The Cache Location Selection Based on Group Betweenness Centrality Maximization

Jianfeng Guan[1(✉)], Zhiwei Yan[2], Su Yao[3], Changqiao Xu[1], and Hongke Zhang[1,3]

[1] State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
{jfguan, cqxu, hkzhang}@bupt.edu.cn

[2] China Internet Network Information Center, Beijing 100190, China
yan@cnnic.cn

[3] National Engineering Laboratory for Next Generation Internet Interconnection Devices, Beijing Jiaotong University, Beijing 100044, China
bjtuyaosu@163.com

Abstract. Content-Centric Networking (CCN) as a content-oriented network architecture can provide efficient content delivery via its in-network caching. However, it is not optimal way to cache contents at all intermediate routers for that the current technology is not yet ready to support an Internet scale deployment. Therefore, in this paper we study the cache location selection problem with an objective to maximize cache delivery performance while minimize the cache nodes. The existing work select cache location based on the important of single node rather than that of entire group, which may result in inefficient problem caused by reduplicative impertinences. Therefore in this paper, we adopt group centrality especially Group Betweenness Centrality (GBC) to select cache locations. To evaluate its performance, we simulate CCN caching under different topologies, and the final results show that GBC-based scheme can provide better performance than others in term of average hop of content delivery.

Keywords: CCN · Cache · Group betweenness centrality

1 Introduction

With the booming of various network technologies, the Internet usage has gradually shifted from resource sharing to content dissemination and retrieval. According to the recent Cisco visual networking report, the video services have already occupied 40% of today's traffic, and it will reach over 60% by the end of 2015. Internet has evolved from a network connecting pairs of end-hosts to a substrate for information dissemination. As a result, the traditional end-point centric model seems to no longer cater current communication demands [1]. Therefore, many systems introduce caching mechanisms [2, 3] as a means to reduce load on access links and shorten the selected content accessing time to acquire better performance. However, the end-to-end design pattern of the current Internet is inefficient to provide these content delivery services. As a

J.-H. Lee and S. Pack (Eds.): QShine 2016, LNICST 199, pp. 269–279, 2017.
DOI: 10.1007/978-3-319-60717-7_27

result, several network architectures are emerging recently. As a promising network architecture, Content Centric Networking (CCN) has got more studied for that it can provide better service suited to today's usage including the mobility, content distribution and more resilient to disruptions and failures [4]. Different to traditional Internet, CCN treats content as primitive and uses new approaches to routing the named content similar to TRIAD [5] and DONA [6]. CCN adopts receiver-driven transport mode in which data are only transmitted in response to content requests expressed by users. *Interest* message is used by end user to express its interest of content identified by content name. While the *Data* message is response for the *Interest* message if it stratifies the uses' Interest. CCN introduces the *Content Store* which is same as the buffer of an IP router but it has a different replacement policy. Each CCN packet is self-identifying and self-authenticating. In practical, CCN adopts the Least Recently Used (LRU) or Least Frequently Used (LFU) replacement policy to maximize the probability of sharing, which minimizes upstream bandwidth demand and downstream latency to store the *Data* packets as long as possible [7].

It is obviously that the cache mechanism has an import impact on performance, and it therefore has been got researched in the context of performance measurement [8], analytical models [9], and energy impacts [10]. Different to traditional web cache [3], CCN caches the very small data chunks (typically packet-size) instead of caching full objects, which can be identified by users (named data chunks). Each router in the network will cache the data chunk and send back to users once an interest packet hits the cache [11]. However, it is not optimal to cache the chunks at all intermediate routers in CCN [12] for that it may introduce large additional deployment cost. Besides, according to the recent research [13], today's technology is not yet ready to support an Internet scale deployment of CCN at a Content Distribution Network (CDN) and ISP scale. So, there will be a long transition period in which the CCN and current Internet will coexist. As a result, the CCN routers will be deployed in selected locations of Internet and cache the heterogeneous contents. Our previous work has studied the content selection problem [14], while in this paper we mainly focus on the cache location selection problem. More specifically, we consider a scenario which is not every CCN router caching the content. Our previous work [15] has proved that Betweenness Centrality has the better performance than others, and it can be used as a metric to select the cache locations of CCN routers, while in this paper we adopt Group Betweenness Centrality (GBC) to further improve its performance.

