Aalto University
School of Science
and Technology

CS-E5745
Mathematical Methods for Network Science

Mikko Kivelä

Department of Computer Science
Aalto University, School of Science
mikko.kivela@aalto.fi

February 9, 2023

ERGMs and SBMs

- Learning goals this week:
  - Learn the basics of exponential random graphs (ERGMs)
  - Learn the basics of stochastic block models (SBMs)
- Materials: Newman 15.2

Graph ensembles with given properties

- Ensembles where graphs have predetermined values for properties $x(G) = x^*$.
- **"Microcanonical"**: G in the ensemble iff $x(G) = x^*$.
  - Otherwise maximally random: $P(G) = 1/c$ if $x(G) = x^*$ and $P(G) = 0$ if $x(G) \neq x^*$
  - Difficult to deal with analytically
- **"(Macro)canonical"**: $\langle x \rangle = x^*$.
  - Otherwise maximally random: $\max_P[-\sum_G P(G) \log P(G)]$
  - Leads to "exponential random graphs" (ERGM)
  - Nice statistical properties
  - Depending on x, might be difficult or easy to deal with analytically

Exponential random graphs (ERGMs)

- Class of network models for which

$$P(G) = \frac{e^{-\sum_i x_i(G)\theta_i}}{Z(\theta)},$$

  - where each x_i is an observation (a number) that we measure from the network

Exponential random graphs (ERGMs)

+ ERGMs are in the *exponential family* of distributions:
  - Desirable statistical properties
  - Maximum entropy derivation
- The normalisation constant Z can be difficult to calculate
  - Sampling from the model can be difficult
  - Fitting the model can be even more difficult

ERGM in the literature

- You already know examples of ERGMs:
  - The (p version of) Erdős-Rényi networks
  - The “soft” configuration model
- Stochastic block models are ERGMs
- The social network analysis literature uses ERGMs extensively
  - Their models don’t usually have a solution for Z
  - Selecting wrong observables x_i leads to computational problems
  - Selecting x_i is an art form by itself

ERGMs from the maximum entropy principle

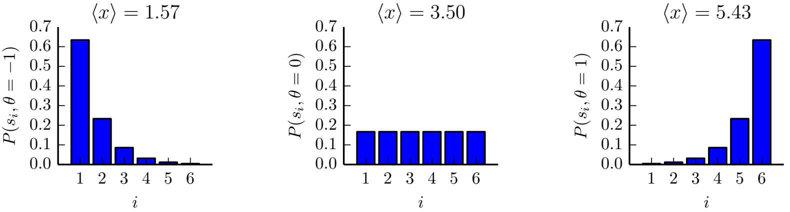
- ERGMs are probability distribution of graphs for which:
  1. The expected value of each observable gets some predetermined value $\langle x_i(G) \rangle = x_i^*$, s.t.
  2. the entropy of the distribution is maximised.

$\rightarrow$ The most random probability distribution with a specified expected value

- If we only know the expected value of the observables, ERGM gives us the "best guess" of the distribution
  - "the least biased estimate possible"
  - "maximally noncommittal with regard to missing information"
- Proof as an exercise

A simple example of exponential distributions

- States of the system: $s \in \{s_1 \dots s_6\}$
- Observable: $x(s_i) = i$
- $P(s_i|\theta) = \frac{e^{i\theta}}{\sum_{j=1}^{6} e^{j\theta}}$
- $P(s_i|0) = \frac{1}{6}$



ERGMs and statistics

- Part of the "exponential family" of distributions
  - Exponential distribution, normal distribution, ...
- x is the vector of "sufficient statistics"
- If the model is defined without fixing parameters θ and you have a single observed nework G_o
  - Choosing $\theta = \hat{\theta}$ such that $\langle x_i(G) \rangle = x_i(G_o)$ *equivalent to* finding the maximum-likelihood estimates $\hat{\theta} = \mathrm{argmax}_\theta\, P(G_o|\theta)$
  - Proof as exercise.

