Chapter 14
Implementation of Social Network Analysis for Web Cache Content Mining Visualization

Sarina Sulaiman, Siti Mariyam Shamsuddin, and Ajith Abraham

Abstract A Web cache content mining is a very important part in analyzing and filtering the internet contents. Essentially, a log data will be used to identify either to cache or not to cache Web contents in a cache server. This data contains dissimilar elements consisting of URL, size, retrieval time, number of hits and other elements for Web contents. In this chapter, we propose a new method and analyses of our cache server data using social network analysis (SNA); and make a number of statistic measurements to reveal the hidden information on E-Learning@UTM (EL) and Boston University (BU) logs dataset. The log dataset was extracted by particular queries, and it was displayed as a connected graph and clustered based on a similarity of characteristics. Later, the statistical properties of dataset network were computed, including speed and complexity. The result shows the SNA important behaviors: data localization in a separate position; and centralized data in a single position approve that the concentration of data to one node. These behaviors are driven by complexity of the dataset and network nodes structure of the chosen log dataset.

Introduction

Social network analysis (SNA) is exploited widely in data mining to make a deep analysis on e-learning log data. Besides, Web caching can be classified through a log dataset that conceals an interesting behavior and hidden information. The Web

S. Sulaiman (✉) • S.M. Shamsuddin
Soft Computing Research Group, Faculty of Computer Science and Information Systems, K-Economy Research Alliance, Universiti Teknologi Malaysia, 81310 Johor, Malaysia
e-mail: sarina@utm.my; mariyam@utm.my

A. Abraham
Machine Intelligence Research Labs (MIR Labs), Scientific Network for Innovation and Research Excellence, Auburn, WA 98092, USA
e-mail: ajith.abraham@ieee.org

A. Abraham (ed.), *Computational Social Networks: Mining and Visualization*, DOI 10.1007/978-1-4471-4054-2_14,

cache content used in this research is a log data from E-Learning@UTM (EL) Web server that has been monitored for 2 days and Boston University (BU) client server for 7 months.

In a previous study, Grissa et al. [1] discovered that the interestingness measures (IM) cluster to assist a user to make a decision on a similar behavior of the dataset. Martinez et al. [2] proposed a combination of qualitative evaluation with social network analysis to investigate the social interaction inside classroom. The relationship inside E-Learning such as: discussing, solving doubts, sharing information and creating a product become parameters to be analyzed with SNA. They succeeded in detecting the model of collaborative design derived from activity inside a class. This research has a similar interest which is to find the vital relationship between variables inside a cache server. This relationship will bring great benefits especially on cache server performance, data size and memory management.

This chapter is organized as follows. Section “Related Studies” explains the related works on SNA, web caching, and formal concept analysis. Section “The Proposed Method” presents the proposed technique. Section “Experimental Results and Analysis” describes results analysis, and section “Discussion and Conclusion” discusses the findings and concluding remark.

Related Studies

Popularity of social network (SN) is growing and it contains a lot of information of people’s behavior all around the world. The popularity of SN becomes a concern for several researchers and has lead to numerous investigations. SN is implemented in multiple applications, including for analyzing terrorist networks, friendly command and control structures, Web content mining [3], e-learning [2, 4] and other applications. SNA assumes that relationships are important. Hidden information and relationship inside SNA attracted researchers to find out what actually happen during the communication between elements in SN. Martinez et al. [2] have chosen E-Learning as a domain problem for their research. They explored the process on exchanging information in their e-learning forum. Another work comes from Shi [5], who used SNA to identify the structures of community. His method starts with selecting data using query and create a model of semantic relation using a statistical approach: collocation, weighted dependence, and mutual information [5].

Conversely, SN is also able to be combined with some other artificial intelligence technique such as the ant colony to get a good analysis result [6]. Al-Fayoumi [6] proposed a clever ant colony metaphor that can make a cluster on the SN by using maximum clique and sub grouping criteria. The expansion of SN researches as a foundation of SNA becomes a popular technique today. SNA analyses the complication of a human system through mapping and classifying a relationship among people, groups or organizations [7].

On Web cache context, many researchers look at many ways to improve the existing caching techniques. Padmanabhan and Mogul [8] recommended a predictive model to be used as a server hint. The proposed model is equipped with a server that can create a markov model by predicting a probability of an object *A* that is tagged with next *n* requests and an object *B* (*n* is a parameter of the algorithm). The server will use the model to produce a forecast for subsequent references. The client will use the forecasted result to pre-fetch an object on a server for the object that is not in the client cache. The simulation results show that the latency is reduced up to 45%. However, the network traffic is even worse which is two times compared to the conventional solution [8]. Therefore, there is a need to provide a justification of why it is important to solve the issue on latency and network reduction simultaneously.

Bestavros et al. [9] presented a model for the speculative dissemination of World Wide Web data. Their work illustrates that reference patterns from a Web server can be used as the main source of information for pre-fetching. Their study shows that latency reduction increases until 50% despite the increment of the bandwidth utilisation. In addition, Pallis et al. [10] provided the solutions on pre-fetching by implementing a clustering method. The technique creates the cluster for pre-fetching into graph-based clustering that consists of correlated and directed clusters. This technique can be adapted in a proxy server to increase the performance. The simulation conducted on a real dataset implies that the technique can increase performance of the Web caching environment efficiently. Kroeger et al. [11] suggested that a local proxy caching can decrease latency up to 26%, while pre-fetching could decrease latency up to 57%. The combination of both methods will give better latency reduction until 60%. Furthermore, the study also describes that the algorithm on pre-fetching contributes to a reduction of latency, and it can provide double improvement on caching. However, it only happens when the latency is decreased.

Furthermore, Harding et al. [12] implemented pre-fetching on a mobile device by enhancing the feature for optimum utilisation. This feature allows users to connect to the server when it is required and few amounts of data are needed to be transferred. Fig. 14.1 shows that the majority of activities are setting up on the desktop, these will be inefficient since communication between desktop and mobile is linked through the internet. When a number of tasks and services are developed in desktop application, the mobile environment must be synchronised and customised. However, the method still has challenges in terms of an application design; user interaction, usability and mentality model. Ye et al. [13] proposed a model to manage Web caching in a multi-device that is connected to two wireless networks: MSS (mobile support station), and the neighbouring peers to form the P2P (peer to peer) network.

