Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 124 (2017) 29–37

Procedia Computer Science

www.elsevier.com/locate/procedia

4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

Community Detection On Citation Network Of DBLP Data Sample Set Using LinkRank Algorithm

Satrio Baskoro Yudhoatmojo*, Muhammad Arvin Samuar

Faculty of Computer Science, Universitas Indonesia, Kampus UI, Depok, 16411, Jawa Barat, Indonesia

Abstract

This paper describes the application of a community detection algorithm, namely LinkRank algorithm, on a citation network. Community detection is a task in network analysis which aims to find sets of tightly connected nodes that are loosely connected with other nodes outside of those sets. In our study, we focused on a citation network which depicts relationships between cited papers and the papers which cite those papers. The objectives of our study are to identify communities of papers based on the citation relationships and analyze the similarities of topics within each community. The approach of our study to reach the objectives is by applying LinkRank algorithm to a citation network. LinkRank algorithm is chosen because it can be applied to a directed network where other algorithms that we have surveyed can only be used on undirected network. The citation network that we used in our study is from Aminer website. In applying the algorithm, we had to port the original source code which is written in C programming language into Python programming language for our convenience in doing the experiment. The result shows that the algorithm able to detect 10,442 communities from 188,514 nodes. Once the communities have been detected, we sampled top three communities (the ones with the largest number of members) and took the top 10 nodes with the highest PageRank score in each of those communities. The samples show that most of the nodes have similar topic, but there are still some nodes with different topics mixed inside the same community. We found the ratio between nodes with similar and different topics to be 7 to 3, that is 70% of the nodes have similar topic while the other 30% have different topics. Thus, the homophily of each community does not reach 100%. Nevertheless, our study confirms that LinkRank algorithm can be used for community detection on directed network.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: Citation Network; Community Detection; Complex Network; Directed Network; LinkRank Algorithm; Social Network Analysis

1. Introduction

A task of citing earlier research publications is a common task and it is a way of acknowledging previous achievement made by other researchers that may become foundations of new researches in the future for discovering new knowledges. Analysis on the citation relationships could reveal several interesting things, one of them is detecting

* Corresponding author. Tel.: +62-21-7863419 ; fax: +62-21-7863415.
E-mail address: satrio.baskoro@cs.ui.ac.id

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017
10.1016/j.procs.2017.12.126

communities of citations. This detection can show how research publications are related to previous research publications which may due to similar research topic or known as topic discovery according to [1]. In Complex Network, researchers take the citations of research publications and model them into networks of citations. This is commonly known as *Citation Network*. Citation network is a graph where the nodes/vertices are composed of a set of documents and the links/edges are the citation relationships of documents to other documents [2]. In a citation network, the links between the cited publications and the publication that cites them are directed and not reciprocals. Directed means a publication cites older publications and an older publication cannot cite a newer publication. Networks with directed links, such as citation network, are called directed networks.

For computer science publications, there is a web application that serves as an open bibliographic information on major computer science journals and proceedings. This web application is called DBLP and the URL is: http://dblp.uni-trier.de. DBLP is a joint service provided by University of Trier and Schloss Dagstuhl [3]. A web site called *Aminer* has kindly pre-processed the DBLP data and format them into a citation network. The resulted data is available at this URL: https://aminer.org/citation [4]. In our work, we analyzed the community structure of the citation network from this data set using a community detection algorithm named LinkRank. We also observed the members of the identified communities to identify whether the publications within those communities are homogeneous or heterogeneous.

This paper is organized as follows. Section 2 describes literature reviews that we have conducted to several academic conference papers, journal papers and textbooks. Section 3 provides the research methodology that we used to work on this research. Section 4 shows the result of our analysis. Section 5 concludes our work.

2. Literature review

This section describes the literature reviews of foundation theories and related topics of our research. There are seven topics that we reviewed. These seven topics are: (a) Social Network Analysis, (b) Network-based Unsupervised Learning, (c) Community Detection, (d) Modularity, (e) PageRank Algorithm, (f) LinkRank Algorithm, and (g) Pajek Data Format. Each of the seven topics is described in the following paragraphs.

