

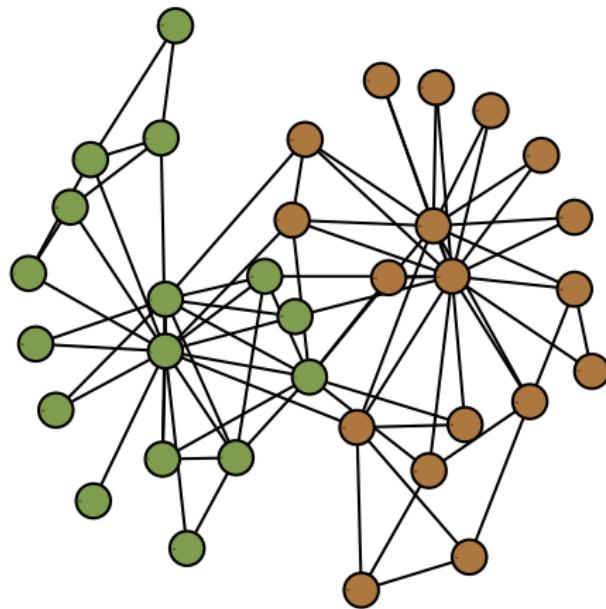
community structure

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

Lovro Šubelj
University of Ljubljana
26th September 2024

community *structure*

karate club *network split* [Zac77]

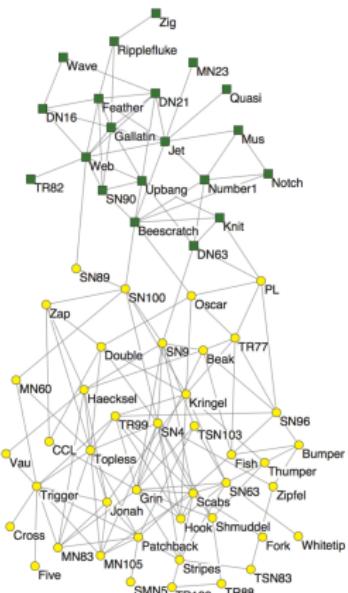


community *detection*

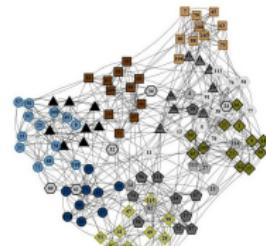
karate club *split detection* [RAK07]

community *examples*

most social networks contain *communities* [GN02]



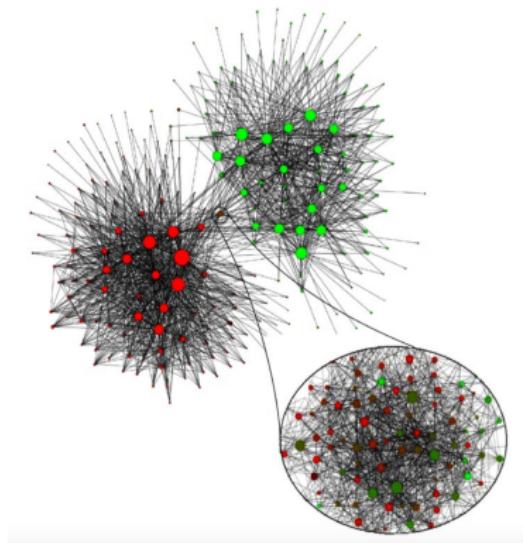
bottlenose dolphins [LSB⁺03]



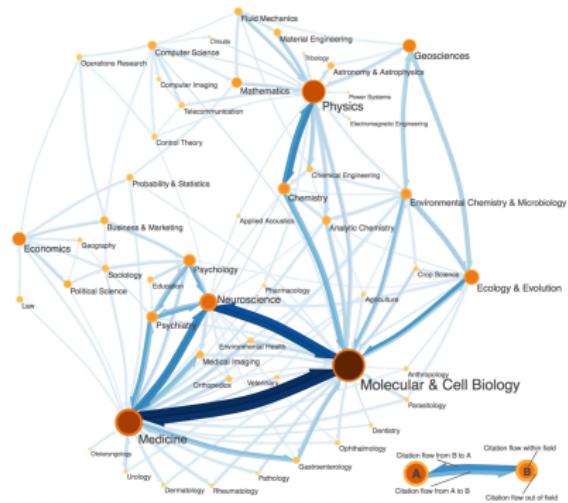
college football [GN02]

community *examples*

many *information networks* contain *communities* [FLG00]



