1 Cut

minimising the number of edges cut [1, 2, 3]

this only cut small subgraphs, so ration cut [4], normalised cut [5] and min-max cut [3]

[6] local parititoning algorithm using a variation of PageRank the ordering of vertices produced by PageRank produces a cut with small conductance

Metis [7] - MQI [8, 9] algorithm metis - splits graph into two equal sized pieces MQI - low

2 modularity

[10] seminal hierarchical divisive algorithm removed edges based on their betweeness until modularity quality function is maximised

$$Q - \sum_{i=1}^{k}(e_{ii} - a_i^2)$$

where e_i is the probability that edge i is in module i and a_i is the probability that a random edge would be in module i

[11] sped up [12] links are iteratively added to produce the largest increase in modularity

[13] same as above but with simulated annealing

[14] multistage algorithm local optimisation of GN modularity on neighbourhood of each node communities are replaced by a single node repeat until modularity stops increasing

[15] dividing hierarchical like GN does not use edge betweenness - uses edge clustering co-efficient based on loops and stops based on quality of communities not modularity strong community - internal ¡ external degree week community - total internal ¡ total external degree

[16] cfinder first algorithm to consider overlapping communities rolls k-cliques across networks computationally expensive – time increases exponentially with graph size

[17] markov clustering algorithm (MCL) two phases to simulate diffusion process on graph - expansion/inflation only k largest elements are kept at each phase for efficiency

[18] translates the problem of clustering to the problem of optimally compressing the information on the graph so that the most information is uncovered when the compression is decoded uses SA to minimise a function that represents compression and data loss slow

[19] same principle as above but with a dynamic process used SA to optimise a compression based on minimum length of a random walk

[20] based on spectral components of graph select g eigen vectors and project points to g-dimensional space group with traditional hierarchgical clustering techiniques and maximise modularity

[21] uses bayesian inference to deduce best model to fit data represented by graph structure advantage: can find best group structure (not necessarily communities) disadvantage: k must be known a-priori

[22] modularity optimisation can lead to incorrect community divisions due to a 'resolution limit' (fails to identify modules smaller than a scale that depends on the size of the network)

[23] potts finding ground state of spin can find hierarchies and overlap 'communities structure interpreted as the spin configuration that minimises energy of spin glass' uses modularity of random graphs as null model

[24] potts method minimise hamiltonian of potts-like spin model resolution parameter for scale of communities to detect

[25] overcomes the resolution limit of modularity optimisation by not comparing to a null model but instead penalising communities for missing edges

[26] two contributions: multi-scale quality functions that can uncover hieracrhical communities and produce several differnent partitions. post-processing of clusters found by hierarchical methods (encoded in a dendogram)

[27] proposes Probabalistically Mining Communities (PMC) two phases - heuristic - random walk to reduce search space for communities optimisation - constrained quadratic objective function integrates advantages of both and offers a trade-off between effectiveness and efficiency

[28] conductance:

$$\phi(S) = \frac{c_s}{\min(Vol(S), Vol(V \setminus S)}$$

with

$$c_s = |\{(u, v) : u \in S, v \notin S\}|$$

best conductance cut

modularity for bipartite networks: [29]

3 real-time

[30] analyse 'label propagation/epidemic' a technique proposed in [31]

4 spectral-based

5 overlapping

[32] clique percolation method (CPM) overlapping large scale networks communities are unions of k-cliques

[33] proposed EAGLE to detect hierarchical and overlapping communities on the word association and scientific collaboration networks

[34] first algorithm for both hierarchical and overlapping communities local algorithm to optimise

$$f_G = \frac{k_{in}^G}{(k_{in}^G + k_{out}^G)^\alpha}$$

[35] bayesian non-negative matrix factorisation 'soft-partitioning' and assigns node participation scores to modules

detailed overview: [36]

6 EA

locus-based adjacency [37]

[38] proposed MOCK a multi objective EA for clustering with automatic k-determination two objective functions - connectedness and deviation two phases - clustering using adapted PESA-II and an encoded that does not require k a priori model selection that selected best clustering based on synthetic unclustered data generated from a poisson distribution

[39] defined quality function based on power mean changing power r changes the weight of densely inter-connected nodes

7 SOM

seminal paper [40] based on earlier work about topological ordering [41] later work by kohonen – LVQ [42]

[43] clustering is under researched successful for classification: medical diagnosis [44] image and character recognition [45, 46] speech recognition [47]

[43] clusters SOM output using contiguity constrained clustering also [48] clusters output of SOM -compared k-means and hierarchical performs well against clustering alone with less computation time

[49] unsupervised and supervised (RBF) SOM for clustering automatically finds network structure and size controlled growth and removal

[50] som with graph input constrained weights to point to edges on a graph shortest path metric is used

$$w = (n_i, n_j, \beta)$$

References

[1] Brian W Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *Bell system technical journal*, 49(2):291–307, 1970.

[2] Charles M Fiduccia and Robert M Mattheyses. A linear-time heuristic for improving network partitions. In *Design Automation, 1982. 19th Conference on*, pages 175–181. IEEE, 1982.

[3] Chris HQ Ding, Xiaofeng He, Hongyuan Zha, Ming Gu, and Horst D Simon. A min-max cut algorithm for graph partitioning and data clustering. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 107–114. IEEE, 2001.

[4] Y-C Wei and C-K Cheng. Ratio cut partitioning for hierarchical designs. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 10(7):911–921, 1991.

[5] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8):888–905, 2000.

[6] Reid Andersen, Fan Chung, and Kevin Lang. Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 475–486. IEEE, 2006.