Social Network Analysis is a field that takes social relationships in the form of network of nodes and edges. The nodes represent actors and the edges represent relationships between those actors [5]. According to [6], Social Network Analysis is part of a model development, specification and testing process for conveying relationships theoretically by purveying formal definition and measurement for expressing structured output. The goal of Social Network Analysis is to find ways of analyzing network data to reveal the structure of the underlying communities that they represent [7]. According to [8], networks in general have four common properties: (a) Small-World Property, (b) Power-Law Distribution Property, (c) Network Transitivity Property, and (d) Community Structure Property. This paper is about the community structure property. The community structure property implies there are subsets of vertices in which within the subsets, the vertex-to-vertex connections are *dense* but between the subsets are less dense [8].

Unsupervised learning is process of learning patterns from a given data which has no prior knowledge about its structure and information [9]. According to [9], two classical example of unsupervised learning are clustering and dimensionality reduction. Clustering is one of the main topics in unsupervised learning and it can be called Community Detection for clustering data in which the data is in the form of network [10].

Communities are sets of nodes in a network which share common characteristics or similar properties [11]. *Community detection* is a common problem in graph data analysis in which the goal is to find sets of tightly connected nodes that are loosely connected with other nodes outside of those sets [12]. Santo Fortunato in [13] also stated that community detection's goal in graph is to identify modules and hierarchical organizations based only on the topological information of the network. Community structure is an important characteristic of networks because it is the key to learn about the complex network topology, to understand the network functions, to find hidden pattern, to do link prediction and to expand the detection [1][14]. The problem of detecting communities can be formalized as an optimization problem by defining and optimizing an appropriate criterion function that catches the intuitive concept of community [14]. According to [15], there are four categories in finding communities of a directed network. Those four categories are: (a) naive graph transformation, (b) transformations maintaining directionality, (c) extending clustering objective functions and methodologies to directed network, and (d) alternative approaches.

Modularity is a quality function used to find the best communities of a network [13]. Modularity is one of the most sensitive detection methods through optimization of the quality function over the possible divisions of a network in which the direct application of this method using, for instance, simulated annealing is computationally costly [5]. In general, quality function works by partitioning graph by giving a score to each graph partition. These scores are used to determine the best graph partition. The calculation of finding communities in general can be done by removing edges out of the communities which the network has been partitioned by the expected removal of edges.

PageRank is a network centrality measurement algorithm created by Sergey Brin and Lawrence Page [16]. This algorithm gives scores to webpages in the Internet from the highest to the lowest score. According to [17], the definition of PageRank algorithm has another intuitive basis in random walks on graphs. A webpage would have a high PageRank score if other webpages that refer to that webpage also have high PageRank score [16].

LinkRank is an algorithm that is related to random walks similar to PageRank algorithm. Furthermore, LinkRank algorithm indicates the importance of links in the network as the probability that a random surfer will follow this link into a stationary state [15]. The LinkRank algorithm is introduced by Kim et al [18]. The algorithm is mathematically defined in equation 1.

$$L_{ij} = \pi_i G_{ij} \tag{1}$$

$$G_{ij} = \alpha \frac{w_{ij}}{w_i^{out}} + \frac{1}{N}(\alpha a_i + 1 - \alpha) \tag{2}$$

where π_i is the *ith* element of PageRank vector π, and G_{ij} is the element of a Google Matrix G (equation 2). In equation 2, the $(1 - \alpha)$ is the teleportation probability where the random walker stops following the edges and move to a random node, a_i is equal to one if and only if node i is a dangling node; otherwise a_i is zero. The value of w_{ij}/w_i^{out} is set to zero when $w_i^{out} = 0$. In LinkRank algorithm, Kim et al [18] established a new definition of community and modularity. The community is defined as a group of nodes in which a random walker is more likely to stay. The new definition of modularity is defined as in equation 3 and equation 4.

