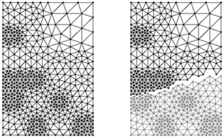
network *blockmodeling*

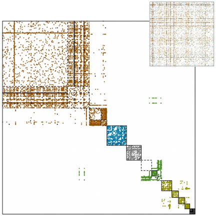
introduction to *network science in Python* (*NetPy*)

Lovro Šubelj
University of Ljubljana
3rd October 2024

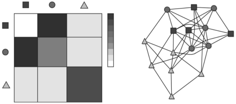
blockmodeling *overview*



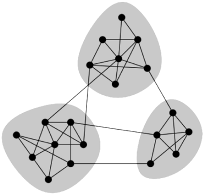
graph partitioning [KL70, Fie73]



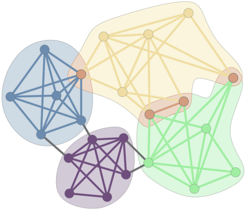
blockmodeling [LW71, WR83]



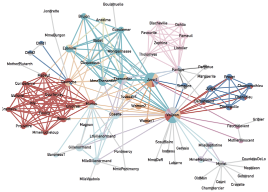
stochastic block models [Pei15]



communities [GN02]



overlapping communities [PDFV05]



link communities [EL09, ABL10]

* assortative & disassortative equivalence blockmodeling

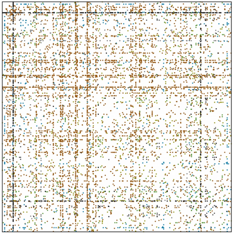
blockmodeling *equivalence*

— *standard equivalence blockmodeling* [DBF05]

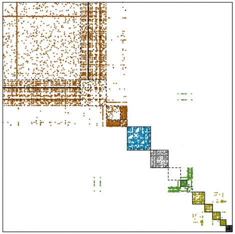
- define *node similarity* as *(structural) equivalence*

$$\sigma_{ij} \sim |\Gamma_i \cap \Gamma_j|$$

1. *blockmodeling* by *(hierarchical) clustering* $\mathcal{O}(n^2)$
2. return *block model* at desired *clustering resolution*



javax adjacency matrix



javax block model

‡ `javax.swing`, `javax.management`, `javax.naming`, `javax.print`, `javax.xml`, `javax.lang` etc.

blockmodeling *structural*

similar nodes have *same neighbors*

— *standard structural equivalence* [LW71] of i and j is

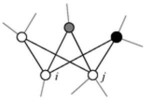
$$\sigma_{ij} = \sum_x A_{ix} A_{xj} = |\Gamma_i \cap \Gamma_j|$$

— *Salton structural equivalence* [SM83] of i and j is

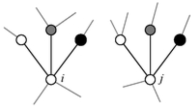
- θ_{ij} is *angle* between neighborhoods A_i and A_j

$$\sigma_{ij} = \cos\theta_{ij} = \frac{\sum_x A_{ix} A_{xj}}{\sqrt{\sum_x A_{ix}^2}\sqrt{\sum_x A_{jx}^2}} = \frac{|\Gamma_i \cap \Gamma_j|}{\sqrt{k_i k_j}}$$

— *Leicht structural equivalence* [LHN06] of i and j is $\sigma_{ij} = \frac{|\Gamma_i \cap \Gamma_j|}{k_i k_j / n}$



structural



regular equivalence

stochastic models

introduction to *network science in Python* (*NetPy*)

Lovro Šubelj
University of Ljubljana
3rd October 2024

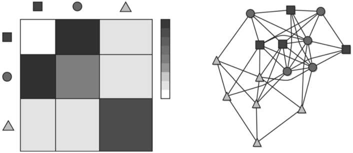
stochastic *models*

— *random graph model* $G(n, m)$ for *network links* m [ER59]

— *configuration model* $G(\{k\})$ for *node degrees* $\{k\}$ [NSW01]