[7] George Karypis and Vipin Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1):359–392, 1998.

[8] Giorgio Gallo, Michael D Grigoriadis, and Robert E Tarjan. A fast parametric maximum flow algorithm and applications. *SIAM Journal on Computing*, 18(1):30–55, 1989.

[9] Kevin Lang and Satish Rao. A flow-based method for improving the expansion or conductance of graph cuts. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 325–337. Springer, 2004.

[10] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[11] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.

[12] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[13] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

[14] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[15] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2658–2663, 2004.

[16] Balázs Adamcsek, Gergely Palla, Illés J Farkas, Imre Derényi, and Tamás Vicsek. Cfinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, 22(8):1021–1023, 2006.

[17] Stijn Marinus Van Dongen. *Graph clustering by flow simulation*. PhD thesis, 2001.

[18] Martin Rosvall and Carl T Bergstrom. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences*, 104(18):7327–7331, 2007.

[19] Martin Rosvall and Carl T Bergstrom. Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123, 2008.

[20] L Donetti and MA Munoz. Detecting network communities: a new and systematic approach. *Journal of Statistical Mechanics: Theory and Experiment P*, 10012:2004, 2004.

[21] Mark EJ Newman and Elizabeth A Leicht. Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences*, 104(23):9564–9569, 2007.

[22] Santo Fortunato and Marc Barthelemy. Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, 104(1):36–41, 2007.

[23] Jörg Reichardt and Stefan Bornholdt. Statistical mechanics of community detection. *Physical Review E*, 74(1):016110, 2006.

[24] Peter Ronhovde and Zohar Nussinov. Multiresolution community detection for megascale networks by information-based replica correlations. *Physical Review E*, 80(1):016109, 2009.

[25] Peter Ronhovde and Zohar Nussinov. Local resolution-limit-free potts model for community detection. *Physical Review E*, 81(4):046114, 2010.

[26] Pascal Pons and Matthieu Latapy. Post-processing hierarchical community structures: Quality improvements and multi-scale view. *Theoretical Computer Science*, 412(8):892–900, 2011.

[27] Bo Yang, Jin Di, Jiming Liu, and Dayou Liu. Hierarchical community detection with applications to real-world network analysis. *Data & Knowledge Engineering*, 83:20–38, 2013.

[28] Jure Leskovec, Kevin J Lang, and Michael Mahoney. Empirical comparison of algorithms for network community detection. In *Proceedings of the 19th international conference on World wide web*, pages 631–640. ACM, 2010.

[29] Michael J Barber. Modularity and community detection in bipartite networks. *Physical Review E*, 76(6):066102, 2007.

[30] Ian XY Leung, Pan Hui, Pietro Lio, and Jon Crowcroft. Towards real-time community detection in large networks. *Physical Review E*, 79(6):066107, 2009.

[31] Usha Nandini Raghavan, Réka Albert, and Soundar Kumara. Near linear time algorithm to detect community structures in large-scale networks. *Physical review E*, 76(3):036106, 2007.

[32] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[33] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1706–1712, 2009.

[34] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.

[35] Ioannis Psorakis, Stephen Roberts, Mark Ebden, and Ben Sheldon. Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E*, 83(6):066114, 2011.

[36] Jierui Xie, Stephen Kelley, and Boleslaw K Szymanski. Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (csur)*, 45(4):43, 2013.

[37] YoungJa Park and ManSuk Song. A genetic algorithm for clustering problems. In *Proceedings of the third annual conference on genetic programming*, pages 568–575, 1998.

[38] Julia Handl and Joshua Knowles. An evolutionary approach to multiobjective clustering. *IEEE transactions on Evolutionary Computation*, 11(1):56–76, 2007.

[39] Clara Pizzuti. Ga-net: A genetic algorithm for community detection in social networks. In *International Conference on Parallel Problem Solving from Nature*, pages 1081–1090. Springer, 2008.

[40] Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[41] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

[42] Teuvo Kohonen. Learning vector quantization. In *Self-Organizing Maps*, pages 175–189. Springer, 1995.

[43] Melody Y Kiang. Extending the kohonen self-organizing map networks for clustering analysis. *Computational Statistics & Data Analysis*, 38(2):161–180, 2001.

[44] L Vercauteren, G Sieben, and M Praet. The classification of brain tumours by a topological map. 1990.

[45] Michael Sabourin and Amar Mitiche. Modeling and classification of shape using a kohonen associative memory with selective multiresolution. *Neural Networks*, 6(2):275–283, 1993.

[46] Alberto Delbimbo, Leonardo Landi, and Simone Santini. Three-dimensional planar-faced object classification with kohonen maps. *Optical Engineering*, 32(6):1222–1234, 1993.

[47] Lea Leinonen, Tapio Hiltunen, Kari Torkkola, and Jari Kangas. Self-organized acoustic feature map in detection of fricative-vowel coarticulation. *The Journal of the Acoustical Society of America*, 93(6):3468–3474, 1993.

[48] Juha Vesanto and Esa Alhoniemi. Clustering of the self-organizing map. *IEEE transactions on neural networks*, 11(3):586–600, 2000.

[49] Bernd Fritzke. Growing cell structuresa self-organizing network for unsupervised and supervised learning. *Neural networks*, 7(9):1441–1460, 1994.

[50] Takeshi Yamakawa, Keiichi Horio, and Masaharu Hoshino. Self-organizing map with input data represented as graph. In *International Conference on Neural Information Processing*, pages 907–914. Springer, 2006.