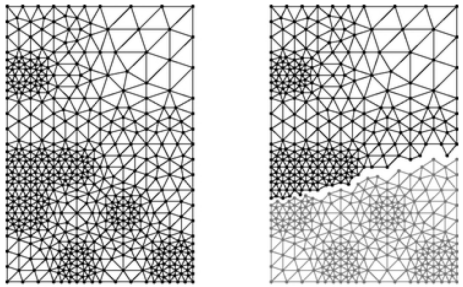
network *clustering*

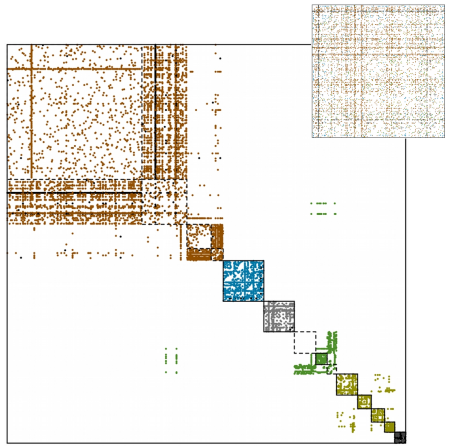
introduction to *network analysis in Python* (*NetPy*)

Lovro Šubelj
University of Ljubljana
10th Dec 2019

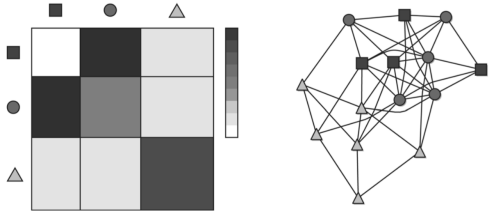
clustering *overview*



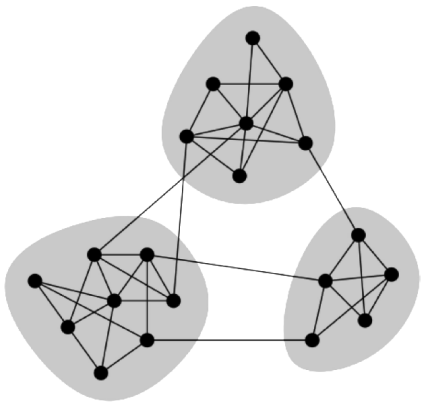
graph partitioning [KL70, Fie73]



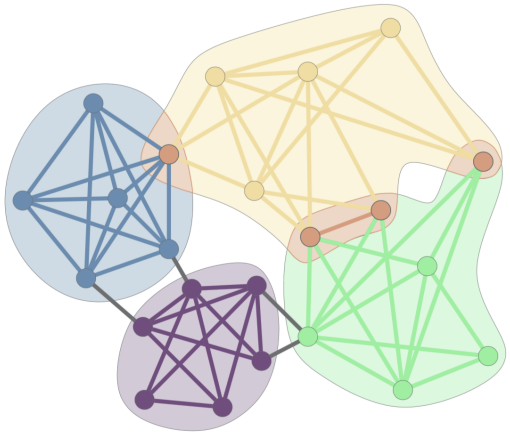
blockmodeling [LW71, WR83]



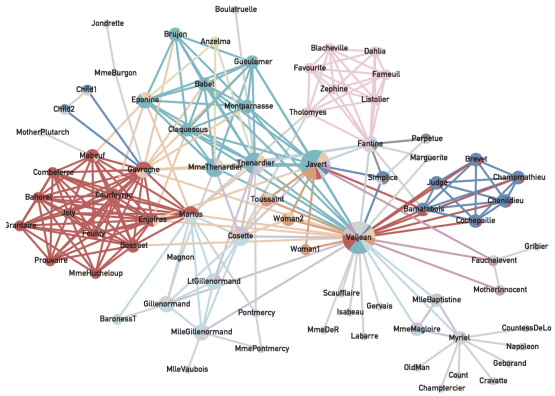
stochastic block models [Pei15]



communities [GN02]



overlapping communities [PDFV05]



link communities [EL09, ABL10]

community detection

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

Lovro Šubelj
University of Ljubljana
10th Dec 2019

community *agglomerative*

— Ravasz *hierarchical clustering* [RSM+02]