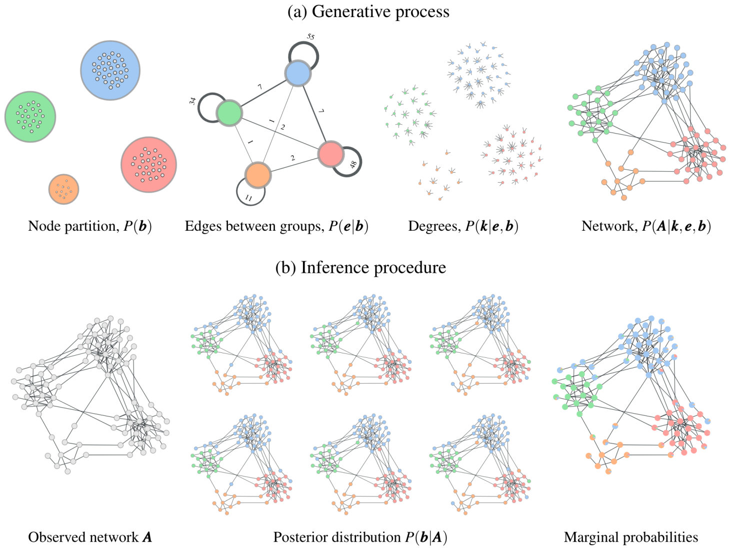
— *exponential* p^**-model* $G(n, \{\langle x\rangle\})$ for *any expectations* $\{\langle x\rangle\}$

— *stochastic block model* $G(\{C\})$ for *node clusters* $\{C\}$ [HLL83]



*assortative & disassortative stochastic block models

stochastic *process*



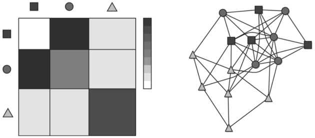
stochastic *SBM*

- $G(\{C\},\{p\})$ *stochastic block model* [HLL83]
- *link* between i and j *placed with probability* $p_{c_ic_j}$
  - $m_{c_ic_j}$ is *number of links* between C_i and C_j
  - $M_{c_ic_j}$ is *maximum* $m_{c_ic_j}$ hence n_in_j or $\binom{n_i}{2}$

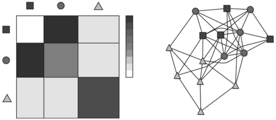
$$\mathrm{P}(G|\{C\},\{p\}) = \prod_{i\leq j} p_{c_ic_j}^{A_{ij}}(1-p_{c_ic_j})^{1-A_{ij}} = \prod_{c_i\leq c_j} p_{c_ic_j}^{m_{c_ic_j}}(1-p_{c_ic_j})^{M_{c_ic_j}-m_{c_ic_j}}$$

- *maximum likelihood* $G(\{C\})$ *block model*
  - $\frac{m_{c_ic_j}}{M_{c_ic_j}}$ is *maximum likelihood estimate* for $p_{c_ic_j}$

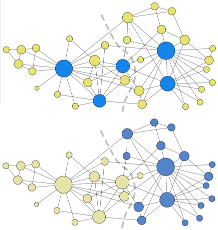
$$\mathcal{L}(G|\{C\}) = \log \mathrm{P}(G|\{C\}) = \sum_{c_i\leq c_j} m_{c_ic_j}\log\frac{m_{c_ic_j}}{M_{c_ic_j}-m_{c_ic_j}} + M_{c_ic_j}\log\frac{M_{c_ic_j}-m_{c_ic_j}}{M_{c_ic_j}}$$



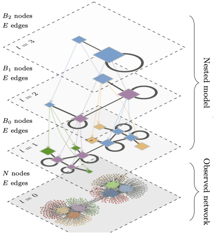
stochastic *overview*



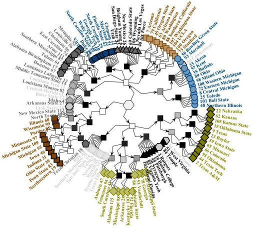
stochastic block model [HLL83]