Based on previous works, we can conclude that SNA can be put into practice to visualize and manage the Web cache content. This is because SNA has been used widely to reveal hidden information, relationship and characteristic of dataset in graph representation for different cases.

Fig. 14.1 Architecture of Web cache server and mobile clients

The Proposed Method

This chapter proposes an analysis of server logs data that was obtained on 13 and 14 January 2008 from one of the E-Learning@UTM (EL) servers at the Centre of Information and Communication Technology (CICT), Universiti Teknologi Malaysia (UTM) [14]. The second logs data came from a combination of 7 months (November 1994 until May 1995) of browser logs data from Boston University (BU) [15]. The investigation is intended to find a relationship between the log data variables and find a solution to increase the performance of Web cache content mining. Previously, we developed our pre-fetching technology on a mobile context and integrated it to the cache server [16, 17]. The architecture of communication between our cache server and mobile clients can be viewed in Fig. 14.1.

In Fig. 14.1, the performance of cache server is determined by the size of data transfer, relationship between elements in the cache server and also Web caching process. The proposed method is done by selecting the potential data from the main dataset and queries based on the rules generated from classification of dataset on proxy cache [18]. Next, the visualization can be perceived through a graph. Later, SNA tool will check a relationship of data and cluster the Web contents.

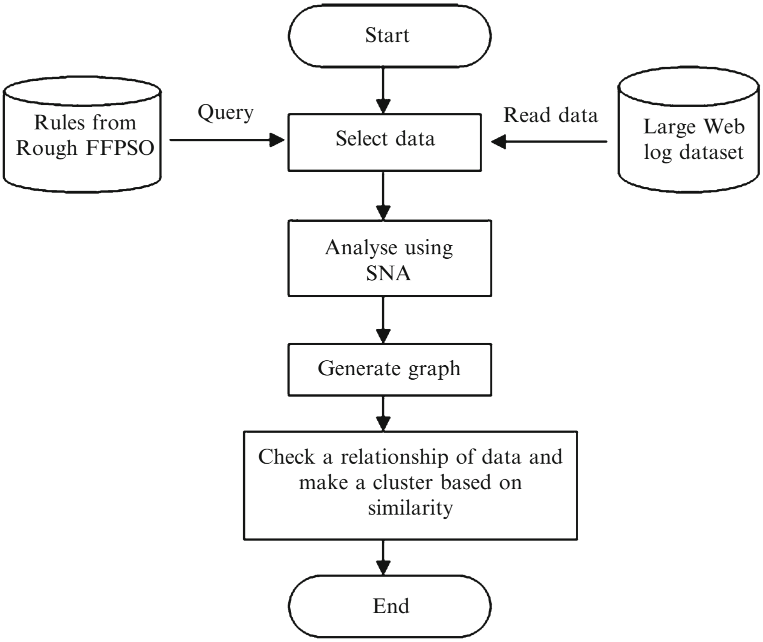

Fig. 14.2 Methodology of the proposed method

In reality, we can learn the behavior of a Web caching process and the result can be used to enhance the cache server for future development. The process of proposed methodology is illustrated in Fig. 14.2.

Experimental Result and Analysis

The data from large dataset is too complex and almost impossible to be analyzed with a standard PC. Therefore, as shown in the methodology Fig. 14.2, filtering procedure was used to obtain the data in small size but rich in terms of interaction. The filtered data was analysed using Organisational Risk Analyser (ORA) tool (http://www.casos.cs.cmu.edu). This tool is a statistical analysis package for analysing complex systems as Dynamic Social Networks. ORA generated a graph based on data relationship and clustered these data based on their node similarity. In this study, the node represents each element of dataset by using SNA similar to agent, resource or knowledge. The log data in Web cache server contain four main elements:

URL – Web content addresses of uniform resource locator.
SIZE – the size is expressed in bytes.

NUMBER_OF_HITS – the number of hits per data. Each complete request for Web content will increase the number of hits for a requested content.

CACHE – decision either to cache or not to cache the Web content.

These four elements were analyzed using SNA by checking the relationships between URL x SIZE, URL x NUMBER_OF_HITS, while cache elements were selected based on a particular scenario. There are several scenarios that discussed in this chapter. The scenario is derived by selecting the data from a main dataset, which is a very large size of data.

Each scenario contains hidden information and relationship between each element. The first and second scenario used EL dataset. At the same time, both scenarios utilized trial and error queries to identify the nature of the relationship between Web cache contents. The third scenario with three different rules applied BU dataset and made queries based on generated Rough FFPSO rules. Consequently, only the first rule used the right proposed rules for BU dataset; however, the second and third rule was queried using the proposed rules for EL dataset not for BU dataset.

First Scenario

The total data row for the main dataset with CACHE = 1 and hit number more than zero is around two thousand records. This will affect the visualization and analysis of the system due to the computational complexity. Hence, in this scenario the dataset is selected using query statement:

```
SELECT * FROM [EL Dataset]
WHERE NUM_OF_HITS > 0 AND CACHE = 1 AND SIZE > 0.00269226
AND SIZE < 1;
```

After the data selection, each column needs to be associated with node class in SNA tool:

1. URL associated as an Agent
2. SIZE associated as an Agent
3. NUM_OF_HITS as an Event
4. CACHE is not included in a process due to the cache value did not have any variation.

The selected data can be visualized in two modes: 2D and 3D graph. Figure 14.3 shows an interesting result in which the data is centralized in some location, and this is similar to cluster model.