The main contributions of this paper are: (1) study CCN router deployment problem during the transition period; (2) propose a GBC-based scheme to select the cache locations, which can maximize the cache delivery performance while minimize the number of participated CCN routers; (3) Evaluate and compare the average hop of content delivery under different network topologies and different network centralities. The rest of this paper is organized as follows. Section 2 investigates the related work of CCN caching mechanisms. Section 3 presents the related research in terms of different network centralities. Section 4 evaluates the performance under different scenarios. Finally, Sect. 5 concludes this paper.

2 Related Work

Some CCN caching schemes are proposed recently [16], and they mainly consist of location selection and content selection. The content selection schemes are generally based on popularity [17–21], priority [22], content relationship [23], user characters of request and distribution [24, 25]. As for cache locations selection or placement, some works have been done in CDN [26–28], web services [29].

The location selection is the well-known p-median or k-center problem, and it is similar to the facility location problem (FLP). However, it is different to solve it for that it is a NP-hard problem, and most researches are focused on better approximation algorithms [34], only few topologies such as line and ring can get the optimal solution in polynomial time [29]. As for CCN, the existing location selection schemes mainly depend on node importance [15, 30–33], node capability [35, 36], node attribution [37]. In term of node importance, Rossi and Rossini [30] adopted the graph-related centrality metrics (e.g., betweenness, closeness, stress) to allocate content store heterogeneously across the CCN, and got that the simple metric such as degree centrality can get modest cache hit gain under different topologies. Guan *et al.* [15] compared the performance of different centralities and found that betweenness centrality has better performance. Wang *et al.* [31, 32] proposed an optimal solution for cache allocation, and comprehensively evaluated impacts of topology character, content characteristic and replacement strategies. Cui *et al.* [33] proposed a cache allocation scheme based on the Request Influence Degree (RID).

However, the existing works are mainly based on single node's importance while little consider the importance of the given cache group. So in this paper, we just consider the topology property, and propose a GBC-based cache location selection scheme to decide the number of deployable CCN router during the transition period.

3 The Proposed Solution

In this section, we first investigate the network centrality, and then describe our GBC-based scheme.

3.1 Node Network Centrality

Network centrality has a long tradition in the analysis of networks, and it is a structural attribute used to measure the contribution of node. There are various types of measures of the centrality of a node to determine the relative importance of a node in the network including centralities of degree, closeness, betweenness and information. For a given network $G = (V, E)$, several typical measures are shown as follows.

(1) Degree Centrality (DC)

The DC of a node v is defined as

$$C_D = \deg(v) \tag{1}$$

Where $deg(v)$ is the number of links incident on node v. For a direct graph, it includes the in-degree and out-degree.

(2) Closeness Centrality (CC)

CC is used to measure the distance of node *v* to all the other nodes in the network, which can be defined as

$$C_{CC} = \frac{1}{\sum_{\forall s \in V \backslash v} d(v,s)} \tag{2}$$

Where the *d*(*v*, *s*) is the shortest path length from node *v* to node *s*. If the graph is not completely connected, this algorithm computes the closeness centrality for each connected part separately.

(3) Betweenness Centrality (BC)

BC reflects how often the node *v* locates on the shortest paths, and it is defined as

$$C_{BC} = \sum_{s \neq v \neq t \in V} \frac{\delta_{st}(v)}{\delta_{st}} \tag{3}$$

Where δ_{st} is total number of shortest paths from node *s* to node *t*, and $\delta_{st}(v)$ is the number of those paths that pass through the node *v*.

(4) Eigenvector Centrality (EC)

EC assigns relative scores to all nodes based on the principle that connections to high-scoring nodes contribute more to the score of node than equal connections to low-scoring nodes. In particular, Google's PageRank is a variant. The definition is shown as follows.

$$C_{EC}(v) = \frac{1}{\lambda} \sum_{t \in M(v)} C_E(t) \tag{4}$$

Where *M(v)* is a set of the neighbors of *v*, and λ is a constant.

(5) Load Centrality (LC)

LC of a node is used to measure the load on each node. If all the traffic flows transmit along the shortest paths, then BC and LC are equivalent.

(6) Subgraph Centrality (SC)

SC of a node *n* is the sum of closed walks of all lengths starting and ending at node *n*, which can be expressed as

$$C_{SC}(u) = \sum_{i=1}^{N} (v_i^u)^2 e^{\lambda_i} \tag{5}$$

Where v_i is an eigenvector of adjacency matrix *A* of *G* corresponding to the eigenvalue λ_i. The communicability centrality of a node can be found using the matrix exponential of the adjacency matrix of *G* [38].

