



دانشکده مهندسی کامپیوتر

یافتن اجتماعات در گراف‌های حجیم به کمک بررسی خواص توپولوژیکی شبکه‌های پیچیده زیرین

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته مهندسی کامپیوتر
گرایش مهندسی نرم‌افزار

حسن عابدی

استاد راهنما

حسن نادری

اردیبهشت ۱۳۹۵



تأییدی هیأت داوران جلسه دفاع از پایان نامه

نام دانشکده: دانشکده مهندسی کامپیوتر

نام دانشجو: حسن عابدی

عنوان پایان نامه: یافتن اجتماعات در گراف‌های حجیم به کمک بررسی خواص توپولوژیکی

شبکه‌های پیچیده زیرین

تاریخ دفاع: اردیبهشت ۱۳۹۵

رشته: مهندسی کامپیوتر

گرایش: مهندسی نرم افزار

ردیف	سمت	نام و نام خانوادگی	مرتبه دانشگاهی	دانشگاه یا مؤسسه	امضاء
۱	استاد راهنما	دکتر حسن نادری	استادیار	دانشگاه علم و صنعت ایران	
۲	استاد داور داخلی	دکتر عین‌اله خنجری	استادیار	دانشگاه علم و صنعت ایران	
۳	استاد داور خارجی	دکتر محمد صنیعی آباده	استادیار	دانشگاه تربیت مدرس	

تأییدی صحت و اصالت نتایج

باسمه تعالی

اینجانب حسن عابدی به شماره دانشجویی ۹۲۷۲۳۱۴۷ دانشجوی رشته مهندسی کامپیوتر مقطع تحصیلی کارشناسی ارشد تأیید می‌نمایم که کلیه نتایج این پایان‌نامه حاصل کار اینجانب و بدون هرگونه دخل و تصرف است و موارد نسخه‌برداری شده از آثار دیگران را با ذکر کامل مشخصات منبع ذکر کرده‌ام. در صورت اثبات خلاف مندرجات فوق، به تشخیص دانشگاه مطابق با ضوابط و مقررات حاکم (قانون حمایت از حقوق مؤلفان و مصنفان و قانون ترجمه و تکثیر کتب و نشریات و آثار صوتی، ضوابط و مقررات آموزشی، پژوهشی و انضباطی) با اینجانب رفتار خواهد شد و حق هرگونه اعتراض در خصوص احقاق حقوق مکتسب و تشخیص و تعیین تخلف و مجازات را از خویش سلب می‌نمایم. در ضمن، مسئولیت هرگونه پاسخگویی به اشخاص اعم از حقیقی و حقوقی و مراجع ذیصلاح (اعم از اداری و قضایی) به عهده‌ی اینجانب خواهد بود و دانشگاه هیچ‌گونه مسئولیتی در این خصوص نخواهد داشت.

نام و نام خانوادگی: حسن عابدی

تاریخ و امضا:

مجوز بهره‌برداری از پایان‌نامه

- بهره‌برداری از این پایان‌نامه در چهارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنما به شرح زیر تعیین می‌شود، بلامانع است:
- ☐ بهره‌برداری از این پایان‌نامه برای همگان بلامانع است.
 - ☐ بهره‌برداری از این پایان‌نامه با اخذ مجوز از استاد راهنما، بلامانع است.
 - ☐ بهره‌برداری از این پایان‌نامه تا تاریخ ممنوع است.

استاد راهنما: حسن نادری

تاریخ:

امضا:

تقديم به:

پدر و مادرم.