Figures 14.3 and 14.4 depict 3D visualization of the data from the first scenario selection mode. Figure 14.3 illustrates graph positioning and zooms into one of

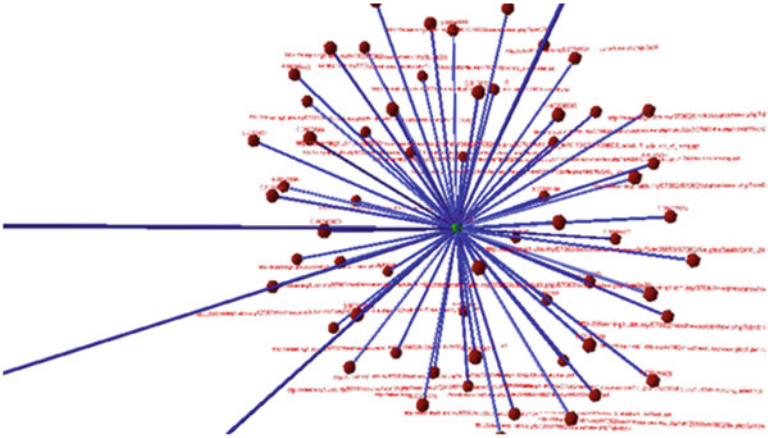

Fig. 14.3 3D visualization of first scenario – zoom in

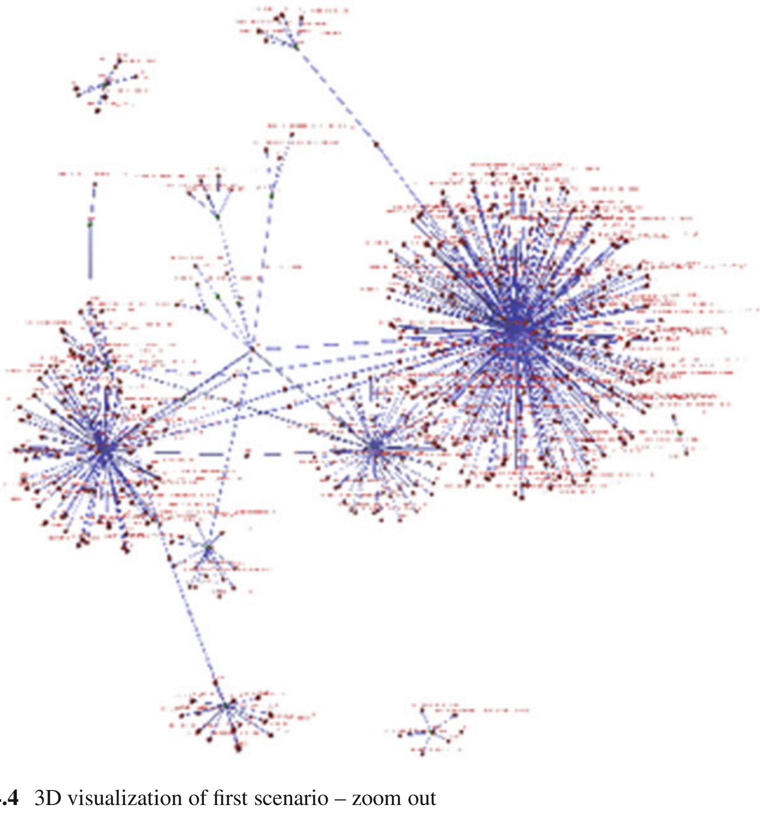

Fig. 14.4 3D visualization of first scenario – zoom out

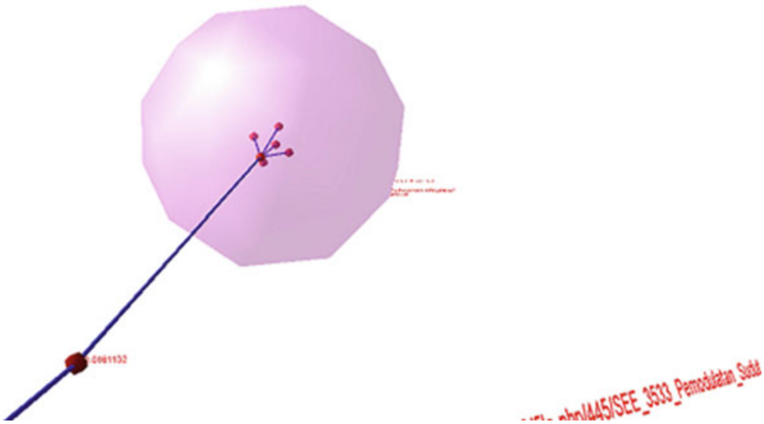

Fig. 14.5 Clustering process on one of centralized data

the groups in dataset. This shows that the log data has a lot of similarity in one value; for example, dataset "NUM_OF_HITS" with value = 0.00175439 belongs to the numbers of rows. This similarity makes the data clustered, similar to the dataset depicted in Fig. 14.4. On the other hand, Fig. 14.5 shows the clustering process to one of the centralized data. This cluster makes the separation on the graph clear and user can easily differentiate each cluster.

The graph rendering can be monitored through networks over time. This step is useful to ensure that the 3D and 2D visualization can be generated. This step can show the duration of a transition and display phase (refer to Fig. 14.6).

Figures 14.7 and 14.8 show the graph analysis of a network level of performance measurement. Figure 14.8 focuses to detect changes in the network. The speed of agent is rising linearly according to the index of agent. The count node of agent also holds similar behavior. The node is linear increasing based on the index count node of agent. Figure 14.8 depicts that counted row between agents also rises linearly until index 2, and it changes to a steady mode after index 2. On the other hand, link count is increasing rapidly on index 4 until index 5.

Second Scenario

The second scenario, the dataset is selected using query statement:

```
SELECT * FROM [EL Dataset]
WHERE NUM_OF_HITS > 0 AND CACHE = 1 AND SIZE > =0 AND
SIZE < = 0.00269226;
```

Fig. 14.6 Network monitoring

After the data selection, each column needs to be associated with node class in SNA tool:

1. URL associated as an Agent
2. SIZE associated as an Agent
3. NUM_OF_HITS as an Event
4. CACHE is not included in a process due to the cache value did not have any variation.

Total number of rows acquired from this query is 2,800 rows. These huge rows are quite hard to be visualized and analyzed. Consequently, we finally succeed to visualize this dataset and found an interesting result as shown in Figs. 14.9 and 14.10.