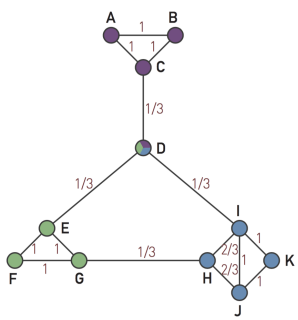
- define *node similarity* as *topological overlap*

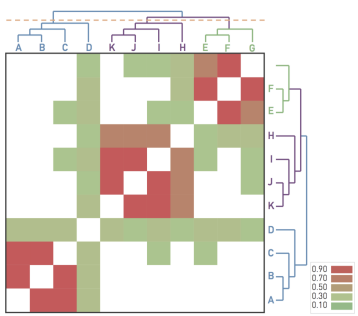
$$s_{ij} = \frac{|\Gamma_i \cap \Gamma_j| + A_{ij}}{\min(k_i, k_j)}$$

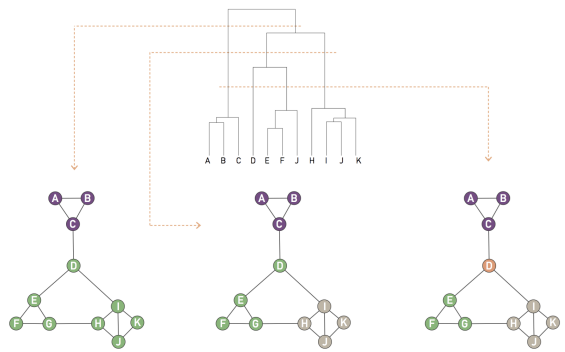
- define *cluster similarity* as *average linkage*

$$S_{ij} = \frac{1}{n_i n_j} \sum_{xy} s_{xy} \delta_{c_x c_i} \delta_{c_y c_j}$$

1. bottom-up *agglomerative hierarchical clustering* $\mathcal{O}(n^2)$
2. cut *cluster dendrogram* at desired *clustering resolution*







community *divisive*

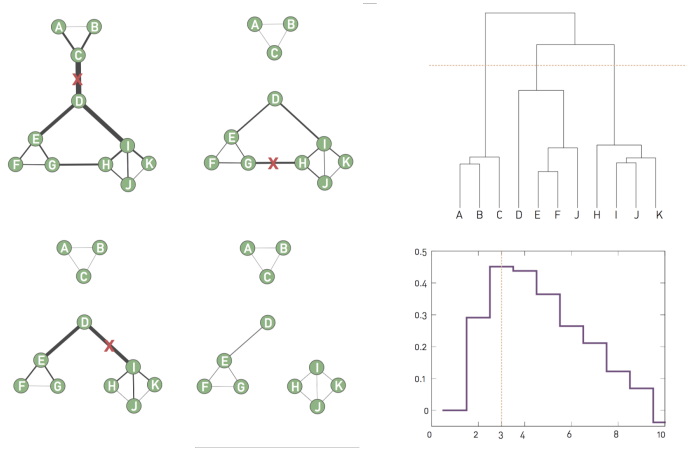
— Girvan-Newman *hierarchical clustering* [GN02]

– define *node dissimilarity* as *link betweenness*

$$\sigma_{ij} = \sum_{st \notin \{i,j\}} \frac{g_{st}^{ij}}{g_{st}}$$

1. top-down *divisive hierarchical clustering* $\mathcal{O}(nm^2)$
2. cut *cluster dendrogram* at *maximum modularity*