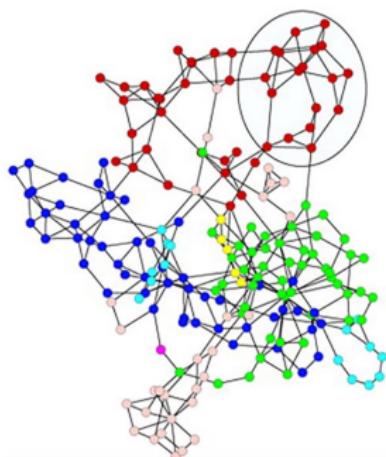
mobile communications [BGLL08]



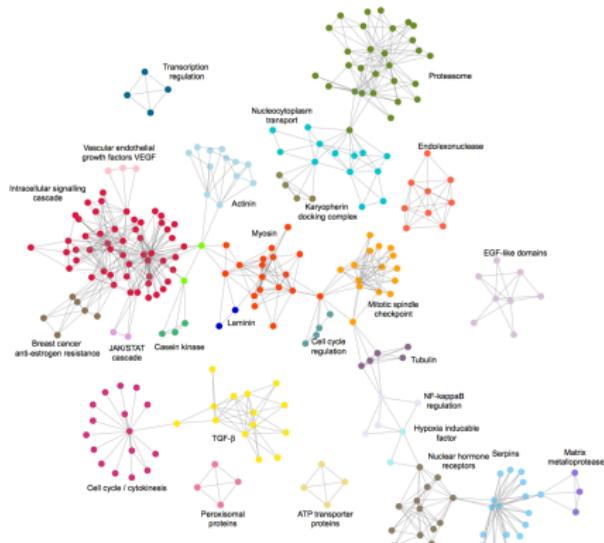
journal citations [RB08]

community examples

many biological networks contain *communities* [RSM⁺02]



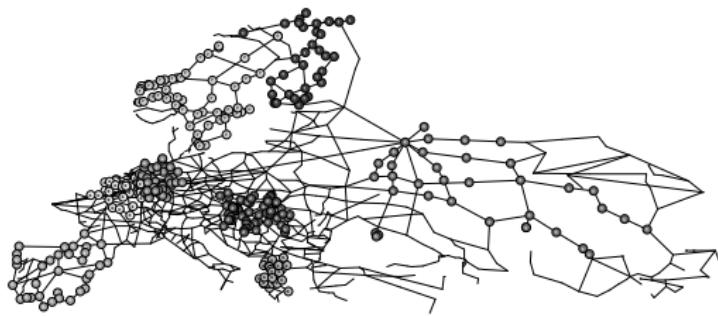
E. coli metabolism [RSM⁺02]



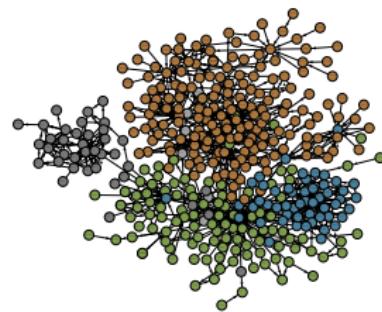
protein interactions [JCZB06]

community *examples*

technological networks rarely contain *communities* [ŠB11a]



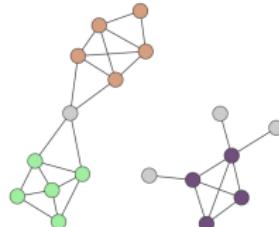
European highways [ŠB11b]



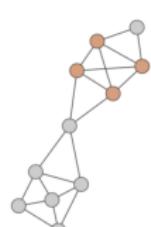
JUNG dependencies [ŠB11a]

community *definition*

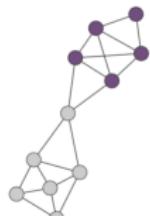
- *clique* is *complete subgraph of some graph*
- *community* is *dense subgraph of sparse network* [GN02]
- *strong* and *weak community* C [FLG00, RCC⁺04] defined as
 - k_i^{int} and k_i^{ext} are *internal* and *external degree* of i
$$\forall i \in C : k_i^{\text{int}} > k_i^{\text{ext}} \quad \sum_{i \in C} k_i^{\text{int}} > \sum_{i \in C} k_i^{\text{ext}}$$
- *community detection* is \gg *graph partitioning* [For10]