Figures 14.9 and 14.10 show an interesting phenomenon that happens in the second scenario. The data is pulled out to the centre of network and act like a nucleus of the dataset. It means that the dataset has a number of rows with the same "SIZE" value. Figure 14.11 illustrates this centralized effect in a 2D mode.

Figures 14.12 and 14.13 depict an analysis result of the dataset for the second scenario. Figure 14.12 shows that the values of speed for node inside the network element of this scenario dataset are very high at the beginning and drop significantly on index 2. Then, it continues to be stable until the next index. At the same time, Fig. 14.13 also shows a similar behavior where the speed of agent drops drastically when complexity starts to go up. Next, the speed of agent stays steadily on low value as long as the complexity is still in high value.

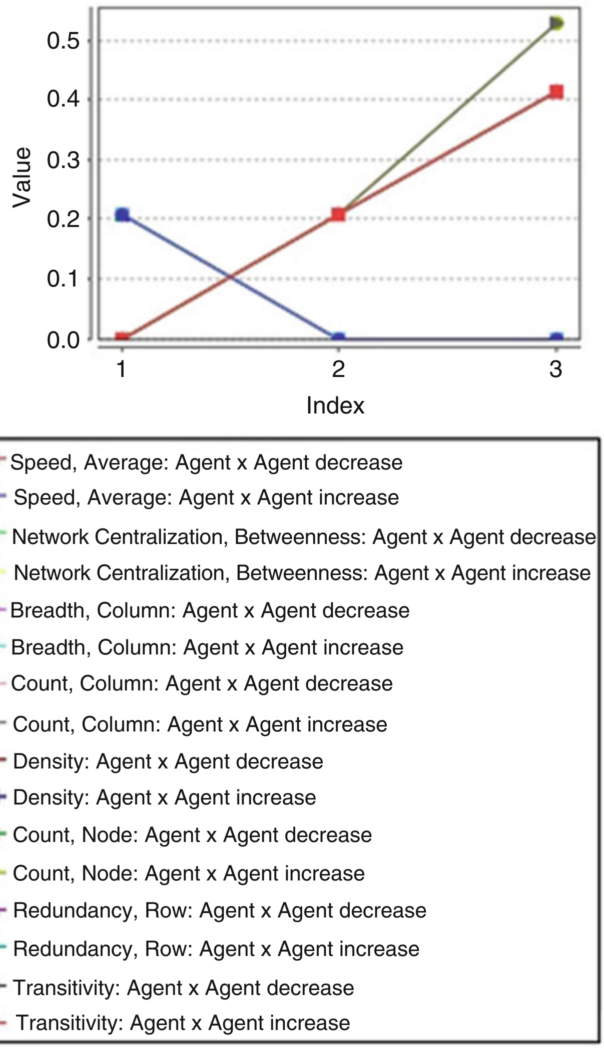

Fig. 14.7 Over time measurement – change detection

Third Scenario

The third scenario is divided into three rules. This scenario uses BU log dataset and each rule represents a particular query to filter the dataset. These rules are generated from Rough FFPSO data warehouse of intelligent Web caching on proxy cache [18, 19].

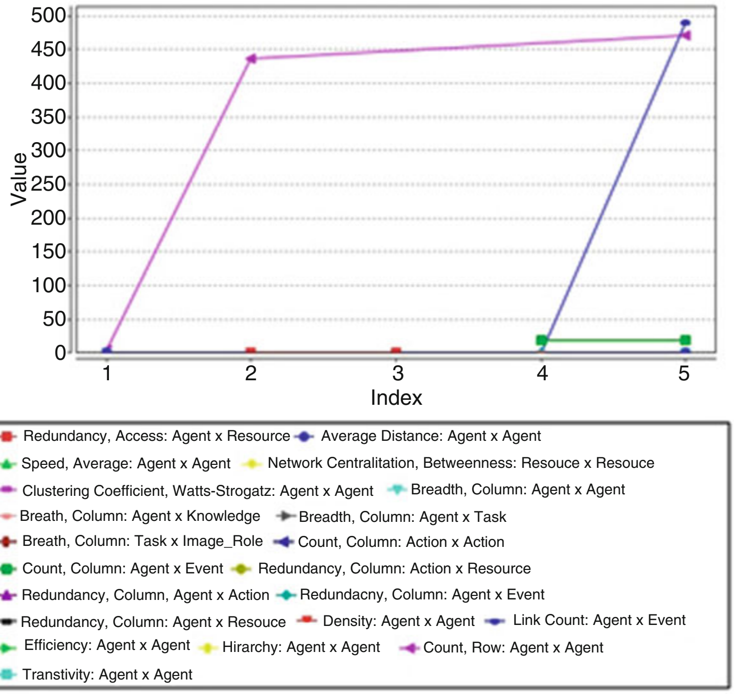

Fig. 14.8 Over time measurement – measure value

Rule 1: SIZE([0.00003, 0.00004)) AND NUM_OF_HITS([0.00015, 0.00045)) = > CACHE(1)

In this rule, the selection query is made through a query as shown below:

```
SELECT * FROM [BU Dataset]
WHERE SIZE > =0.00003 AND SIZE < =0.00004 AND
NUM_OF_HITS > =0.00015 AND NUM_OF_HITS < =0.00045;
```

The visualization of this data is centralized and focused in one position as depicts in Figs. 14.14 and 14.15.

The speed of agent is steady at the beginning, and it increases rapidly starting from index 7. It reaches the climax at index 10 then starts to decrease (see Fig. 14.16).