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta_{c_i c_j}$$

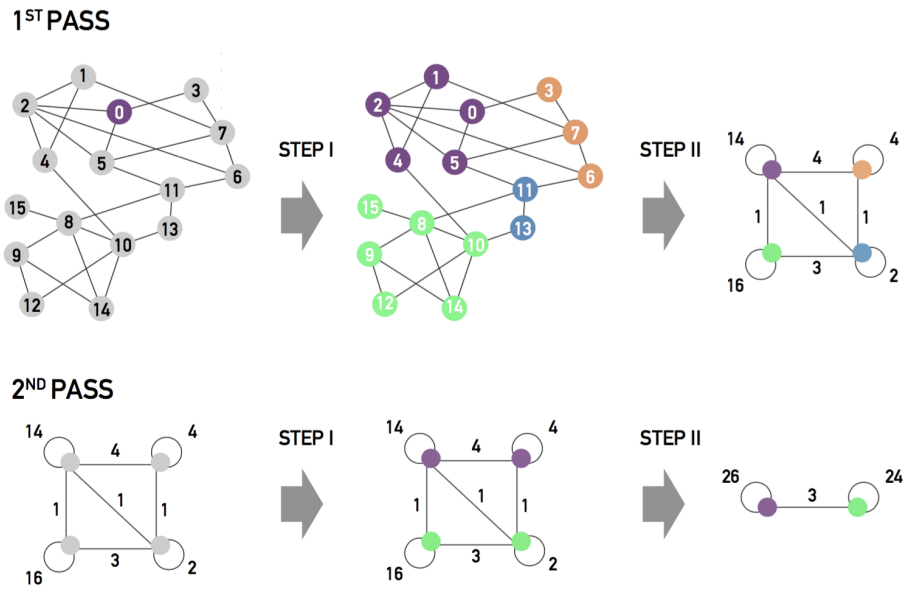


community *modularity*

— Louvain *modularity optimization* [BGLL08]

1. set *node community* by *modularity optimization* $\mathcal{O}(cm)$
2. *aggregate community nodes into supernodes* and repeat 1.
3. return *community structure maximizing modularity*

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta_{c_i c_j}$$



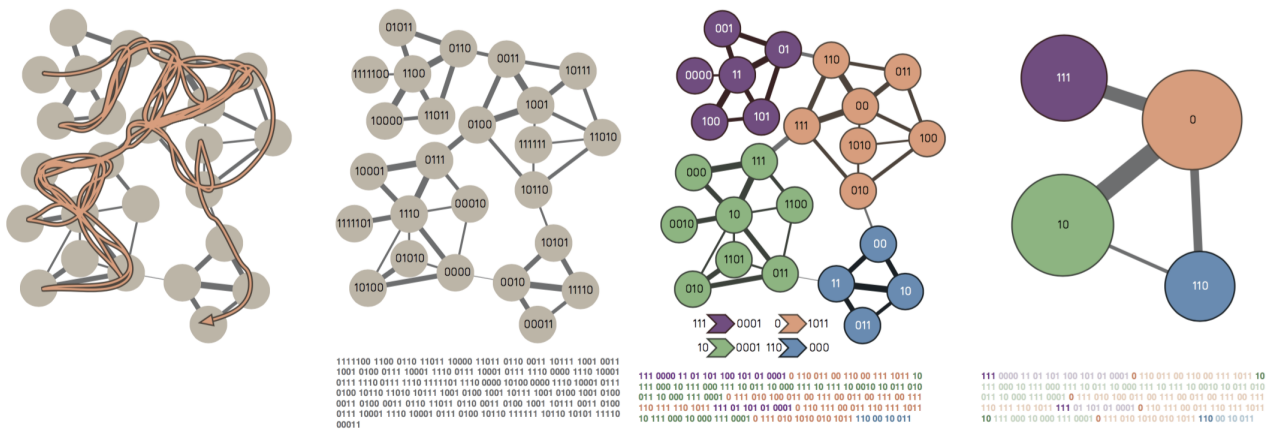
see findcommunities implementation

community *map equation*

— Infomap *map equation compression* [RB08]