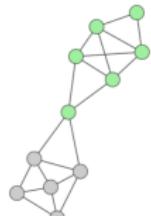
connected communities



maximum clique



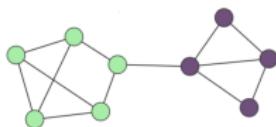
strong community



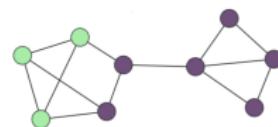
weak community

community *modularity*

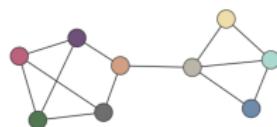
- random graphs should lack community structure
- modularity Q [GN02] of communities $\{C\}$ defined as
$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - p_{ij}) \delta_{c_i c_j} = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta_{c_i c_j}$$
- $p_{ij} = \frac{k_i k_j}{2m}$ is link probability in configuration model [NSW01]
- modularity Q popular quality/optimization function [For10]



optimal $Q = 0.41$



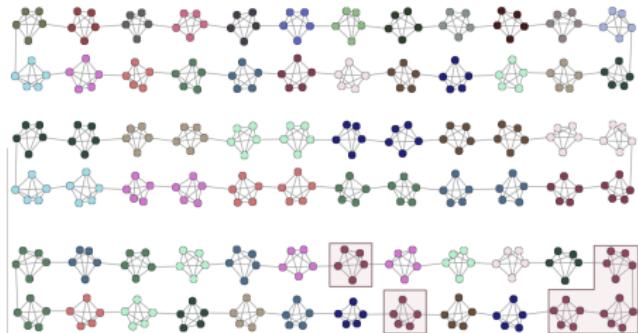
suboptimal $Q = 0.22$



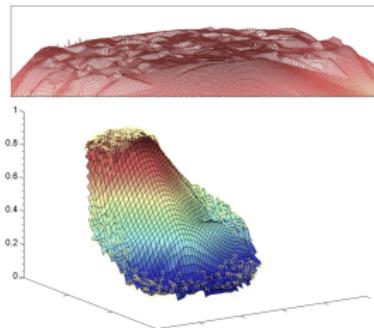
isolates $Q = -0.12$

community \neg modularity

- modularity $Q \gg 0$ also in random graphs [GSPA04]
- modularity Q has resolution limit at $k_c \leq \sqrt{2m}$ [FB07]
- modularity Q lacks clear optimum in real networks [GdMC10]

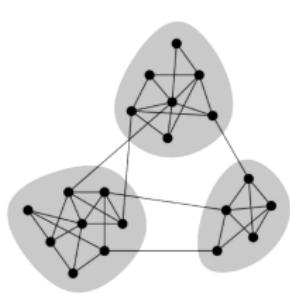


intuitive $Q = 0.867$, optimal $Q = 0.871$ and random $Q = 0.8$

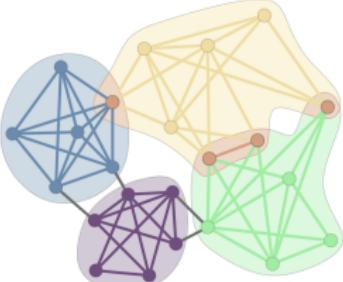


Q plateau and maxima

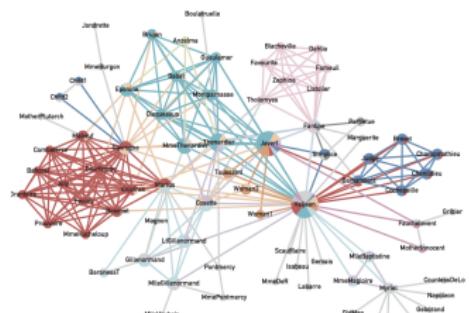
community *overview*



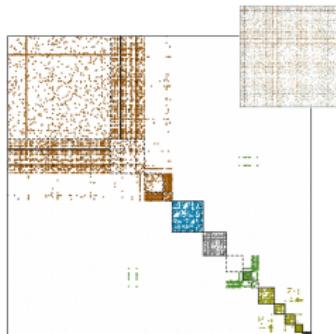
communities [GN02]



overlapping communities [PDFV05]