Fig. 14.9 3D visualization – zoom out

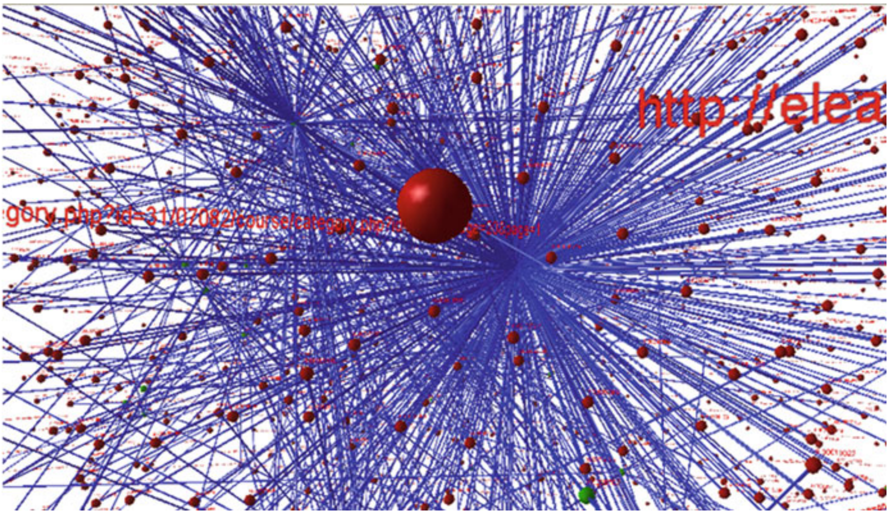

Fig. 14.10 3D visualization – zoom in

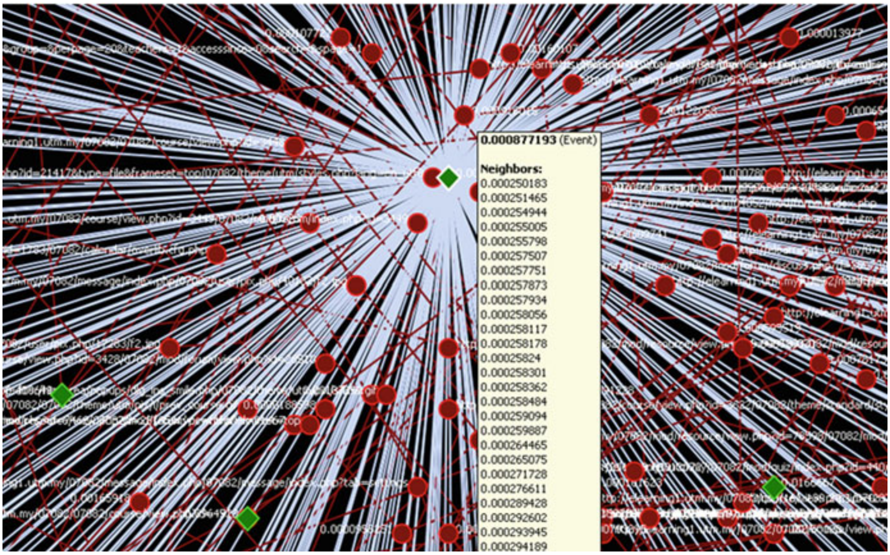

Fig. 14.11 2D visualization- zoom in

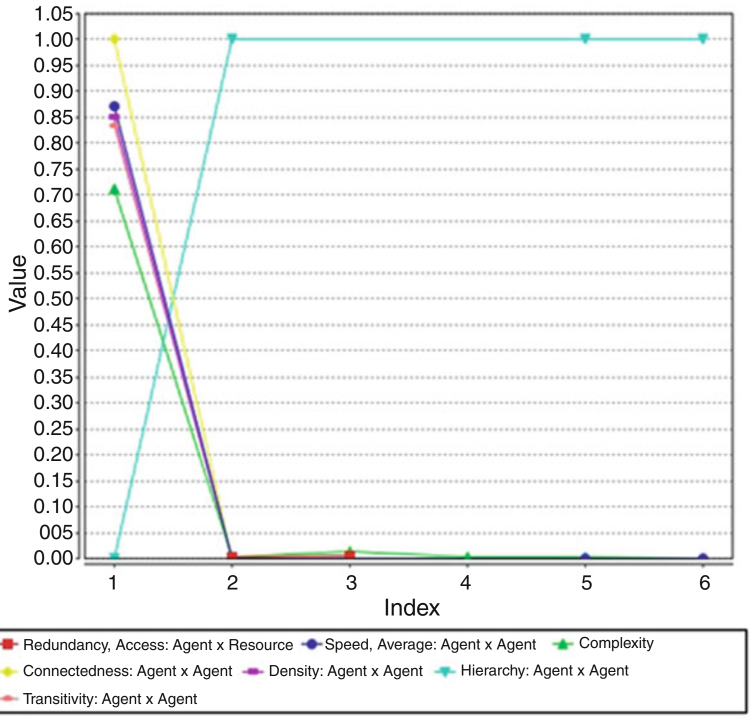

Fig. 14.12 Over time measurement – measure value

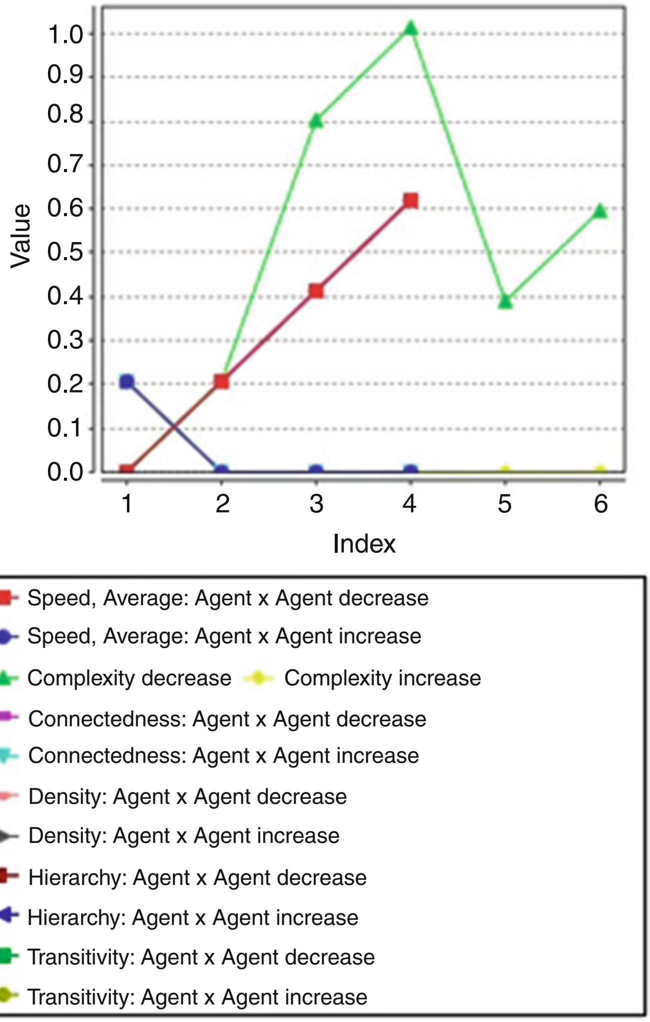

Fig. 14.13 Over time measurement – change detection

Rule 2: SIZE([0.00001, 0.00002)) AND NUM_OF_HITS([0.00044, 0.00132)) = > CACHE(0)