$$Q^{lr} = (\textit{fraction of time spent walking within communities by a random walker}) - (\textit{expected value of this fraction}) \tag{3}$$

$$Q^{lr} = \sum_{i,j} L_{i,j}\delta_{c_i c j} - \sum_{i,j} E(L_{i,j})\delta_{c_i c j} \tag{4}$$

Equation 4 shows how LinkRank algorithm for community detection works. The L_{ij} states the probability of a random walker following the link from i to j. In equation 4, the first term is the fraction of time spent by a random walker walking within communities and the second term is the expected value of that fraction [18].

Pajek is a program for analysis and visualization of large networks [19]. This program has specific data format that can be read by the program. This data format aptly also named Pajek. This data format can represent a network as defined by a graph theory: list of nodes and list of edges where edges will have some values [20].

3. Research methodology

This section describes the research methodology that we used in this research. Our research methodology has eight stages as you can see in Figure 1. The research starts with *Problem Definition* activity by acknowledging the existence of an abundant research publications. Within each publication, the paper cites other previous publications that are related or part of the foundations of the work done in that paper. The paper citations can be represented as a network which is called citation network. The problem that is the focus of our research is understanding the structure of the citation network by detecting communities within the network. In the preparation to embark a process of detecting communities within the citation network, we conducted *Literature Review* activity as the second activity in this research. We gathered and reviewed several research publications about social network analysis, especially that are related to citation network, and community detection methods for finding communities in the network.

Following the literature review, we proceed to the next activity that is the *Data Collection and Processing*. We used a citation network dataset provided by a website called Aminer [4]. The dataset is filtered by focusing on papers published in 2015 and have citation relationships. We filtered the data due to the limitation of our computing hardware. In

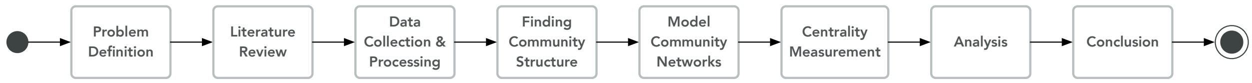

Fig. 1. Research methodology

our filtered data, the citation network contains 188,514 nodes or publications, 198,263 edges or citation relationships and it is a directed network. Following the filtering task, we reformatted the data in a format that we can use in our program that is the Pajek format [19].

Once the data is ready, we conducted the next activity called *Finding Community Structure*. We used LinkRank algorithm, a community detection algorithm introduced by Kim et al [18], because it can be used on directed network. Once the communities are found, we moved to the next activity called *Model Community Networks*. In this activity, we build new networks based on the communities detected by the algorithm. After the new networks are built, we progressed to the next activity called *Centrality Measurement*. In measuring the centrality of each community, we used PageRank algorithm. Once the PageRank centrality score is calculated, we sort the score from highest to the lowest and take the top 10 from each community for ground truth testing. This takes us to the next research methodology activity called *Analysis*. We analysis the result of previous activities (the community detection, new network model, and PageRank centrality calculation). Finally, we concluded the result of our analysis in an activity called *Conclusion*.

4. Result

This section describes the result of our work. The following subsections describe the result of finding community structure using LinkRank algorithm, which includes building new networks based on the communities found by the algorithm, centrality measurement of the new communities, and the ground-truth testing of our work.

4.1. Community detection result

In this subsection, we describe the result of using LinkRank algorithm in finding community structure of our citation network dataset. At first, we ran the algorithm from the original source code given by the creator of the algorithm (Kim et al [18]). We had performance issue in running the code that is the length of calculations went to several days. We decided to terminate the running program. To alleviate this issue, we had to analyze the original source code and modified it. Based on our analysis, the cause of the performance issue was the number of iterations of random walks to determine the community. Hence, we had to decreased the number of iterations.