3.2 Group Network Centrality

The node centrality just reflect the importance of a node, while in some applications, the centrality of a group is more attractive. For example, in social networks, Borgatti [39] proposes key player problem which includes the Key Player Problem/Positive (KPP-POS) and Key Player Problem/Negative (KPP-NEG). KPP-POS is used to identify the key players for purpose of optimally diffusing something through the network, while the KPP-NEG is defined to identify the key players to disrupt the network by removing the key nodes. As for Internet, groups of routers or links that has maximal potential to control over traffic to increase the effectiveness of network measurements or intrusion detection [40]. So, the group network centrality can be applied to select the cache location. Everett and Borgatti [41] defined GBC as a natural extension of the betweenness measure, which is used to estimate the influence of a group of nodes over the information flow in the network. Some research has shown that finding a group with maximal GBC is a NP-hard problem. Therefore, Puzis *et al*. [42, 43] propose a method for rapid computation of group betweenness centrality to locate the most prominent group of nodes in a network.

Let $S \subseteq V$ be a group of nodes, the GBC(S) stands for the group betweenness centrality, which can be expressed as

$$GBC(S) = \sum_{s,t \in V | s \neq t \in V} \frac{\ddot{\bar{\delta}}_{s,t}(S)}{\delta_{s,t}} \tag{6}$$

Where $\ddot{\bar{\delta}}_{s,t}(S)$ is the number of shortest paths between s and t that traverse at least one member of the group S. The centrality of group is not simply the sum of centralities of its members. Ishakian *et al*. [44] define a generic algorithm for computing the generalized centrality measure for every node and every group of nodes in the network to identify the subset of a given network that has the largest group centrality, which is called as *K-Group Centrality Maximization* (k-GCM) problem. So, in this paper, we adopt the group between centrality maximization as a metric to select the prominent group in the network to deploy the CCN caches.

4 Performance Evaluation

4.1 Evaluation Metric

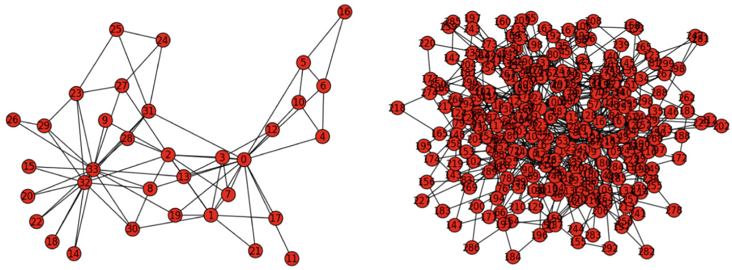
The objectives of caching are to lower content delivery latency, reduce the traffic and congestion and alleviate server load. In the performance evaluation, we adopt the average hop of content delivery. Assuming that the network topology consists of N nodes, and we deploy M CCN routers in the network, and the cache of each CCN router has the same capability, and the average hop of content delivery is defined as

$$D = E(d(N, S)) = \frac{1}{N}\sum_{i=1}^{N} d(n_i, S) \tag{7}$$

Where $d(n_i, S)$ denotes the shortest path between node n_i and a CCN routers set S. The average hop of content delivery will reduce with the increase of the number of deployed CCN routers M obviously. The extreme case ($M = N$) is that all the routers in the networks support the CCN.

4.2 Methodology

We use the Networkx [45] and MATLAB to analyze the performance, and we implemented the GBC maximization algorithm in C++ to select the prominent group of nodes for the given network and the given group size. In the analysis, we adopt three networks to evaluate its applicability including the Zachary's Karate Club (ZKC) [46], Barabási-Albert (BA) network and scale-free network. These topologies can be used to represent the social network, random graph and Internet, respectively. Figure 1 shows the ZKC topology and BA topology, respectively.



(a) ZKC network (b) BA network

Fig. 1. Demonstration of ZKC and BA topologies

Table 1 shows the main characteristics of each network including the number of node and edge, average degree.

The users' requests follow a uniform distribution and are generated in each node of the test topologies.