چکیده

ساختار اجتماع خصوصیتی فراگیر در شبکه‌های پیچیده است. مساله‌ی یافتن اجتماعات در این شبکه‌ها جزو مسایل مورد توجه محققین در چند سال اخیر بوده است. اجتماع مجموعه‌ای از گره‌های گراف می‌باشد که در عین حالی که با هم دارای اتصالات زیادی می‌باشند از بقیه گراف به خوبی مجزا هستند. گراف شبکه‌های پیچیده دارای خواص ساختاری مانند کوتاهی فاصله دو گره دلخواه هستند که آن‌ها را از گراف‌های تصادفی مطالعه شده در گذشته متمایز می‌نماید. از طرفی الگوریتم‌های یافتن اجتماعات را می‌توان به دو دسته‌ی الگوریتم‌های محلی و سراسری تقسیم کرد. یکی از چالش‌های الگوریتم‌های محلی انتخاب گره‌های دانه است. در این پایان‌نامه روشی برای یافتن اجتماعات روی گراف شبکه‌های پیچیده به کمک انتخاب دانه‌های مرغوب و بسط این گره‌ها توسط یک الگوریتم محلی ارائه شده است. روش پیشنهادی ۳ مرحله دارد، در مرحله‌ی نخست گره‌های گراف ورودی به کمک یک استراتژی حریصانه به چندین افراز تقسیم می‌شوند. این افرازها نشان دهنده‌ی اجتماعات اولیه گراف ورودی خواهند بود. در گام دوم درون هر زیرگراف حاصل از گره‌های درون یک افراز و اتصالات میان گره‌های آن به دنبال گره‌هایی که به احتمال زیادی به خوبی در بطن یک اجتماع واقع شده‌اند می‌گردیم. در این مرحله در هر زیرگراف به صورت موازی به کمک بررسی همسایگی گره‌های با درجه بالا، گره‌هایی را که می‌توانند به خوبی نشانگر اجتماع خود باشند را به عنوان گره دانه برمی‌گزینیم. در گام آخر اجتماعاتی که هر گره دانه در آن قرار گرفته است را به کمک الگوریتمی که بر پایه محاسبه‌ی بردار Personalized Pagerank عمل می‌کند، می‌یابیم. برای آزمایش کیفیت اجتماعات خروجی روش پیشنهادی خود، از ۲۲ گراف استاندارد که در ۷ دسته مختلف از شبکه‌های پیچیده قرار می‌گیرند استفاده نموده ایم. برای سنجش کیفیت اجتماعات روش پیشنهادی میزان ۵ معیار مختلف سنجش اجتماعات برای خروجی روش ما و ۳ روش دیگر آزمایش شده است. کیفیت اجتماعات روش پیشنهادی برای ۲ معیار از ۵ معیار بسیار بهتر از خروجی دیگر روش‌هاست، برای ۳ معیار دیگر هم عملکرد روش پیشنهادی بسیار شبیه عملکرد روش با بهترین خروجی بوده است. نتایج نشان می‌دهد که نباید تنها درجه یک گره را به عنوان معیار صرف دانه بودن آن انتخاب کرد.

واژگان کلیدی: یافتن اجتماعات، شبکه‌های پیچیده، الگوریتم‌های گراف

فهرست مطالب

چ	فهرست تصاویر
ح	فهرست جداول
خ	فهرست الگوریتم‌ها
د	فهرست علائم اختصاری
۱	فصل ۱: فصل مقدمه
۲	فصل ۲: مروری بر ادبیات و کارهای انجام شده
۳	فصل ۳: روش پیشنهادی
۴	فصل ۴: آزمایش و نتایج
۵	فصل ۵: جمع‌بندی و پیشنهادها
۶	مراجع
۱۱	واژه‌نامه فارسی به انگلیسی
۱۲	واژه‌نامه انگلیسی به فارسی

فهرست تصاویر

فهرست جداول

فهرست الگوریتم‌ها

فهرست علایم اختصاری

C	اجتماع
Φ	رسانش
Ω	ماژولاریتی
G	گراف
V	مجموعه گره‌های گراف G
E	مجموعه اتصالات گراف G
$n = V $	تعداد گره‌ها
$m = E $	تعداد یال‌ها
$deg(v)$	درجه گره v
$N_{deg(x)}$	تعداد گره‌هایی که x یال دارند
a_G	میانگین تعداد یال‌های هر گره گراف G
CC	Clustering Coefficient
Dia	قطر حقیقی گراف
Dia_{ef}	قطر موثر گراف
V_g	مجموعه گره‌های گراف g
E_g	مجموعه اتصالات گراف g

فصل ۱

فصل مقدمه

فصل ۲

مروری بر ادبیات و کارهای انجام شده

فصل ۳

روش پیشنهادی

فصل ۴

آزمایش و نتایج

فصل ۵

جمع‌بندی و پیشنهادها

مراجع

- [1] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol.2008, no.10, p.P10008, 2008.
- [2] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E*, vol.74, no.1, p.016110, 2006.
- [3] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol.70, no.6, p.066111, 2004.
- [4] C. Biemann, "Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems," in *Proceedings of the first workshop on graph based methods for natural language processing*, pp.73–80, Association for Computational Linguistics, 2006.
- [5] J. Kleinberg and E. Tardos. *Algorithm Design*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- [6] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *Proceedings of the National Academy of Sciences*, vol.104, no.1, pp.36–41, 2007.
- [7] J. Yang and J. Leskovec, "Overlapping community detection at scale: a nonnegative matrix factorization approach," in *Proceedings of the sixth ACM international conference on Web search and data mining*, pp.587–596, ACM, 2013.
- [8] M. G. Everett, "Classical algorithms for social network analysis: Future and current trends," in *Encyclopedia of Social Network Analysis and Mining*, pp.88–94, 2014.
- [9] J. Leskovec, K. J. Lang, and M. Mahoney, "Empirical comparison of algorithms for network community detection," in *Proceedings of the 19th international conference on World wide web*, pp.631–640, ACM, 2010.