In modifying the code, we end up porting the original code from C-based programming language to Python-based programming language which more familiar to us and enables us to take advantage of Python library called NetworkX for further analysis. Besides porting the program, we also change the representation of the Google Matrix in the original code from using adjacency matrix to adjacency list. We tested the accuracy of our ported program with the original program using a random synthetic network with 100 nodes and 100 edges. In terms of the duration of program to run, we found that the original program took 23,884,247 milliseconds or around 6.6 hours but our program took only 478 milliseconds to finish. The result showed no decreasing accuracy between the original program and our program. Thus, we ran our program against the citation network dataset. The change from using adjacency matrix to adjacency list showed significant effect on size of data to be processed in the memory. In the original program, the Google Matrix of $n \times n$, where n is the number of nodes, is represented as an array. For our dataset with 188,514 nodes, it took an array with the size of 35,537,528,196 which requires a memory with the size of 142.1 gigabyte (with an assumption of each pointer has the size of 4 bytes). Opposite to this, by changing the adjacency matrix to adjacency list, the size of Google Matrix equals to number of edges and the required memory for storing this also equals to number of edges. In our dataset with 198,263 edges requires around only 795 kilobytes. These modifications are believed to be part of our contributions in this work.

The result of running the LinkRank algorithm using our program to the citation network dataset, the algorithm is able to detect 10,442 communities from 188,514 nodes of the citation network. Within these communities, we found that the range of number of nodes in most communities are from 2 to 24 nodes except for one community having 34,848 nodes. Further analysis on this anomaly showed that the cause might be the result of the reduction of the number of iterations for finding communities in our program. When we reduced the number of iterations, the modularity process of giving score to the communities is not optimal. Thus, the graph partitioning process stops when the number of iterations reaches the maximum which leaves a community with large number of nodes because it cannot be partitioned further.

With the communities that we have identified, we took three communities with the largest number of node inside the community as samples and named them, respectively, *Community A*, *Community B*, and *Community C*. *Community A* has 34,848 nodes, *Community B* has 24 nodes, and *Community C* has 23 nodes. Within each of the community, we calculated the PageRank score of each node. Next, we took the top 10 node with the highest PageRank score in each community as samples. Each top 10 nodes with their PageRank score from each community can be seen in Table 1, Table 2, and Table 3.

Table 1. Top 10 nodes with their PageRank score in Community A

No.	Node ID	PageRank Score
1	108763	9.996488169282493e-05
2	150813	9.996488169282493e-05
3	169166	9.756680720959117e-05
4	150756	9.742739202770912e-05
5	114452	9.709345070997364e-05
6	5180	9.68841370830814e-05
7	63907	9.684308713512753e-05
8	107706	9.665648445529386e-05
9	7010	9.583252002249037e-05
10	77649	9.578358562107545e-05

Table 2. Top 10 nodes with their PageRank score in Community B

No.	Node ID	PageRank Score
1	11906	0.03225806451612903
2	22532	0.03225806451612903
3	1289	0.04707327012331479
4	183052	0.03225806451612903
5	150294	0.03225806451612903
6	38168	0.03225806451612903
7	22047	0.03225806451612903
8	42529	0.03225806451612903
9	174755	0.03225806451612903
10	659598	0.03225806451612903

Table 1 shows that the nodes have diverse PageRank scores but those scores are very small. In contrast to Table 1, the nodes in Table 2 and Table 3 have a more saturated and identical PageRank scores and those scores are bigger than the scores in Table 1. Our analysis concluded the reason behind this is due to the complexity of the network of Community A. Community A has 34,848 nodes with complex citation relationship. This is the opposite of Community B and Community C where they have 24 nodes and 23 nodes respectively with a simpler citation relationship. The complexity of the relationship may affect the PageRank score, because the process of giving the PageRank score is based on the citation relationship. Next, we analyze each of the community sample to see the topic of each publication to get an understanding whether we have a homogenous community or rather a heterogenous community. This analysis is part of the ground-truth checking process which is described in the next sub-section.