ERGMs and statistical physics

- The ERGMs are of the same form as canonical ensembles, the Boltzmann distribution, ...
  - Distribution of energy levels of a system (at state S and observables x_i)
- Hamiltonian: $H = \sum_i x_i(S)\theta_i$
- Partition function: $Z(\theta)$
- Free energy: $F = -\ln Z$
- Chemical potentials, inverse temperature, ...: θ_i

ERGM version of the configuration model

- Observables: degree of each node $k_i = k_i(G)$
  - In our ensemble $\langle k_i \rangle = k_i^*$ (k_i^* are the target values)
- Our Hamiltonian is:

$$H(G, \theta) = \sum_i k_i(G)\theta_i \tag{1}$$

- So the distributions is:

$$P(G|\theta) = \frac{e^{-H(G,\theta)}}{Z(\theta)} = \frac{e^{-\sum_i k_i(G)\theta_i}}{Z(\theta)} \tag{2}$$

ERGM version of the configuration model

- The Hamiltonian can be written as:

$$H(G,\theta) = \sum_i \theta_i k_i = \sum_i \theta_i \sum_j A_{ij} = \sum_{i<j} (\theta_i + \theta_j) A_{ij}\,. \qquad (3)$$

- In this case the partition function can be written without the sum over all graphs!

$$Z(\theta) = \sum_{G\in\mathcal{G}} e^{-H(G,\theta)} = \cdots = \prod_{i<j} \left(1 + e^{-(\theta_i+\theta_j)}\right). \qquad (4)$$

  - Similar derivation as an exercise.

ERGM version of the configuration model

- In total the factors can be reorganised in a way that:

$$P(G|\theta) = \prod_{i<j} p_{ij}^{A_{ij}} (1 - p_{ij})^{1-A_{ij}} \, , \tag{5}$$

where the model parameters have been transformed s.t.

$$p_{ij} = \frac{1}{1 + e^{\theta_i + \theta_j}} \, . \tag{6}$$

- When we require that $\langle k_i \rangle = k_i^*$, we need to solve θ_i from

$$k_i^* = \sum_j p_{ij} = \sum_j \frac{1}{1 + e^{\theta_i + \theta_j}} \, , \forall i \tag{7}$$

ERGM version of the configuration model

- In the "sparse limit", where $e^{\theta_i+\theta_j} \gg 1$ we can write

$$k_i^* = \sum_j \frac{1}{1+e^{\theta_i+\theta_j}} \approx \sum_j e^{-\theta_i} e^{-\theta_j} \,. \tag{8}$$

- Solution:

$$e^{-\theta_i} \approx \frac{k_i^*}{\sqrt{2m}}$$
$$p_{ij} \approx e^{-\theta_i} e^{-\theta_j} = \frac{k_i^* k_j^*}{2m}$$

- This is the "soft" configuration model from the first lecture!
- The sparse limit approximation can be written $1/p_{ij} \gg 1$

About the partition function

- In the configuration model we could write $Z(\theta)$ without the sum over all graphs
  - One can always do it IF the Hamiltonian can be written in form $H = \sum_{ij} \Theta_{ij} A_{ij}$
  - This doesn't always happen!
- It is difficult to do calculations if $Z(\theta)$ cannot be solved
  - MCMC methods for sampling and inference

Stochastic block model (SBM)

- Each node i belongs to block $b_i \in \{1, \ldots K\}$
- Links with probability depending on their blocks p_{rs} (prob. of link between block r and s)

$$P(G|b, \{p_{rs}\}) = \prod_{i<j} p_{b_i b_j}^{A_{ij}} (1 - p_{b_i b_j})^{1-A_{ij}} \,. \tag{9}$$

- p_{sr} is sometimes called the "block matrix"
  - One can think of it spanning a new more simple "block network"

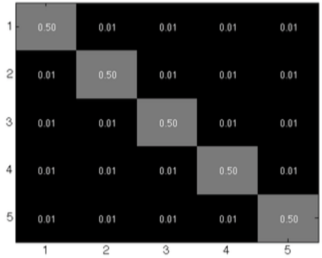
SBM as ERGM

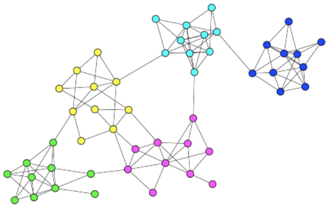
- SBM is an ERGM!
- The observations are the number of links between blocks r and s: e_{rs}
- The Z can be solved and the form in the previous slide is returned with change of variables

$$p_{rs} = \frac{1}{1 + e^{\lambda_{rs}}} \tag{10}$$

- Derivation as an exercise

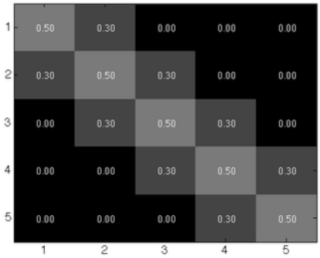
SBM examples[1] (1/4)

