

STOR601 - Research Topics

Network Modelling

Clement Lee

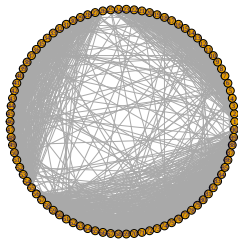
2021-01-14 (Thu)

Welcome

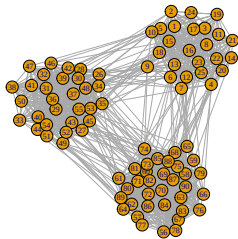
- Outline of this hour
 - Stochastic block models
 - Modelling degree distribution

Stochastic block models

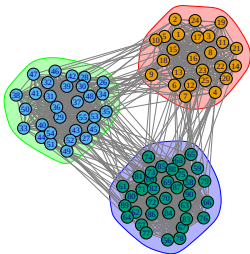
Example: 90 nodes, 1192 edges



Visually better



Colouring the groups



- Nodes are tightly connected to each other **within** each group
 - Less connected **between** groups
- More importantly, the connection pattern is **similar** for all nodes within the same group
 - Approx. same number of edges with nodes in the same group
 - Approx. same number of edges with nodes in another group
- I simulated the data & made the communities visible
 - Less clear-cut for real-life data

- Given: a network
 - Undirected or directed
 - There could be node/edge attributes
 - **Not** considered as classified labels though
- To detect:
 - Groups in which nodes are tightly connected (community detection); **and/or**
 - Groups in which the connectivity pattern is similar
- Node clustering / unsupervised learning

Community detection **algorithms**

- Related to stochastic block models
- Heuristic algorithms
- Quick, based on simple rules
- Comprehensive review by **Fortunato (2010, Physics Reports)**
- Let's look at an example

Label propagation algorithm

- Raghavan, Albert and Kumara (2007, *Physical Review E*)
- Initially each node is randomly assigned to one of the K groups
- Each node takes turn to join the group to which the maximum number of neighbours belong
- The iterative process continues until no node changes its group membership anymore

Issues of community detection algorithms

- Different initial configurations may lead to different results
- Different algorithms give different results
 - Which one is better / best?
- Pre-specifying the number of groups beforehand

Stochastic block models (SBMs)

- Proper statistical models
 - Theoretically one set of results, once the model is specified
- Idea: To put nodes that have **similar** properties together
 - The concept of **stochastic equivalence**

Stochastic equivalence

- Given nodes p and q in the same group
- $\Pr(\text{edge from } p \text{ to } r) = \Pr(\text{edge from } q \text{ to } r)$
 - No matter r is in the same group with p & q or not
 - The two probabilities are independent, **given the group memberships**
- This is the one I used for the example

	Group 1	Group 2	Group 3
Group 1	0.8	0.05	0.05
Group 2	0.05	0.8	0.05
Group 3	0.05	0.05	0.8

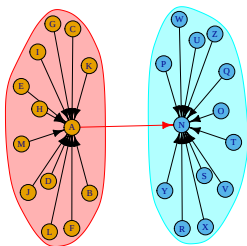
SBMs do not necessarily achieve community detection

- The model equation says nothing about the probabilities' magnitude
- High number of connections between groups but not within groups

	Group 1	Group 2	Group 3
Group 1	0.03	0.06	0.80
Group 2	0.90	0.05	0.01
Group 3	0.02	0.70	0.04

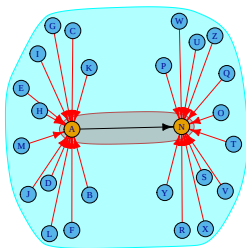
- SBMs *usually* achieve community detection
 - Able to cluster tightly connected nodes
- This means the following are assumed:
 - Stochastic equivalence for tightly connected nodes
 - Degrees of nodes within each group follow the same distribution
 - Quite strong assumptions
- Not quite realistic for actual data
 - Degree heterogeneity
 - Lack of fit

Illustration



- Community detection achieved but poor fit

Illustration (cont'd)



- Nodes stochastic equivalent but not tightly connected

- Karrer and Newman (2011, *Physical Review E*)
- An extra parameter / latent variable for each node
 - To correct for the heterogeneity in their degrees
 - Nodes within the same group “stochastically equivalent” after correcting
- Community detection + improved fit

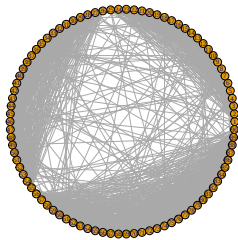
Not the universal fix though

- In some cases, selecting between the original and degree-corrected versions is not clear cut
- Each version reveals a different underlying structure about the same network
- Instead of choosing between the two versions, they could be combined in a weighted / mixture model

- What data do we have?
 - The network, represented by e.g. the adjacency matrix
- What quantities do we have to infer?
 - The group memberships of the nodes
 - The block matrix

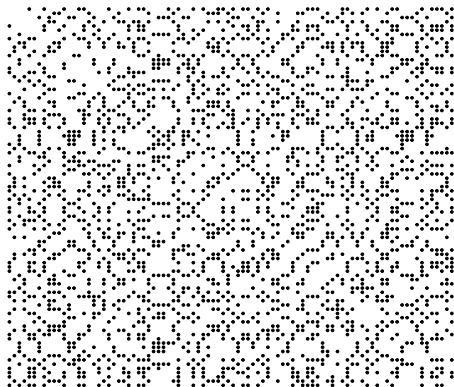
	Group 1	Group 2	Group 3
Group 1	??	??	??
Group 2	??	??	??
Group 3	??	??	??