4.2. Ground-truth checking

Ground-truth checking is a task that checks similarity of nodes within a community. The similarity of the nodes that we check is the similarity of the publication's topic. The ground-truth check was conducted manually. For each

Table 3. Top 10 nodes with their PageRank score in Community C

No.	Node ID	PageRank Score
1	24067	0.03125
2	1686	0.03125
3	10631	0.03125
4	19083	0.03125
5	28690	0.03125
6	129300	0.03125
7	117910	0.03125
8	3247	0.03125
9	69022	0.03125
10	677	0.03125

publication, we checked its publisher's websites and glanced through the publication to identify the topic. We did the ground-truth check to the 10 nodes samples of the three communities in Table 1, Table 2, and Table 3. The result of the ground-truth checking can be seen in Table 4, Table 5, and Table 6.

Table 4. Ground-truth check result on community A

No.	Node ID	Publication Title	Topic
1	108763	MultiClust special issue on discovering, summarizing and using multiple clusterings	Machine learning
2	150813	Collective problem solving: Features and affordances of creative online communities	Collaborative Online, Collective Intelligence
3	169166	Exact Algorithms for Intervalizing Coloured Graphs	Graph algorithm, Exact algorithm
4	150756	Mudslinging and Manners: Unpacking Conflict in Free and Open Source Software.	Virtual teams, Virtual Works
5	114452	Level-Up; Motorized Stilts that Simulate Stair Steps in Virtual Reality	Virtual Reality, Algorithm
6	5180	Latent Dirichlet Allocation	Machine Learning
7	63907	Gamifying Research: Strategies, Opportunities, Challenges, Ethics	Gamification, Performance
8	107706	Image re-ranking with an alternating optimization	Visual word selection, Image re-ranking, Alternating optimization
9	7010	Distinctive Image Features from Scale-Invariant Keypoints	Image Matching, Object Recognition
10	77649	An Evaluation of Tail Loss Recovery Mechanism for TCP	Networks, Congestion Control

Table 4 shows the top 10 nodes and their topics for Community A. It shows most of the topics have similarities around machine learning, image processing and visualization. Nevertheless, we believed there are three publications derailed from those similar topics. Node 114452 has virtual reality topic, node 63907 has gamification and performance topic and node 77649 has networks and congestion control topic.

Table 5 shows the top 10 nodes and their topics for Community B. It shows most of the topics have similarity around artificial intelligence, image processing and machine learning. There are two publications that have a topic about software engineering and logic which are far from the topic mentioned before. There are two nodes that represent proceedings instead of single publications (node 38168 and node 42529). Node 42529 is a proceeding about data science and management which is slightly out of the topic. There two nodes (node 22532 and node 1298) that represent the name of conferences but not single publications.

Table 6 shows the top 10 nodes and their topics for Community C. It shows there two common themes: image processing and data processing. Other less common topics are HCI, and data encryption. Six nodes (node 1686, node 10631, node 19083, node 117910, node 3247 and node 677) represent conferences but not single publications.

Based on the result of ground-truth check, Community A, B and C through their samples show reasonable similarities in terms of the topics within each community, that is 70% of the nodes are similar while the other 30% are varies.

Table 5. Ground-truth check result on community B

No.	Node ID	Publication Title	Topic
1	11906	Proceedings of the twenty Fourth International Joint Conference on Artificial Intelligence.	Artificial Intelligence
2	22532	11th IEEE International Conference on e-Science	Computer Science
3	1289	INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association	Technology
4	183052	Robust Object Tracking Enhanced by Correction Dictionary	Machine Learning, Tracking
5	150294	On Quantitative Modelling and Verification of DNA Walker Circuits using Stochastic Petri Nets	Programming, Software engineering, logic and program
6	38168	Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia	Multimedia, Information retrieval, Machine learning
7	22047	Application and Theory of petri nets and Concurrency	Software engineering, logic and meaning of program
8	42529	Proceedings of the Third International Conference on Information technology and Quantitative management	Data Science, management
9	174755	Experiments on Behavioral Coordinated Control of an Unmanned Aerial Vehicle manipulator System	Robotics, Artificial Intelligence, machine Learning
10	659598	Continuous Arousal Self-Assessment Validating Using Real-Time Physiological Responses	Multimedia, Image Processing, Machine Learning