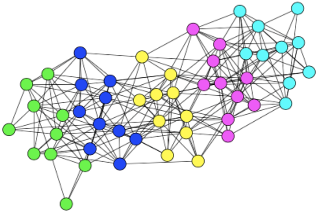




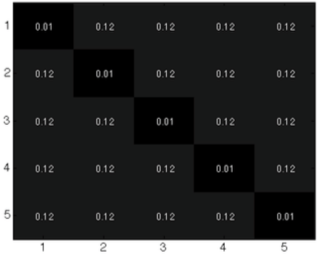
[1] http://tuvalu.santafe.edu/~aaronc/courses/5352/

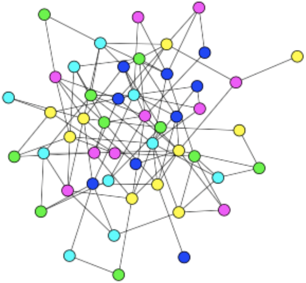
SBM examples (2/4)



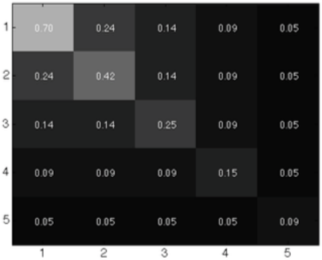


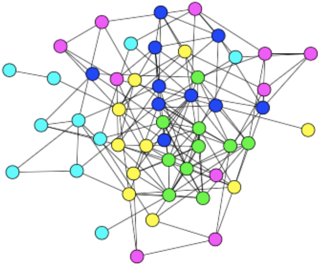
SBM examples (3/4)





SBM examples (4/4)



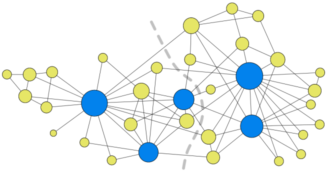


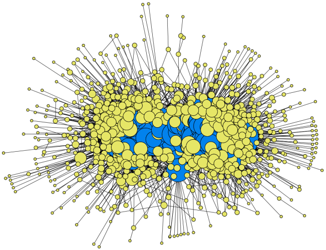
Inference with SBM

- SBM produces a network with the planted partition b_i and the block matrix p_{rs}
- Inference: we want to know the most likely model to produce the data
- Finding p_{rs} is easy given a network G and b_i (exercise)
- Finding b_i difficult $\rightarrow$ heuristic algorithms

Problem with SBM: degree distributions

- Real networks have fat-tail degree distributions, SBM finds this structure[2]





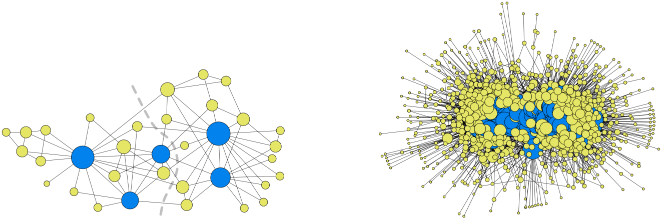
[2] Karrer & Newman, PRE 83, 016107

Degree-corrected SBM

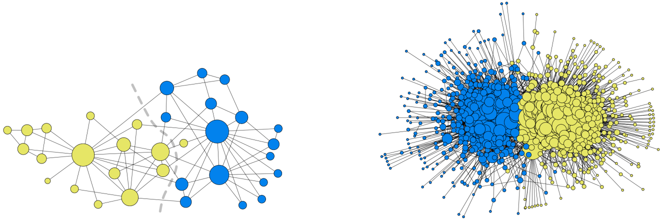
- Idea: combine the ERGM configuration model and SBM
- Observables: the degrees of nodes AND number of links between blocks
  - Model parameters related to degree θ_i and blocks λ_{rs}
  - The best fit to data explains degrees with θ and blocks with λ_{rs}

Problem with SBM: degree distributions

No degree correction:

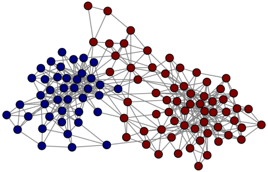


With degree correction:

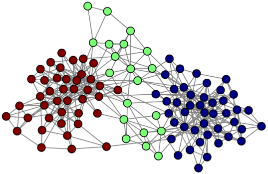


Problem with SBM: overfitting

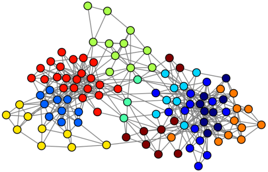
- More blocks → better likelihood[3]



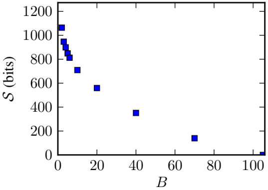
$B = 2$



$B = 3$



$B = 10$



$S = -\log P(G|\theta)$, B number of blocks

[3] Figures from Peixoto, Como'16

Minimum description length and SBM

- Instead of maximising (log) likelihood $P(G|\theta)$ maximise the posterior $P(\theta|G) = \frac{P(G|\theta)P(\theta)}{P(G)}$