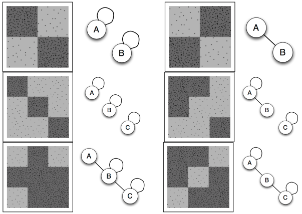
degree-corrected SBM [KN11]



nested SBM [Pei15]



hierarchical models [CMN08, ŠB14]



role models [RW07, NL07, GSPA07]

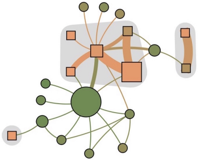


image graphs [ŠB12]

† overlapping & corrected models also known as mixture & mixed membership models

blockmodeling *references*

Yong-Yeol Ahn, James P. Bagrow, and Sune Lehmann.
Link communities reveal multiscale complexity in networks.
Nature, 466(7307):761–764, 2010.

Aaron Clauset, Cristopher Moore, and M. E. J. Newman.
Hierarchical structure and the prediction of missing links in networks.
Nature, 453(7191):98–101, 2008.

Patrick Doreian, Vladimir Batagelj, and Anuska Ferligoj.
Generalized Blockmodeling.
Cambridge University Press, Cambridge, 2005.

T. S. Evans and R. Lambiotte.
Line graphs, link partitions and overlapping communities.
Phys. Rev. E, 80(1):016105, 2009.

P. Erdős and A. Rényi.
On random graphs I.
Publ. Math. Debrecen, 6:290–297, 1959.

M. Fiedler.
Algebraic connectivity of graphs.
Czech. Math. J., 23:298–305, 1973.

M. Girvan and M. E. J Newman.
Community structure in social and biological networks.
P. Natl. Acad. Sci. USA, 99(12):7821–7826, 2002.

Roger Guimerà, Marta Sales-Pardo, and Luis A. N. Amaral.
Classes of complex networks defined by role-to-role connectivity profiles.
Nat. Phys., 3(1):63–69, 2007.

blockmodeling *references*

Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt.
Stochastic blockmodels: First steps.
Soc. Networks, 5(2):109–137, 1983.

Brian W. Kernighan and S. Lin.
An efficient heuristic procedure for partitioning graphs.
Bell Sys. Tech. J., 49(2):291–308, 1970.

Brian Karrer and M. E. J Newman.
Stochastic blockmodels and community structure in networks.
Phys. Rev. E, 83(1):016107, 2011.

E. A. Leicht, Petter Holme, and M. E. J. Newman.
Vertex similarity in networks.
Phys. Rev. E, 73(2):026120, 2006.

F. Lorrain and H. C. White.
Structural equivalence of individuals in social networks.
J. Math. Sociol., 1(1):49–80, 1971.

M. E. J Newman and E. A Leicht.
Mixture models and exploratory analysis in networks.
P. Natl. Acad. Sci. USA, 104(23):9564–9569, 2007.

M. E. J. Newman, S. H. Strogatz, and D. J. Watts.
Random graphs with arbitrary degree distributions and their applications.
Phys. Rev. E, 64(2):026118, 2001.

Gergely Palla, Imre Derényi, Illes Farkas, and Tamas Vicsek.
Uncovering the overlapping community structure of complex networks in nature and society.
Nature, 435(7043):814–818, 2005.

blockmodeling *references*

Tiago P. Peixoto.
Model selection and hypothesis testing for large-scale network models with overlapping groups.
Phys. Rev. X, 5(1):011033, 2015.

J. Reichardt and D. R. White.
Role models for complex networks.
Eur. Phys. J. B, 60(2):217–224, 2007.

Lovro Šubelj and Marko Bajec.
Ubiquitousness of link-density and link-pattern communities in real-world networks.
Eur. Phys. J. B, 85(1):32, 2012.

Lovro Šubelj and Marko Bajec.
Group detection in complex networks: An algorithm and comparison of the state of the art.
Physica A, 397:144–156, 2014.

G. Salton and M. J. McGill.
Introduction to Modern Information Retrieval.
McGraw-Hill, 1983.

D. R. White and K. P. Reitz.
Graph and semigroup homomorphisms on networks of relations.
Soc. Networks, 5(2):193–234, 1983.