In this rule, the selection query is made through a query as shown below:

```
SELECT * FROM [BU Dataset]
WHERE SIZE > =0.00001 AND SIZE < =0.00002 AND
NUM_OF_HITS > =0.00044 AND NUM_OF_HITS < =0.00132;
```

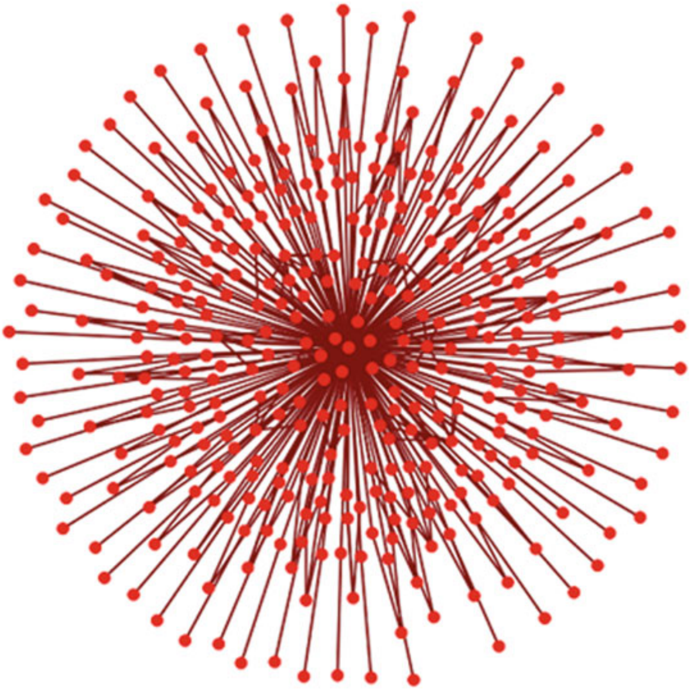

Fig. 14.14 Visualization of third scenario for rule 1

The second rule illustrates another interesting behavior. The BU data is centralized in two big localizations as depicted in Figs. 14.17 and 14.18.

Figure 14.19 describes the behavior of data in rule 2 and shows the increasing power in frequency 5.0. However, the power decreases in frequency between 5.0 and 7.5.

Rule 3: SIZE([0.00006, 0.00007)) AND NUM_OF_HITS([0.00044, 0.00132)) = > CACHE(1)

In this rule, the selection query is made through a query as stated below:

```
SELECT * FROM [BU Dataset]
WHERE SIZE > =0.00006 AND SIZE < =0.00007 AND
NUM_OF_HITS > =0.00044 and NUM_OF_HITS < =0.00132;
```

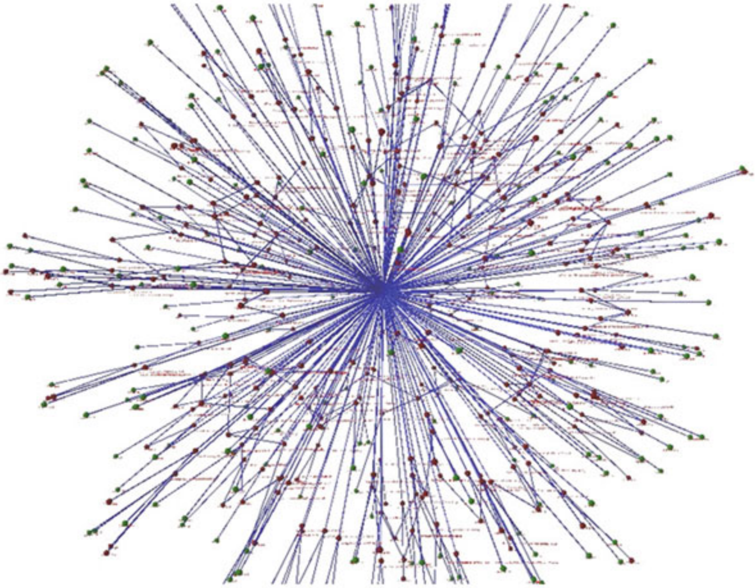

Fig. 14.15 Visualization of third scenario rule 1 – 3D mode

The third rule has made two connected localizations and one separated localization. The separated data has shown that their data has no similarity at all other localization (see Figs. 14.20 and 14.21). Figure 14.22 is clustering mode to one of the localization in this rule 3 dataset.

Figure 14.23 describes the measurement of rule 3 dataset using change detection. Speed reaches the climax value at index 6, nevertheless starts to decrease at the next index.