Table 6. Ground-truth check result on community C

No.	Node ID	Publication Title	Topic
1	24067	Detecting Depression of Cancer Patients with Daily Mental Health Logs from Mobile Applications	Image Processing, Clustering, Mobile
2	1686	19th International Conference in knowledge Based and Intelligent Information and Engineering System	Technology, Computer Science
3	10631	Human-Computer-Interaction: Design and Evaluation 17th International Conference	HCI, UX, UI
4	19083	2016 International Conference on Electronics, Communications, and Computers	Technology
5	28690	Development of Emotion-Weather Maps	Information Visualization, Data Processing, Emotion
6	129300	Profiling and Tracking a Cyberlocker Link Sharer in a Public Web Forum	Piracy, Data Encryption
7	117910	Advances in Digital Forensics XI 11th IFIP WG 11.9 International Conference	Data Security, Data Encryption
8	3247	13th International Conference on Document Analyst and Recognition	Graphics Analysis, Character and Symbol Recognition
9	69022	Exploiting Visual Similarities for Ontology Alignment	Image Processing, Visual Similarity
10	677	31st IEEE International Conference on Data Engineering, ICDE	Data Processing, Data Engineering

Thus the homophily property does not reach 100%. Although, we thought that they would be tightly knit around one similar topic, especially with the communities that have smaller number of nodes (Community B and C) but it does not seem to be this way. Our conclusion that LinkRank algorithm is useful to find communities of directed network, such as citation network.

5. Conclusion

Our work on finding communities using LinkRank algorithm in a citation network is successful. Based on the top 10 samples, 70% of the nodes within a community are similar while the other 30% are varies. Thus, the homophily property does not reach 100%. We ported the LinkRank algorithm from the original source code in C programming language to Python programming language, reduced the number of iterations of the random walk and change the representation of the Google Matrix from using adjacency matrix to adjacency list that reduced the size of data to be calculated. The reduction of the number of iterations affect the modularity process of giving score to the communities which may leave a community with a large number of nodes that can not be partitioned further. We reduced the number of iterations due to limited performance of the computing hardware that we used.

6. Future works

There are several some improvements that we could do in the future for this research. First, we could use different dataset where the size is larger and the citation relationships are more complex. Furthermore, dataset which has ground-truth information about communities formed in that dataset could also be beneficial for validating our study. Second, the LinkRank algorithm in our code can be optimized by experimenting on the number of random walk iterations and changing the graph data structure into a more efficient structure. Lastly, we can combine with text mining methods to analyze the title (and possibly the content) of the publication for identifying its topic. Thus, we can improve the validation of the result of the community detection.

Acknowledgements

The authors would like to thank the Faculty of Computer Science, Universitas Indonesia for funding this publication and the accommodation for attending the Information System International Conference 2017 (ISICO 2017) at Sanur Paradise Plaza Hotel from November 6, 2017 to November 8, 2017 in Bali Indonesia.

References

[1] F. Huang, X. Li, S. Zhang, J. Zhang, J. Chen, Z. Zhai, Overlapping community detection for multimedia social networks, IEEE Transactions on Multimedia 19 (8) (2017) 1881–1893. doi:10.1109/TMM.2017.2692650.

[2] S. R. Goldberg, H. Anthony, T. S. Evans, Modelling citation networks, Scientometrics 105 (3) (2015) 1577–1604. doi:10.1007/s11192-015-1737-9.
URL `http://dx.doi.org/10.1007/s11192-015-1737-9`

[3] dblp - computer science bibliography: Welcome to dblp (05 2017) [cited 2017-05-03].
URL `http://dblp.uni-trier.de`

[4] Citation network dataset (05 2017) [cited 2017-04-05].
URL `https://aminer.org/citation`

[5] M. Newman, Networks: An Introduction, Oxford University Press, Inc., New York, NY, USA, 2010.