Table 1. The information of selected topology types

Network types	Number of nodes	Number of edges	Average degree
ZKC	34	78	4.5882
BA	300	596	3.9773
Scale-free	500	1489	5.9560

4.3 Results

In the evaluation, we select a prominent group with size from 1 to 10, and compare the GBC with the Degree Centrality (DC), Closeness Centrality (CC), Betweenness Centrality (BC), Eigenvector Centrality (EC), Load Centrality (LC), Subgraph Centrality (Noted it as SC).

(1) ZKC Network

Figure 2 shows the average hop of content delivery in ZKC network. We can get that the average hop decrease greatly with the increase of the number of CCN routers. To describe it more detail, we adopt the relative value to show the differences in Fig. 2(a). And Fig. 2(b) compares the BC and GBC. We can get that in this small social topology, the BC is more attractive that other centralities.

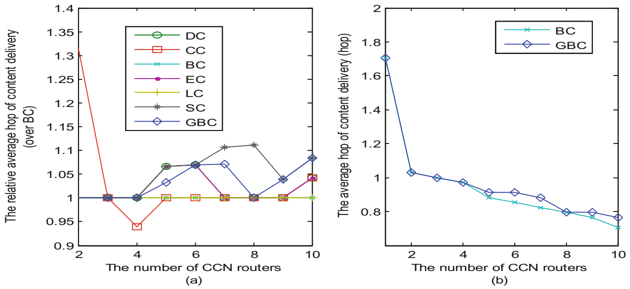


Fig. 2. Comparison of BC and GBC under ZKC network

(2) BA Network

Figure 3 shows the relative average hop over BC in Fig. 3(a), and more detailed comparison between BC and GBC in Fig. 3(b). We can get that the GBC has the smallest average hop than the others, which shows that GBC can choose the prominent group in the network to get the best performance.

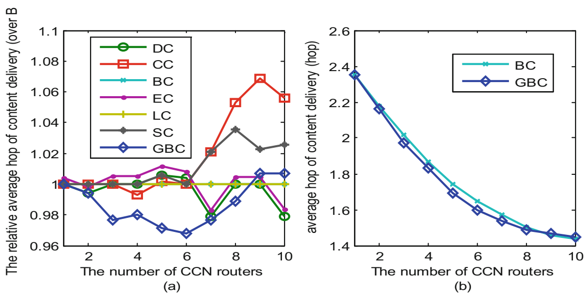


Fig. 3. Comparison of BC and GBC under BA network

(3) **Scale-Free Network**

Figure 4 shows the relative average hop over BC in Fig. 4(a), and more detailed comparison between BC and GBC in Fig. 4(b). Similarly, we can get that the GBC has the smallest average hop.

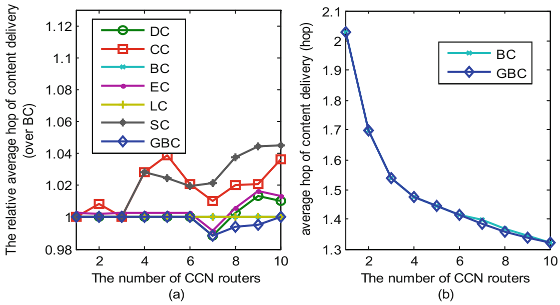


Fig. 4. Comparison of BC and GBC under Scal-free network

Table 2 shows the cache nodes list of the prominent group under different networks based on BC and GBC (group size is 10).

Table 2. Cache locations list under different topology

Network types	Centrality	Nodes list
ZKC	BC	0 33 32 2 31 8 1 13 19 5
	GBC	0 33 32 2 5 0 6 31 0 27
BA	BC	2 0 3 8 14 7 17 22 26 4
	GBC	2 3 8 14 17 22 26 19 11 28
Scale-free	BC	1 2 3 9 8 27 4 20 44 17
	GBC	1 2 3 9 8 27 20 44 17 30

Based on the above information, we can get the cache location to deploy the CCN routers in the transition period.

5 Conclusion

In this paper, we study cache location selection problem of CCN during the transition period. We investigate the existed cache schemes and different network centralities, and propose a GBC-based scheme to choose the prominent group of CCN router in the network. The simulation shows that compared with other network centralities, the proposed scheme can maximize the cache delivery performance in the same group size. Our future work is to set up more complicated network models and user models to evaluate the different user behaviors and service behaviors to further study the cache performances.

Acknowledgments. This work is partially supported by the National Basic Research Program of China (973 Program) under Grant No. 2013CB329102, and the National Natural Science Foundation of China (NSFC) under Grant Nos. 61003283, 61232017, and 61303242.