Another measurement analysis that explains the performance of the graph for each scenario is as shown in Appendices (refer to Appendices 1–5). These five appendices summarize the top scoring nodes from the dataset and calculated based on (1–7) [20, 21]:

(a) Betweenness Centrality

The betweenness centrality of node k in a network is defined as: across all node pairs that have a shortest path containing k, the percentage that pass through k. Individuals or organizations that are potentially influential are positioned to broker connections between groups and to bring to bear the influence of one group on

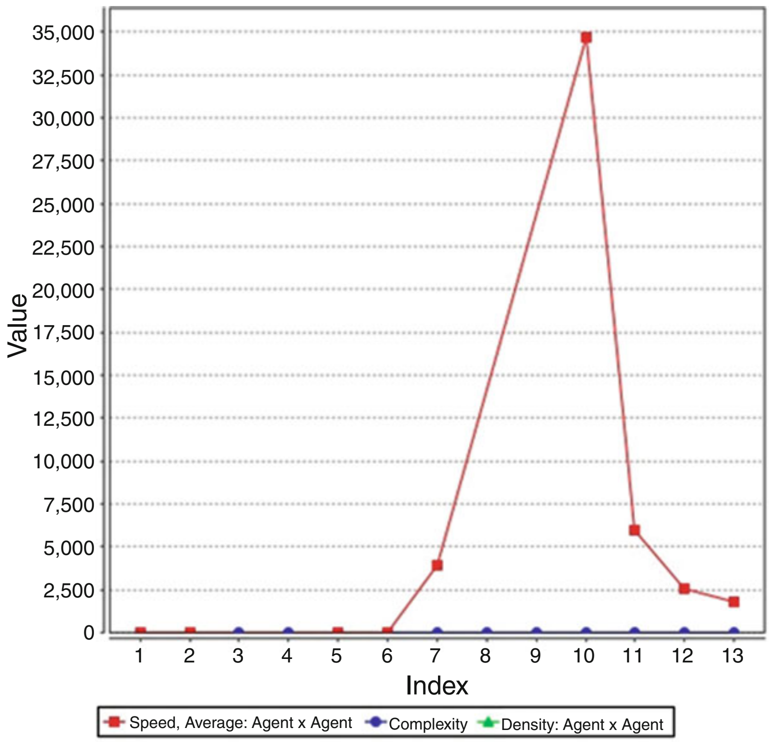

Fig. 14.16 Over time measurement – rule1

another or serve as a gatekeeper between groups. This agent occurs on many of the shortest paths between other agents. The scientific name of this measure is betweenness centrality and it is calculated on agent by agent matrices as formulated in (14.1):

$$C_k^{\mathrm{BET}} = \sum_i \sum_j \frac{g_{ikj}}{g_{ij}}, i \neq j \neq k \tag{14.1}$$

where g_{ij} is number of geodesic path from i to j, while g_{ijk} is number of geodesic path through node k.

(b) Closeness Centrality

This type of centrality calculates the average for closeness of a node to the other nodes in a network. Loosely, closeness is the inverse of the average distance in

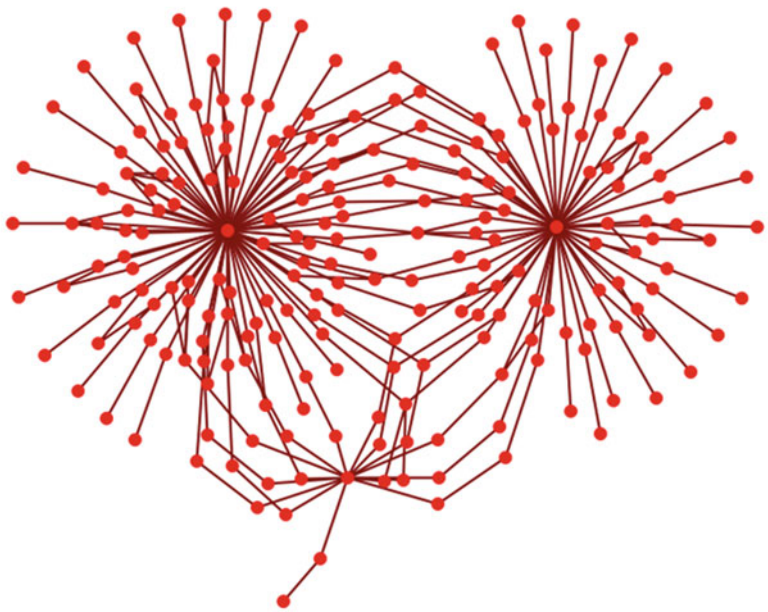

Fig. 14.17 Visualization of third scenario for rule 2

the network between the node and all other nodes. The specific equation for this centrality as shown below in (14.2):

$$C_i^{\mathrm{CLO}} = \sum_j dij \tag{14.2}$$

where d_{ij} is total geodesic distance from a given node to all other nodes.

(c) Eigenvector Centrality

This kind of centrality calculates the principal eigenvector of the network. A node is central to the extent that its neighbors are central. Leaders of strong cliques are individuals who or organizations who are collected to others that are themselves highly connected to each other. In other words, if we have a clique then the individual most connected to others in the clique and other cliques, is the leader of the clique. Individuals or organizations who are connected to many otherwise isolated individuals or organizations will have a much lower score in this measure then those that are connected to groups that have many connections themselves. The scientific name of this measure is eigenvector centrality, and it is calculated on agent by agent matrices as follows in (14.3):

$$\lambda v = Av \tag{14.3}$$

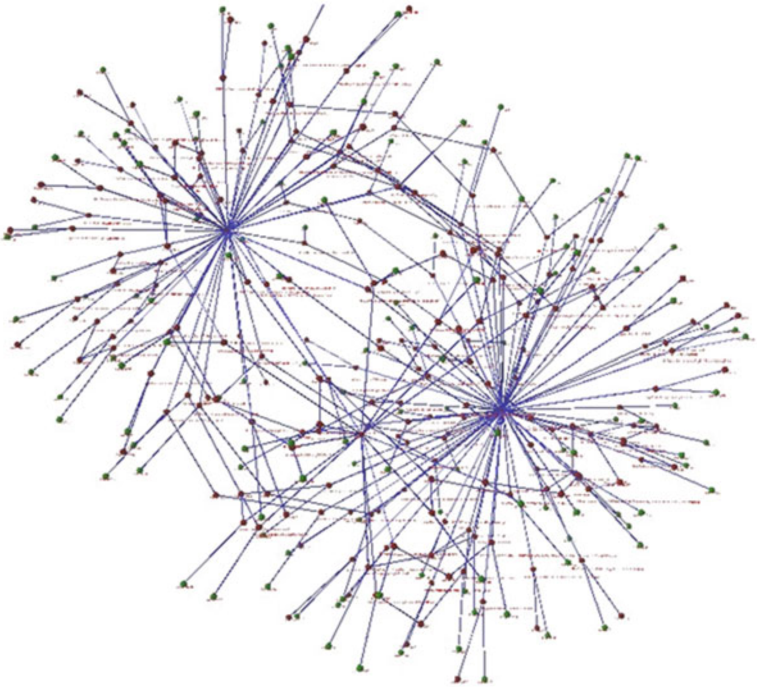

Fig. 14.18 Visualization of third scenario rule 2 – 3D Mode

where A is the adjacency matrix of the graph, λ is a constant (the eigenvalue), and v is the eigenvector.