[6] S. Wasserman, K. Faust, Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences, Cambridge University Press, 1994.

[7] M. Newman, M. Girvan, Mixing Patterns and Community Structure in Networks, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 66–87. doi:10.1007/978-3-540-44943-0_5.
URL `http://dx.doi.org/10.1007/978-3-540-44943-0_5`

[8] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences 99 (12) (2002) 7821–7826. arXiv:http://www.pnas.org/content/99/12/7821.full.pdf, doi:10.1073/pnas.122653799.
URL `http://www.pnas.org/content/99/12/7821.abstract`

[9] Z. Ghahramani, Unsupervised Learning, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 72–112. doi:10.1007/978-3-540-28650-9_5.
URL `http://dx.doi.org/10.1007/978-3-540-28650-9_5`

[10] T. C. Silva, L. Zhao, Network-Based Unsupervised Learning, Springer International Publishing, Cham, 2016, pp. 143–180. doi:10.1007/978-3-319-17290-3_6.
URL `http://dx.doi.org/10.1007/978-3-319-17290-3_6`

[11] F. Mourchid, A. Kobbane, J. B. Othman, M. E. Koutbi, A game-theoretic approach for non-overlapping communities detection, in: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 2017, pp. 1315–1320. doi:10.1109/IWCMC.2017.7986475.

[12] S. Harenberg, G. Bello, L. Gjeltema, S. Ranshous, J. Harlalka, R. Seay, K. Padmanabhan, N. Samatova, Community detection in large-scale networks: a survey and empirical evaluation, Wiley Interdisciplinary Reviews: Computational Statistics 6 (6) (2014) 426–439. doi:10.1002/wics.1319.
URL `http://dx.doi.org/10.1002/wics.1319`

[13] S. Fortunato, Community detection in graphs, Physics Reports 486 (3â“5) (2010) 75 – 174. doi:https://doi.org/10.1016/j.physrep.2009.11.002.
URL `http://www.sciencedirect.com/science/article/pii/S0370157309002841`

[14] C. Pizzuti, Evolutionary computation for community detection in networks: a review, IEEE Transactions on Evolutionary Computation PP (99) (2017) 1–1. doi:10.1109/TEVC.2017.2737600.

[15] F. D. Malliaros, M. Vazirgiannis, Clustering and community detection in directed networks: A survey, CoRR abs/1308.0971.
URL `http://arxiv.org/abs/1308.0971`

[16] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Computer Networks and ISDN Systems 30 (1) (1998) 107 – 117. doi:http://dx.doi.org/10.1016/S0169-7552(98)00110-X.
URL `http://www.sciencedirect.com/science/article/pii/S016975529800110X`

[17] L. Page, S. Brin, R. Motwani, T. Winograd, The pagerank citation ranking: Bringing order to the web., Technical Report 1999-66, Stanford InfoLab, previous number = SIDL-WP-1999-0120 (November 1999).
URL `http://ilpubs.stanford.edu:8090/422/`

[18] Y. Kim", Finding communities in directed networks, Physical Review E 81 (1). doi:10.1103/PhysRevE.81.016103.

[19] A. Mrvar, V. Batagelj, Analysis and visualization of large networks with program package pajek, Complex Adaptive Systems Modeling 4 (1) (2016) 6. doi:10.1186/s40294-016-0017-8.
URL `http://dx.doi.org/10.1186/s40294-016-0017-8`

[20] P. Doreian, Exploratory social network analysis with pajek, w. de nooy, a. mrvar, v. batagelj. cambridge university press, new york (2005), Social Networks 28 (3) (2006) 269–274. doi:10.1016/j.socnet.2005.12.002.
URL `https://doi.org/10.1016/j.socnet.2005.12.002`