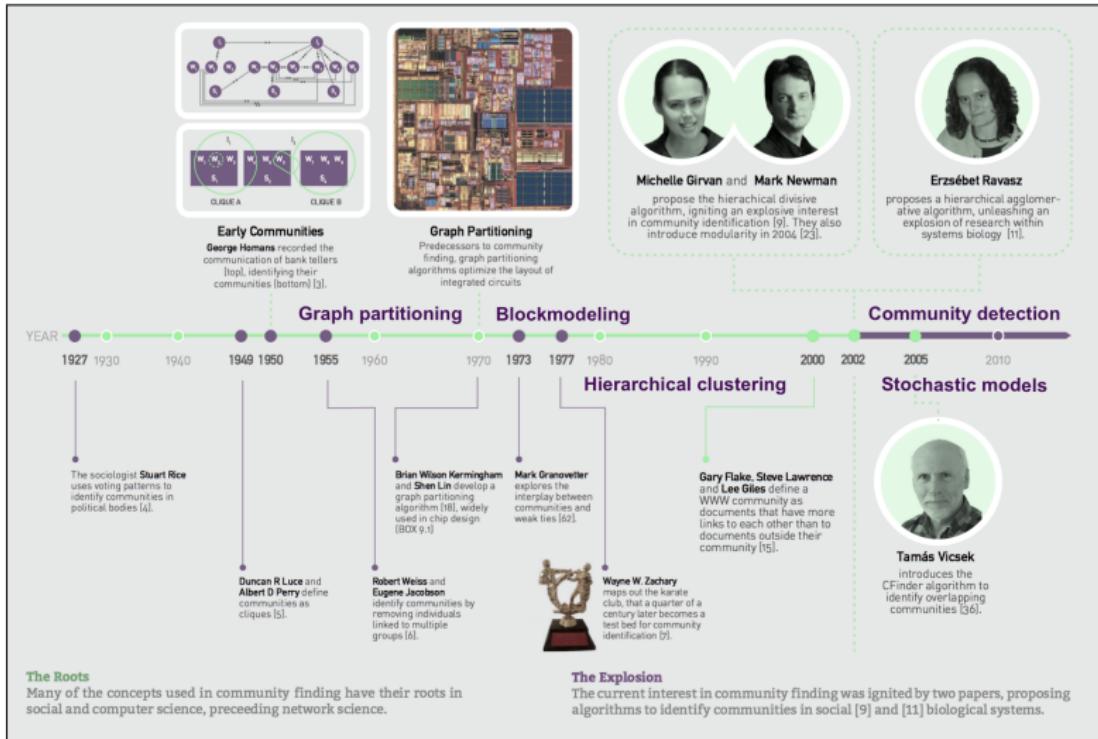
link communities [EL09, ABL10]



block models, blockmodeling etc.

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

community *history*



community *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.
-  Santo Fortunato and Marc Barthelemy.
Resolution limit in community detection.
P. Natl. Acad. Sci. USA, 104(1):36–41, 2007.

community *references*

-  Gary William Flake, Steve Lawrence, and C. Lee Giles.
Efficient identification of web communities.
In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 150–160, Boston, MA, USA, 2000.
-  Santo Fortunato.
Community detection in graphs.
Phys. Rep., 486(3-5):75–174, 2010.
-  Benjamin H. Good, Yves Alexandre de Montjoye, and Aaron Clauset.
Performance of modularity maximization in practical contexts.
Phys. Rev. E, 81(4):046106, 2010.
-  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 Luís A. Nunes Amaral.
Modularity from fluctuations in random graphs and complex networks.
Phys. Rev. E, 70(2):025101, 2004.
-  Pall F. Jonsson, Tamara Cavanna, Daniel Zicha, and Paul A. Bates.
Cluster analysis of networks generated through homology: Automatic identification of important protein communities involved in cancer metastasis.
BMC Bioinformatics, 7:2, 2006.
-  D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson.
The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait?
Behav. Ecol. Sociobiol., 54(4):396–405, 2003.

community *references*

-  Mark E. J. Newman.
Networks: An Introduction.
Oxford University Press, Oxford, 2010.
-  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.
-  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.
-  Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi.
Defining and identifying communities in networks.
P. Natl. Acad. Sci. USA, 101(9):2658–2663, 2004.
-  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.
-  Lovro Šubelj and Marko Bajec.
Community structure of complex software systems: Analysis and applications.
Physica A, 390(16):2968–2975, 2011.

community *references*

-  Lovro Šubelj and Marko Bajec.
Robust network community detection using balanced propagation.
Eur. Phys. J. B, 81(3):353–362, 2011.
-  Wayne W. Zachary.
An information flow model for conflict and fission in small groups.
J. Anthropol. Res., 33(4):452–473, 1977.