- [10] D. Hric, R. K. Darst, and S. Fortunato, "Community detection in networks: Structural communities versus ground truth," *Physical Review E*, vol.90, no.6, p.062805, 2014.
- [11] M. E. Newman, "Communities, modules and large-scale structure in networks," *Nature Physics*, vol.8, no.1, pp.25–31, 2012.
- [12] D. Easley and J. Kleinberg. *Networks, crowds, and markets reasoning about a highly connected world*. New York: Cambridge University Press, 2010.
- [13] S. L. Tauro, C. Palmer, G. Siganos, and M. Faloutsos, "A simple conceptual model for the internet topology," in *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, vol.3, pp.1667–1671, IEEE, 2001.
- [14] K. H. Rosen. *Handbook of Discrete and Combinatorial Mathematics, Second Edition*. Chapman & Hall/CRC, 2nd ed. , 2010.
- [15] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *nature*, vol.393, no.6684, pp.440–442, 1998.
- [16] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E*, vol.76, no.3, p.036106, 2007.
- [17] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, "Detect overlapping and hierarchical community structure in networks," *Physica A: Statistical Mechanics and its Applications*, vol.388, no.8, pp.1706–1712, 2009.
- [18] B. Abrahao, S. Soundarajan, J. Hopcroft, and R. Kleinberg, "On the separability of structural classes of communities," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.624–632, ACM, 2012.
- [19] D. F. Gleich and C. Seshadhri, "Vertex neighborhoods, low conductance cuts, and good seeds for local community methods," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.597–605, ACM, 2012.
- [20] I. M. Kloumann and J. M. Kleinberg, "Community membership identification from small seed sets," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp.1366–1375, ACM, 2014.
- [21] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Data mining (ICDM), 2013 IEEE 13th international conference on*, pp.1151–1156, IEEE, 2013.

- [22] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol.51, no.4, pp.661–703, 2009.
- [23] P. Pons and M. Latapy, "Computing communities in large networks using random walks," in *Computer and Information Sciences-ISCIS 2005*, pp.284–293, Springer, 2005.
- [24] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol.69, no.2, p.026113, 2004.
- [25] A. Clauset, "Finding local community structure in networks," *Physical review E*, vol.72, no.2, p.026132, 2005.
- [26] D. Gibson, J. Kleinberg, and P. Raghavan, "Inferring web communities from link topology," in *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space—structure in hypermedia systems: links, objects, time and space—structure in hypermedia systems*, pp.225–234, ACM, 1998.
- [27] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol.46, no.5, pp.604–632, 1999.
- [28] Y. J. Wu, H. Huang, Z. F. Hao, and F. Chen, "Local community detection using link similarity," *Journal of computer science and technology*, vol.27, no.6, pp.1261–1268, 2012.
- [29] M. J. Olsen, "Community detection in large social networks," 2014.
- [30] Q. Chen and M. Fang, "An efficient algorithm for community detection in complex networks," in the *6th Workshop on Social Network Mining and Analysis*, 2012.
- [31] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E*, vol.76, no.3, p.036106, 2007.
- [32] J. Yang and J. Leskovec, "Defining and evaluating network communities based on ground-truth," *Knowledge and Information Systems*, vol.42, no.1, pp.181–213, 2015.
- [33] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol.2008, no.10, p.P10008, 2008.
- [34] L. G. Jeub, P. Balachandran, M. A. Porter, P. J. Mucha, and M. W. Mahoney, "Think locally, act locally: Detection of small, medium-sized, and large communities in large networks," *Physical Review E*, vol.91, no.1, p.012821, 2015.