References

1. Katsaros, K., Xylomenos, G., Polyzos, G.C.: MultiCache: an overlay architecture for information-centric networking. Comput. Netw. **55**(4), 936–947 (2011)
2. Che, H., Wang, Z., Tung, Y.: Analysis and design of hierarchical web caching systems. In: INFOCOM 2001, pp. 1416–1424 (2001)
3. Podlipnig, S., Böszörmenyi, L.: A survey of web cache replacement strategies. ACM Comput. Surv. **35**(4), 374–398 (2003)
4. Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., Ohlman, B.: A survey of information-centric networking. IEEE Commun. Mag. **50**(7), 26–36 (2011)
5. Cheriton, D., Gritter, M.: TRIAD: a new next-generation internet architecture. Stanford Technical report (2000)
6. Koponen, T., Chawla, M., Chun, B.-G., Ermolinskiy, A., Kim, K.H., Shenker, S., Stoica, I.: A data-oriented (and beyond) network architecture. SIGCOMM **2007**, 181–192 (2007)
7. Choi, J., Han, J., Cho, E., Kwon, T.T., Choi, Y.: A survey on content-oriented networking for efficient content delivery. IEEE Commun. Mag. **49**(3), 121–127 (2011)
8. Jacobson, V., Smetters, D.K., Briggs, N.H., Plass, M.F., Stewart, P., Thornton, J.D., Braynard, R.L.: Voice over content centric network. In: ACM ReArch, pp. 1–6 (2009)
9. Rosensweig, E.J., Kurose, J., Towsley, D.: Approximate models for general cache networks. In: INFOCOM 2010 (2010)
10. Lee, V.H.U., Rimac, I.: Greening the internet with content centric networking. In: 1st International Conference on Energy-Efficient Computing and Networking (2010)
11. Jacobson, V., Smetters, D.K., Briggs, N.H., Thornton, J.D., Plass, M.F., Braynard, R.L.: Networking named content. In: ACM CoNEXT, pp. 1–9 (2009)
12. Li, Y., Lin, T., Tang, H., Sun, P.: A chunk caching location and searching scheme in content centric networking. In: ICC, pp. 2655–2659 (2012)
13. Perino, D., Varvello, M.: A reality check for content centric networking. In: SIGCOMM Workshop on ICN 2011, pp. 44–49 (2011)
14. Guan, J., He, Y., Wei, Q., Neng, Z.: A classification-based wisdom caching scheme for content centric networking. In: INFOCOM 2016 Workshop, pp. 876–881 (2016)
15. Guan, J., Quan, W., Xu, C., Zhang, H.: The location selection for CCN router based on the network centrality. In: IEEE CCIS 2012, pp. 568–582 (2012)
16. Zhang, G., Li, Y., Lin, T.: Caching in information centric network: a survey. Comput. Netw. **57**(16), 3128–3141 (2013)
17. Lim, S.-H., Ko, Y.-B., Jung, G.-H., Kim, J., Jang, M.-W.: Inter-chunk popularity-based edge-first caching in content-centric networking. IEEE Commun. Lett. **18**(8), 1331–1334 (2014)
18. Bernardini, C., Silverston, T., Festor, O.: MPC: popularity-based caching strategy for content centric networks. In: ICC 2013, pp. 3619–3623 (2013)