1. set *node community* by *optimal coding* $\mathcal{O}(m \log m)$
2. *compress community nodes into supernodes* and repeat 1.
3. return *community structure maximizing map equation*

$$\mathcal{L} = \sum_i p_{i\rightsquigarrow} H(\widetilde{\mathcal{C}}) + \sum_i p_{i\hookleftarrow} H(\mathcal{C}_i)$$



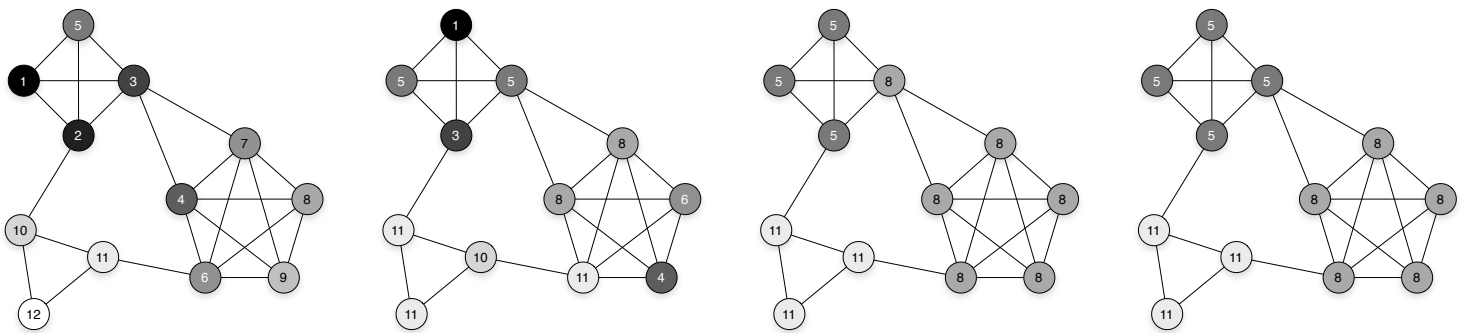
see mapequation implementation

community *propagation*

— Raghavan *label propagation* [RAK07, ŠB11]

1. set *node community* by *neighbors frequency* $\mathcal{O}(cm)$
2. *randomly shuffle nodes* and repeat 1. *until convergence*
3. return *community structure connected components*

$$\forall i : c_i = \arg\max_c \sum_j A_{ij}\delta_{c_j c}$$



see balanced implementation

clustering *references*

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

A.-L. Barabási.
Network Science.
Cambridge University Press, Cambridge, 2016.

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre.
Fast unfolding of communities in large networks.
J. Stat. Mech., P10008, 2008.

David Easley and Jon Kleinberg.
Networks, Crowds, and Markets: Reasoning About a Highly Connected World.
Cambridge University Press, Cambridge, 2010.

Ernesto Estrada and Philip A. Knight.
A First Course in Network Theory.
Oxford University Press, 2015.

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

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.

clustering *references*

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

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

Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.

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.

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

Usha Nandini Raghavan, Reka Albert, and Soundar Kumara.
Near linear time algorithm to detect community structures in large-scale networks.
Phys. Rev. E, 76(3):036106, 2007.

M. Rosvall and C. T. Bergstrom.
Maps of random walks on complex networks reveal community structure.
P. Natl. Acad. Sci. USA, 105(4):1118–1123, 2008.

E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and Albert László Barabási.
Hierarchical organization of modularity in metabolic networks.
Science, 297(5586):1551–1555, 2002.

clustering *references*

Lovro Šubelj and Marko Bajec.

Robust network community detection using balanced propagation.
Eur. Phys. J. B, 81(3):353–362, 2011.

D. R. White and K. P. Reitz.

Graph and semigroup homomorphisms on networks of relations.
Soc. Networks, 5(2):193–234, 1983.