- [35] C. L. Staudt and H. Meyerhenke, "Engineering parallel algorithms for community detection in massive networks," 2014.
- [36] R. Andersen and K. J. Lang, "Communities from seed sets," in Proceedings of the 15th international conference on World Wide Web, pp.223–232, ACM, 2006.
- [37] B. Collingsworth and R. Menezes, "A self-organized approach for detecting communities in networks," Social Network Analysis and Mining, vol.4, no.1, pp.1–12, 2014.
- [38] C. L. Staudt, Y. Marrakchi, and H. Meyerhenke, "Detecting communities around seed nodes in complex networks," in Big Data (Big Data), 2014 IEEE International Conference on, pp.62–69, IEEE, 2014.
- [39] S. Harenberg, G. Bello, L. Gjeltma, S. Ranshous, J. Harlalka, R. Seay, K. Padmanabhan, and N. Samatova, "Community detection in large-scale networks: a survey and empirical evaluation," Wiley Interdisciplinary Reviews: Computational Statistics, vol.6, no.6, pp.426–439, 2014.
- [40] J. P. Bagrow and E. M. Bollt, "Local method for detecting communities," Physical Review E, vol.72, no.4, p.046108, 2005.
- [41] F. Moradi, T. Olovsson, and P. Tsigas, "A local seed selection algorithm for overlapping community detection," in Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on, pp.1–8, IEEE, 2014.
- [42] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in Proceedings of the 2nd ACM workshop on Online social networks, pp.37–42, ACM, 2009.
- [43] J. J. Whang, D. F. Gleich, and I. S. Dhillon, "Overlapping community detection using neighborhood-inflated seed expansion," Transactions on Knowledge and Data Engineering, p.Accepted, 2016.
- [44] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," science, vol.286, no.5439, pp.509–512, 1999.
- [45] S. Fortunato, "Community detection in graphs," Physics Reports, vol.486, no.3, pp.75–174, 2010.
- [46] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: The state-of-the-art and comparative study," ACM Computing Surveys (csur), vol.45, no.4, p.43, 2013.
- [47] D. Eppstein and D. Strash, "Listing all maximal cliques in large sparse real-world graphs," in Experimental Algorithms, pp.364–375, Springer, 2011.

- [48] E. Estrada and J. A. Rodriguez-Velazquez, "Subgraph centrality in complex networks," *Physical Review E*, vol.71, no.5, p.056103, 2005.
- [49] D. A. Spielman and S. H. Teng, "Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems," in *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pp.81–90, ACM, 2004.
- [50] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol.101, no.9, pp.2658–2663, 2004.
- [51] R. Andersen, F. Chung, and K. Lang, "Local graph partitioning using pagerank vectors," in *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pp.475–486, IEEE, 2006.
- [52] J. J. Whang, D. F. Gleich, and I. S. Dhillon, "Overlapping community detection using seed set expansion," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pp.2099–2108, ACM, 2013.
- [53] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [54] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

واژه‌نامه فارسی به انگلیسی

Probabilistic	احتمالی
Measure	اندازه
Heuristic	هیوریستیک
Topology	توپولوژی
Cut	برش
Experiment	آزمایش
Social Networks	شبکه‌های اجتماعی
Program Fragment	قطعه برنامه
Data Mining	داده کاوی
Graph	گراف
Edge	یال
Node	گره
Centrality	مرکزیت
Global	جهانی
Local	محلی
Lovain	لوون
Unstable	ناپایدار
Weblog	وبلاگ
Post	پست
Partition	افراز
Cluster	خوشه
Overlapping	همپوشان
Bridge	پل
Partition	افراز
CLuster	خوشه

واژه‌نامه انگلیسی به فارسی

Component	مولفه
Community	اجتماع
Complex Networks	شبکه‌های پیچیده
Power-Law Distribution	توزیع توانی
Small-World Phenomenon	پدیده‌ی دنیای کوچک
Probabilistic	احتمالی
Random	تصادفی
Benchmark	محک
Bar-Chart	نمودار میله‌ای
Seed Selection	انتخاب دانه
Expansion	بسط
Social System	سیستم اجتماعی
Real-World	دنیای واقعی
Seed Node	گره دانه
Expansion	بسط
Boundary	پیرامون
Performance	کارایی
Articulation-Point	گره برشی
Cut-Node	گره برشی
Cut-Edge	یال برشی
Expert	متخصص
Cohesiveness	انسجام
Separability	جدایی پذیری

Abstract:

Exploring community structure is an appealing problem that has been drawing much attention in the recent years. One serious problem regarding many community detection methods is that the complete information of real-world networks usually may not be available most of the time, also considering the dynamic nature of such networks(e.g. web pages, collaboration networks and users friendships on social networks), it is most probable possibility that one could detect community structure from a certain source vertex with limited knowledge of the entire network. The existing approaches can do well in measuring the community quality, Nevertheless they are largely dependent on source vertex chosen for the process. Additionally, using unsuitable seed vertices may lead to finding of low quality or erroneous communities for output of many of the algorithms. This paper proposes a method to find better source vertices to be used as seeds to construct community structures locally. Inspired by the fact that many gargantuan real-world networks and respectively their graphs contain a myriad of lightly connected vertices, we explore community structure heuristically by giving priority to vertices which have a high number of links pertaining to the core structure of the network. Experimental results prove that our method can perform effectively for finding high quality seed vertices.

Keywords: Community Detection, Complex Networks, Graph Algorithms



Iran University of Science and Technology
Computer Engineering Department

Community detection on graphs of large real-world complex networks

**A Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master
of Science in Computer Engineering**

By:

Hassan Abedi

Supervisor:

Dr. Hassan Naderi

May 2016