Using the first example



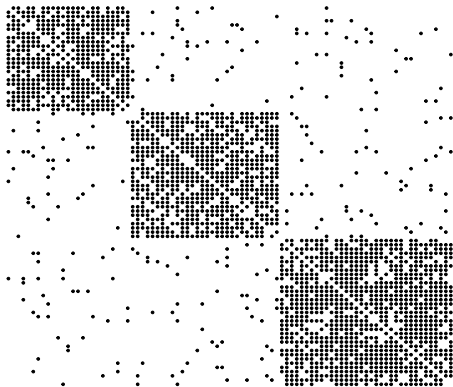
Rearranging the adjacency matrix

- From ...



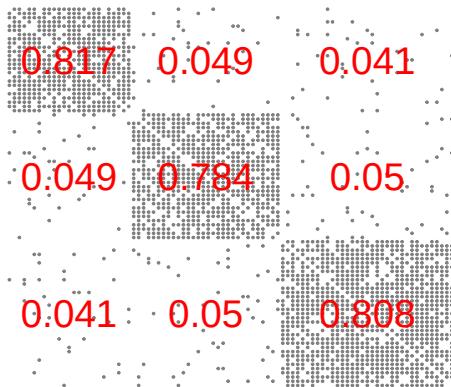
Inference achieved

- ... to ...



Adjacency matrix $\rightarrow$ block matrix

- Densities in each block = point estimates of the parameters



- Computational cost of naive inference algorithm usually $O(n^2)$
 - n = number of nodes
- Number of groups K
 - Infer it as a parameter?
 - Select it using criteria?
- Label switching issue
 - Common in clustering
- Bayesian approach by **Peixoto (2020, *ArXiv:1705.10225*)**
 - Biggest development since the degree-corrected SBM
 - Seems to solve most modelling & inference issues

- Assortative SBMs
- Mixed membership SBMs
- Overlapping SBMs
- Models for temporal / dynamic networks
- With textual information

- We have seen community detection *algorithms*
- Modifying SBMs specifically for community detection purposes
 - Large probabilities within groups (along the major diagonal)
 - Small probabilities between groups (off-diagonal)
 - Assortativeness
- Lu and Szymanski (2019, *Scientific Reports*) models the assortativeness

Mixed membership / overlapping SBMs

- Hard clustering so far
 - Each node belongs to one group only
- Mixed membership models
 - A membership **vector** for each node e.g. (0.2, 0.5, 0.3)
 - Hard clustering a special case: a K -vector with one 1 & the rest 0's
 - [Airoldi et al. \(2008, JMLR\)](#)
- Overlapping models
 - A node can belong to one or more groups at the same time
 - $2^K - 1$ choices: $\{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}$
 - [Ranciati et al. \(2020, AoAS\)](#)

+ Temporal / textual information

- Instances of the network at different times
- Textual information associated with:
 - The edges: email exchanges network
 - The nodes: citation network
- Incorporate topic and/or longitudinal modelling
- Corneli et al. (2018, *Statistics and Computing*)

- Community detection algorithms - we have seen them
- Latent feature models
- Latent space models
- Hypergraph models

- K features instead of K groups
- Having none, one or multiple features
 - 2^K choices (compared to $2^K - 1$ for overlapping SBMs)
- Miller et al. (2009, *NIPS*)

Latent space models

- Instead of being put into groups
- Projected onto an Euclidean space
- Being connected to SBMs by e.g.

Sanna Passino and Heard (2020, *Statistics and Computing*)

- Consider coauthorship networks
 - Each node represents an author
 - Traditionally: edge between two authors if they have coauthored one or more articles
- How to interpret pairwise edges between nodes 1, 2 and 3?
 - Each pair have collaborated separately?
 - All of them have written one or more articles as a whole?
 - A combination of both?
- There is information lost using SBM

Hypergraph models (cont'd)

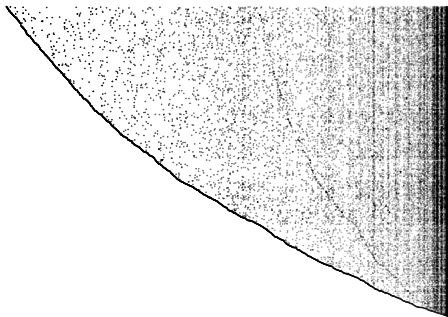
- In a hypergraph:
 - Each pair have collaborated separately: 3 hyperedges, namely $\{1, 2\}$, $\{2, 3\}$, $\{3, 1\}$
 - All of them have written one or more articles as a whole: $\{1, 2, 3\}$
- Lunagomez et al. (2017, *JASA*)

Transition: my work

- Lee and Wilkinson (2019, *Applied Network Science*)
 - Review of stochastic block models
 - All references before (& more)
- Networks that are **directed acyclic graphs** (DAGs)
 - Citation network of all the papers in a conference
 - **Dependency network of R packages**

Modifying SBM for DAGs

- Adjacency matrix can be upper triangular upon rearranging



Investigating the degree heterogeneity

- Instead of considering it in the context of SBMs
- How about we look at the degree distribution directly?

