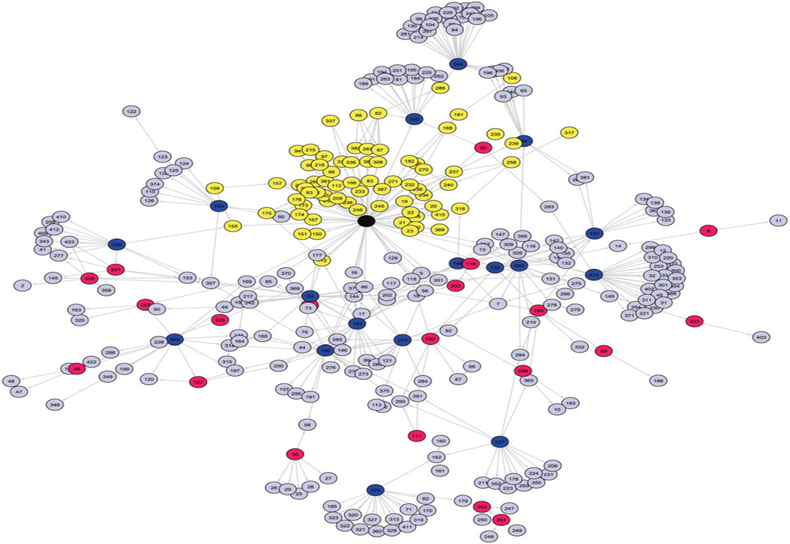
TRN description

- nodes = operon (groups of genes acting together)
- link if one operon encodes a transcription factor that directly regulates another operon

Analysis

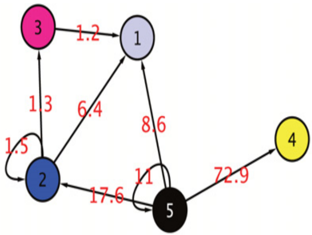
- Clustering of the graph with SBM, using 5 groups (ICL criterion)

TRN of *E. coli* [PMDCR09]



TRN of *E. coli* [PMDCR09]

Summarized through



Summary graph structure indicates that the majority of operons are regulated by very few nodes: At this resolution level, the network is summarized into regulated operons (groups 1 and 4), which receive edges only. These two groups are distinguished based on their regulatory elements: operons of group 4 are regulated by crp only (which makes its own group), whereas operons of group 1 are regulated by many cross-talking elements (group 2, 3, and 5).

TRN of *E. coli* [PMDCR09]

Estimated connectivity matrix

Table 1: Connectivity matrix for E. Coli TRN with 5 classes. The probabilities of connexion are given in percentage, and probabilities lower than 1% are not displayed.

	MixNet Classes				
	1	2	3	4	5
1	.	.	.	.	.
2	6.40	1.50	1.34	.	.
3	1.21	.	.	.	.
4	.	.	.	.	.
5	8.64	17.65	.	72.87	11.01
alpha	65.49	5.18	7.92	21.10	0.30

- empty rows : some groups are made of strictly regulated operons (nodes that receive edges only),
- small diagonal elements : no community structure.