САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Математико-механический факультет

Кафедра Информатики

Проданов Тимофей Петрович

Адаптивный рандомизированный алгоритм выделения сообществ в графах

Бакалаврская работа

Допущена к защите.
Зав. кафедрой:

Научный руководитель:
д. ф.-м. н., профессор О.Н. Граничин

Рецензент:
В.А. Ерофеева

Санкт-Петербург
2015

SAINT-PETERSBURG STATE UNIVERSITY
Mathematics & Mechanics Faculty

Department of Computer Science

Timofey Prodanov

Adaptive randomised algorithm for community detection in graphs

Bachelor's Thesis

Admitted for defence.
Head of the chair:

Scientific supervisor:
Professor Oleg Granichin

Reviewer:
Victoria Erofeeva

Saint-Petersburg
2015

Оглавление

0.1. Размер возмущения

Коэффициент d отвечает за то, насколько сильно будет возмущаться центральная точка для получения следующих измерений. То есть, насколько k_n^+ и k_n^- будут отличаться от $\hat{k}_{n-1}$. Зависимость модулярности от размера возмущения при $f(Q,k) = -10\ln Q,\ \sigma = 500,\ k_0 = 10$ будет выглядеть следующим образом:

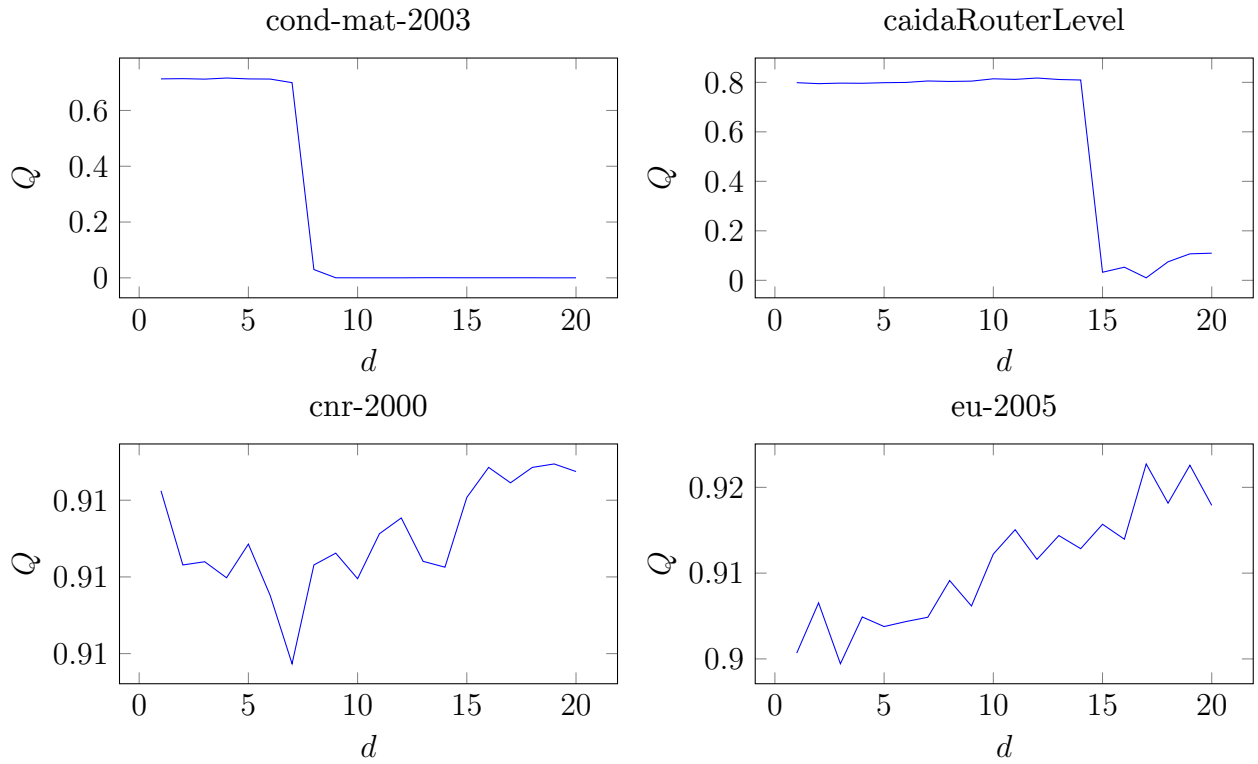
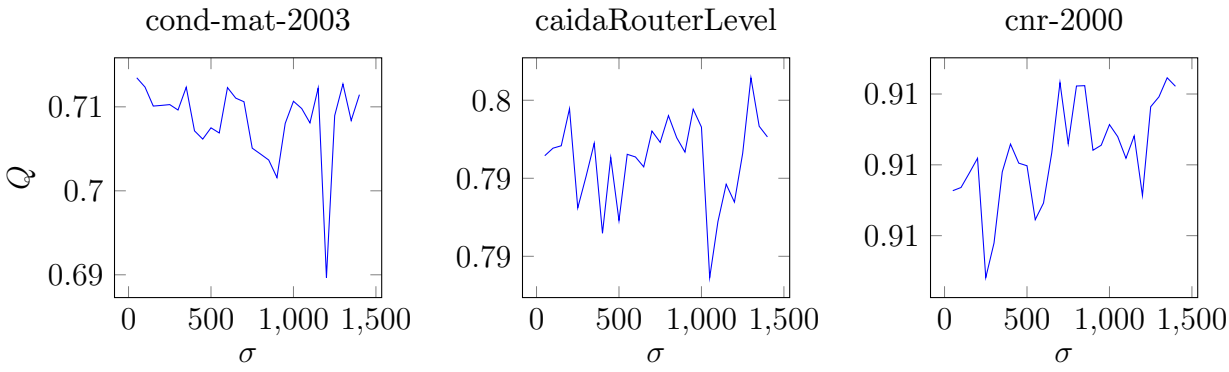


Рис. 1: Зависимость модулярности от размера возмущения на четырёх графах

На графах *cnr-2000* и *eu-2005* значения модулярности не очень сильно менялись в зависимости от параметра d, хотя некоторые значения d и давали более большие значения. Однако на графах *cond-mat-2003* и *caidaRouterLevel* после некоторого порогового значения возмущения модулярность показывала, что получившееся разбиение не лучше случайного.