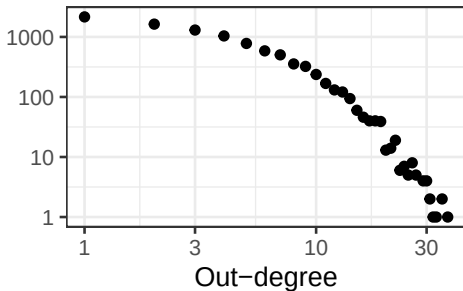
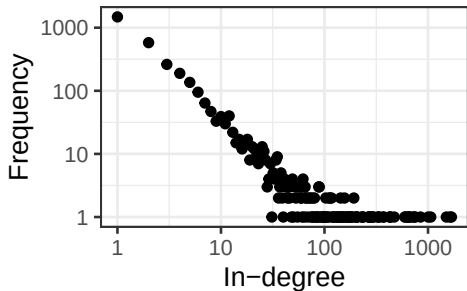
Modelling degree distribution

What does degree heterogeneity mean?

- Nodes not following the same degree distribution
- There may be underlying reasons for such heterogeneity
 - (Unknown) node covariates
 - But they are not the focus here
- Simply fit a univariate distribution to each kind of degrees:
 - In-degree
 - Out-degree
 - Total degree

Degrees of R packages

- In-degree seems to follow the **power law**



The power law

- In probability terms:

$$p(x) \propto x^{-\alpha}, \quad x > u$$
$$\log p(x) = -\alpha \log x + c$$

- Straight line on the log-log scale
 - $\alpha = -1 \times \text{slope}$
- Related concepts:
 - “The rich gets richer”
 - Preferential attachment
 - 80-20 rule / Pareto principle
 - Scale-free networks

- Barabasi and Albert (1999, *Science*)
- Growing the network by simple rules
 - A new node comes in and connect to existing nodes
 - Probability of connecting to the new node $\propto$ current (in-)degree
- Degree distribution follows the power law with $2 < \alpha < 3$
 - Clauset et al. (2009, *SIAM Review*)
- These two papers are themselves embodiment of preferential attachment

Criticisms of the power law

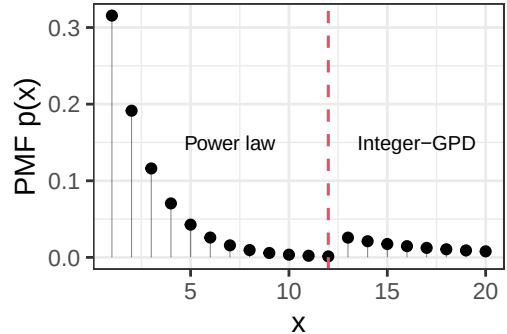
- Not adequate to describe the degree distribution
 - Usually poorer fit in the right tail, where it is more of interest
- Alternative distributions e.g. log-normal better / sufficient
 - Stumpf and Porter (2012, *Science*)
 - Broido and Clauset (2019, *Nature Communications*)

Connecting to extremes

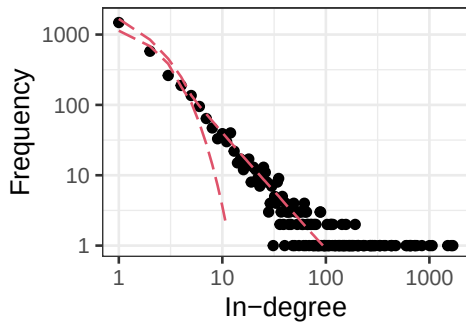
- Continuous power law = Pareto distribution
- Pareto distribution $\rightarrow$ Generalised Pareto distribution (GPD)
 - Popular in extreme value theory
 - Can model the tail better / flexibly
- GPD $\rightarrow$ Integer-GPD
 - Discretised
 - **Rohrbeck et al. (2018, AoAS)**
- Why don't we use this then?

Sidetrack: extreme value theory

- Mixture distribution
- Power law for the left “tail”
 - $\text{data} \leq u$
- GPD for the right tail
 - $\text{data} > u$
 - More flexible
- u is also inferred



Good fit



Next steps

- Can we modify the preferential attachment model to lead to such in-degree distribution?
- How about modifying an SBM? Or a combination of both?
- Work in progress: [Lee and Eastoe \(2020, *ArXiv:2008.03073*\)](#)

Thank you

- Questions?