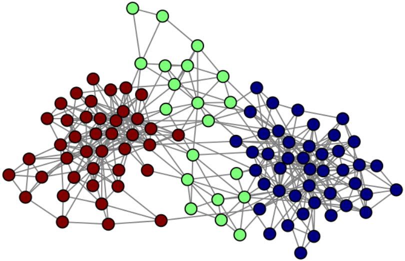
$\rightarrow$ Minimise: $-\ln P(\theta|G) = -\ln P(G|\theta) - \ln P(\theta) + \ln(P(G))$,

  - $P(G)$ is constant
  - $S = -\ln P(G|\theta)$: information needed to describe G when model known
  - $L = -\ln P(\theta)$: information needed describe to the model
  - Description length $S + L$

- Calculating L based on giving each partition b equal probability (uniform prior) etc.

Minimum description length and SBM

- MDL finds a compromise between the model fit S and complexity of the model L [4]



$$B = 3$$

[4] Peixoto, PRL 110, 148701 (2013)

ERGMs in social network analysis (SNA)

- ERGMs are a popular tool for analysing small social networks
  1. select the observables x_i (“network statistics”) based on a research question (often includes metadata on nodes),
  2. fit the model to data, and
  3. look at the θ_i to interpret the results
- The $Z(\theta)$ not solvable $\rightarrow$ numerical methods to find MLE θ
  - Find numerically θ s.t. $\langle x_i \rangle = x_i^*$, with MCMC methods
  - Selecting wrong observables x_i might lead to serious computational problems (“degeneracy”: multiple parameter combinations might explain the data)
  - Often p-values are calculated for testing a null-model where $\theta_i = 0$

Example: ERGMs in SNA

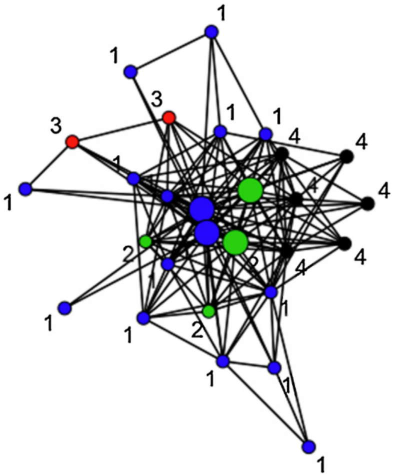
- Data: 6+6 classes, around 24 nodes per class[5]

	"3-year-olds"			"4-year-olds"		
	$\hat{\mu}_{WLS}$	SE	σ^2	$\hat{\mu}_{WLS}$	SE	σ^2
Reciprocity	4.61**	.28	.00	4.59**	.33	.26
Alternating k-instar	1.10**	.22	.02	1.12**	.20	.00
Alternating k-outstar	−1.75**	.09	2.82**	−1.12**	.48	.35
Alternating k-triangle t	.30**	.08	.02*	.21**	.04	.00
Alternating k-two-path	−.51**	.10	.03	−.60**	.07	.01
Ego sex (♂)	−.63**	.21	.00	−.54**	.21	.00
Alter sex (♂)	−.27	.16	.00	−.13	.15	.01
Sex similarity	.87**	.15	.00	1.21**	.17	.00

[5] Daniel et al., Social Net. 35(1), 25 (2013)

Example: ERGMs in SNA

- Social network of judges[6]



[6] Lazega et al., Social Net. 48, 10 (2017)

Example: ERGMs in SNA

Effects	Parameter estimate	Standard error
Variables of interest		
Judges apply the same rule	**−0.579**	**0.272**
Judges belong to same capitalism block	−0.707	0.452
Judges apply the same rule AND belong to continental Europe capitalism block	**1.242**	**0.346**
Judges apply the same rule AND belong to UK capitalism block	0.673	0.335
Judges apply the same rule AND belong to Scandinavia capitalism block	**0.951**	**0.402**
Judges apply the same rule AND belong to southern Europe capitalism block	**0.945**	**0.351**
Endogenous network controls		
Density	**−4.537**	**1.039**
Reciprocity	**1.261**	**0.394**
Indegree control 1(Markov)	**0.012**	**0.001**
Outdegree control 1(Markov)	**0.012**	**0.001**
Twopath	**−0.087**	**0.025**
Indegree control 2	−0.061	0.331
Outdegree control 2	−0.340	0.350
Transitive closure	**1.167**	**0.211**
Cyclic closure	0.029	0.120
Transitive connectivity	−0.058	0.032