0.2. Количество итераций в одном шаге

Параметр σ указывает, как часто меняется k в рандомизированного жадном алгоритме. Так как в функции качества используется медиана прироста модулярности, а не прирост модулярности за все σ шагов — при изменении σ нет необходимости менять функцию качества, модулярность прироста будет оставаться приблизительно такой же по величине, в то время как прирост модулярности линейно зависит от σ. Зависимость модулярности от количества итераций в одном шаге при $d = 5,\ f(Q,k) = -10\ln Q,\ k_0 = 10$ принимает такой вид:


cond-mat-2003
caidaRouterLevel
cnr-2000
Q
0.71
0.7
0.69
0.8
0.79
0.79
0.91
0.91
0.91
0
500
1,000
1,500
σ

Список литературы

[1] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.

[2] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *ACM SIGCOMM Computer Communication Review*, volume 29, pages 251–262. ACM, 1999.

[3] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Computer networks*, 33(1):309–320, 2000.

[4] Cristopher Moore and Mark EJ Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000.

[5] Jing Zhao, Hong Yu, Jianhua Luo, ZW Cao, and Yixue Li. Complex networks theory for analyzing metabolic networks. *Chinese Science Bulletin*, 51(13):1529–1537, 2006.

[6] Wang Hong, Wang Zhao-wen, Li Jian-bo, and Qiu-hong Wei. Criminal behavior analysis based on complex networks theory. In *IT in Medicine & Education, 2009. ITIME'09. IEEE International Symposium on*, volume 1, pages 951–955. IEEE, 2009.

[7] John Scott. *Social network analysis*. Sage, 2012.

[8] Stanley Wasserman. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.

[9] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, Feb 2004.

[10] Stefanie Muff, Francesco Rao, and Amedeo Caflisch. Local modularity measure for network clusterizations. *arXiv preprint cond-mat/0503252*, 2005.

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

[12] Ulrik Brandes, Daniel Delling, Marco Gaertler, Robert Gorke, Martin Hoefer, Zoran Nikoloski, and Dorothea Wagner. On modularity clustering. *Knowledge and Data Engineering, IEEE Transactions on*, 20(2):172–188, 2008.

[13] Mark E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69:066133, Jun 2004.

[14] Michael Ovelgönne and Andreas Geyer-Schulz. A comparison of agglomerative hierarchical algorithms for modularity clustering. In *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*, pages 225–232. Springer, 2012.

[15] Michael Ovelgönne and Andreas Geyer-Schulz. Cluster cores and modularity maximization. In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pages 1204–1213. IEEE, 2010.

[16] Michael Ovelgönne and Andreas Geyer-Schulz. An ensemble learning strategy for graph clustering. *Graph Partitioning and Graph Clustering*, 588:187, 2012.

[17] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

[18] Jack Kiefer and Jacob Wolfowitz. Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[19] Julius R Blum. Multidimensional stochastic approximation methods. *The Annals of Mathematical Statistics*, pages 737–744, 1954.

[20] James C Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

[21] James C. Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control*, volume 65. John Wiley & Sons, 2005.

[22] Boris T. Polyak. *Introduction to optimization*. Optimization Software New York, 1987.

[23] Oleg Granichin and Natalia Amelina. Simultaneous perturbation stochastic approximation for tracking under unknown but bounded disturbances. *IEEE Transactions on Automatic Control*, 60(5), 2015.

[24] Wayne W Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, pages 452–473, 1977.

[25] David Lusseau, Karsten Schneider, Oliver J Boisseau, Patti Haase, Elisabeth Slooten, and Steve M Dawson. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, 2003.

[26] Daniel Baird and Robert E Ulanowicz. The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecological Monographs*, pages 329–364, 1989.

[27] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.

[28] 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.

[29] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):027104, 2005.

[30] Pablo M. Gleiser and Leon Danon. Community structure in jazz. *Advances in complex systems*, 6(04):565–573, 2003.

[31] Roger Guimerà Manrique, L Danon, Albert Díaz Guilera, Francesc Giralt, and Àlex Arenas. Self-similar community structure in a network of human interactions. *Physical Review E, 2003, vol. 68, núm. 6, p. 065103-1-065103-4*, 2003.

[32] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.

[33] Marián Boguñá, Romualdo Pastor-Satorras, Albert Díaz-Guilera, and Alex Arenas. Models of social networks based on social distance attachment. *Phys. Rev. E*, 70:056122, Nov 2004.

[34] Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.

[35] Paolo Boldi and Sebastiano Vigna. The WebGraph framework I: Compression techniques. In *Proc. of the Thirteenth International World Wide Web Conference (WWW 2004)*, pages 595–601, Manhattan, USA, 2004. ACM Press.

[36] Paolo Boldi, Marco Rosa, Massimo Santini, and Sebastiano Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *Proceedings of the 20th international conference on World Wide Web*. ACM Press, 2011.

[37] Paolo Boldi, Bruno Codenotti, Massimo Santini, and Sebastiano Vigna. Ubicrawler: A scalable fully distributed web crawler. *Software: Practice & Experience*, 34(8):711–726, 2004.