19. Liu, Z., Dong, M., Gu, B., Zhang, C., Ji, Y., Tanaka, Y.: Inter-domain popularity-aware video caching in future internet architectures. In: QSHINE 2015, pp. 404–409 (2015)
20. Cho, K., Lee, M., Park, K., Kwon, T.T., Choi, Y., Pack, S.: WAVE: popularity-based and collaborative in-network caching for content-oriented networks. In: INFOCOM Workshops, pp. 316–321 (2012)
21. Li, H., Nakazato, H., Detti, A., Melazzi, N.B.: Popularity proportional cache size allocation policy for video delivery on CCN. In: EuCNC 2015, pp. 434–438 (2015)
22. Tu, Y., Qiao, X., Nan, G., Chen, J., Li, S.: A priority-based dynamic web requests scheduling for web servers over content-centric networking. In: Third IEEE Workshop on Hot Topics in Web Systems and Technologies, pp. 43–48 (2015)
23. Antunes, R.S., Lehmann, M.B., Mansilha, R.B.: CCNrel: leveraging relations among objects to improve the performance of CCN. In: IFIP/IEEE International Symposium on Integrated Network Management (IM), pp. 199–206 (2015)
24. Bernardini, C., Silverston, T., Festor, O.: Socially-aware caching strategy for content centric networking. In: IFIP Networking 2014, pp. 1–9 (2014)
25. Zhou, Y., Cui, L., Jiang, Y., Xu, M.: Modeling and optimizing the cache deployment with filter effect in multi-cache system. In: ISCC 2013, pp. 561–566 (2013)
26. Chen, Z., Lin, C., Yin, H., Li, B.: On the server placement problem of P2P live media streaming system. In: Huang, Y.-M.R., Xu, C., Cheng, K.-S., Yang, J.-F.K., Swamy, M.N.S., Li, S., Ding, J.-W. (eds.) PCM 2008. LNCS, vol. 5353, pp. 178–187. Springer, Heidelberg (2008). doi:10.1007/978-3-540-89796-5_19
27. Zhang, Y., Tatipamula, M.: The freshman handbook: a hint for the service placement of social networks. In: WWW 2011, pp. 173–174 (2011)
28. Yang, D., Fang, X., Xue, G.: ESPN: efficient server placement in probabilistic networks with budget constraint. In: INFOCOM 2011, pp. 1269–1277 (2011)
29. Krishnan, P., Raz, D., Shavitt, Y.: The cache location problem. IEEE/ACM Trans. Netw. **8** (5), 568–582 (2000)
30. Rossi, D., Rossini, G.: On sizing CCN content stores by exploiting topological information. In: INFOCOM Workshop, pp. 280–285 (2012)
31. Wang, Y., Li, Z., Tyson, G., Uhlig, S., Xie, G.: Optimal cache allocation for content-centric networking. In: ICNP 2013, pp. 1–10 (2013)
32. Wang, Y., Li, Z., Tyson, G., Uhlig, S., Xie, G.: Design and evaluation of the optimal cache allocation for content-centric networking. IEEE Trans. Comput. **65**(1), 95–107 (2016)
33. Cui, X., Liu, J., Huang, T., Chen, J., Liu, Y.: The network characters: a novel metric for cache size allocation scheme in content centric networking. In: National Doctoral Academic Forum on Information and Communications Technology, pp. 1–6 (2013)
34. Dohan, D., Karp, S., Matejek, B.: K-median algorithms: theory in practice (2015). http://www.cs.princeton.edu/courses/archive/fall14/cos521/projects/kmedian.pdf
35. Mishra, G.P., Dave, M.: Cost effective caching in content centric networking. In: NGCT-2015, pp. 198–202 (2015)
36. Xu, Y., Wang, Z., Li, Y., Lin, T., An, W., Ci, S.: Minimizing bandwidth cost of CCN: a coordinated in-network caching approach. In: ICCCN 2015, pp. 1–7 (2015)
37. Sourlas, V., Tassiulas, L., Psaras, I., Pavlou, G.: Information resilience through user-assisted caching in disruptive content-centric networks. In: IFIP Networking Conference, pp. 1–9 (2015)
38. Estrada, E., Hatano, N.: Communicability in complex networks. Phys. Rev. E **77**, 036111 (2008). http://arxiv.org/abs/0707.0756
39. Borgatti, S.P.: Identifying sets of key players in a social network. Comput. Math. Organ. Theory **12**(1), 21–34 (2006)

40. Puzis, R., Tubi, M., Elovici, Y., Glezer, C., Dolev, S.: A decision support system for placement of intrusion detection and prevention devices in large-scale networks. ACM Trans. Model. Comput. Simul. **22**(1), Article ID: 5 (2011)
41. Everett, M.G., Borgatti, S.P.: The centrality of groups and classes. Math. Sociol. **23**(3), 181–201 (1999)
42. Puzis, R., Elovici, Y., Dolev, S.: Fast algorithm for successive computation of group between centrality. Phys. Rev. E **76**, 056709 (2007)
43. Puzis, R., Elovici, Y., Dolev, S.: Finding the most prominent group in complex networks. AI Commun. **20**, 287–296 (2007)
44. Ishakian, V., Erdos, D., Terzi, E., Bestavros, A.: A framework for the evaluation and management of network centrality. In: SDM 2012 (2012)
45. NetworkX. http://networkx.lanl.gov/index.html
46. Data File. http://vlado.fmf.uni-lj.si/pub/networks/data/Ucinet/UciData.htm