(d) Degree Centrality

The degree centrality is divided into three centralities: in-degree, out-degree and total centrality. Equation 14.4 is a basic formula to calculate the degree centrality:

$$C_d^{\mathrm{DEG}}(v, G) \equiv |N(v)| \text{ for undirected } G \tag{14.4}$$

In the directed case, three notions of degree are generally encountered:

(e) In-degree Centrality

The in-degree centrality of a node is its normalized in-degree. For any node, e.g. an individual or a resource, the in-links are the connections that the node of interest receives from other nodes. For example, imagine an agent by knowledge matrix

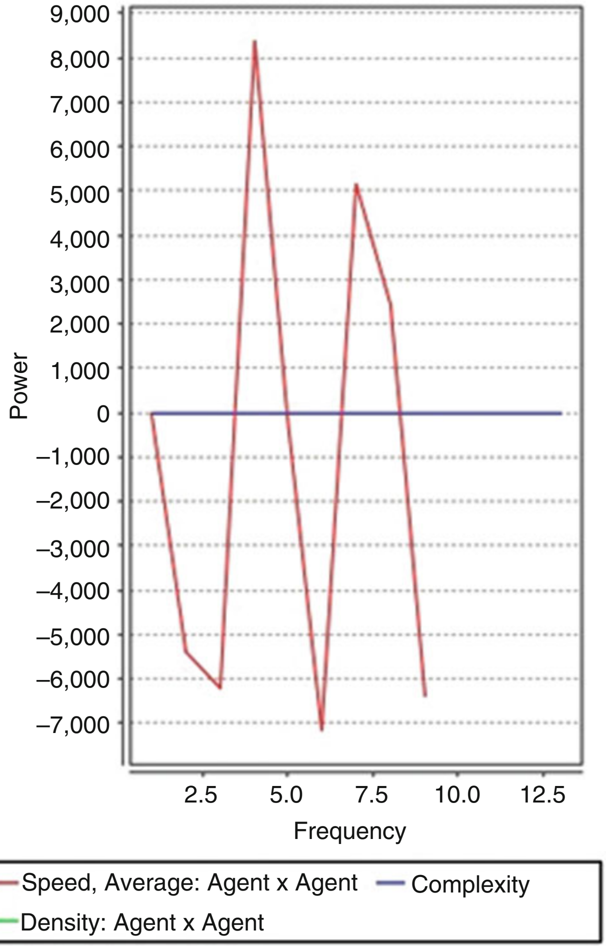

Fig. 14.19 Over time measurement – Fast Fourier Transform (FFT)

then the number of in-links a piece of knowledge has is the number of agents that are connected to. The scientific name of this measure is in-degree and it is calculated on the agent by agent matrices as shown in (14.5):

$$\left(C_{d-}^{\mathrm{DEG}}\left(v,G\right) \equiv \left|N^{-}(v)\right|\right) \tag{14.5}$$

(f) Out-degree Centrality

For any node, e.g. an individual or a resource, out-links are the connections that the node of interest sends to other nodes. For example, imagine an agent by knowledge

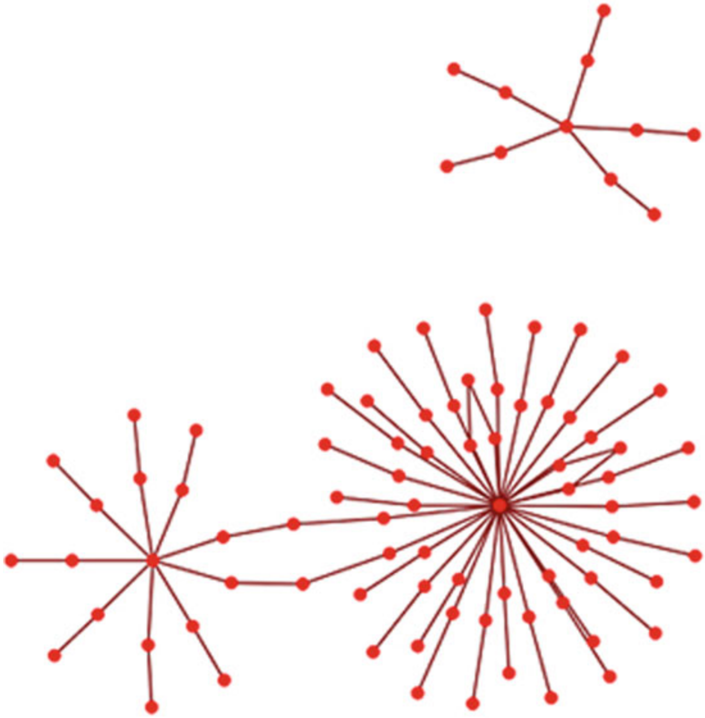

Fig. 14.20 Visualization of third scenario for rule 3

matrix then the number of out-links an agent would have is the number of pieces of knowledge it is connected to. The scientific name of this measure is out-degree and it is calculated on the agent by agent matrices. Individuals or organizations who are high in most knowledge have more expertise or are associated with more types of knowledge than are others. If no sub-network connecting agents to knowledge exist, then this measure will not be calculated. The scientific name of this measure is out degree centrality and it is calculated on agent by knowledge matrices. Individuals or organizations who are high in "most resources" have more resources or are associated with more types of resources than are others. If no sub-network connecting agents to resources exist, then this measure will not be calculated. The scientific name of this measure is out-degree centrality and it is calculated on agent by resource matrices as formulated in (14.6):

$$\left(C_{d+}^{\mathrm{DEG}}(v, G) \equiv \left|N^{+}(v)\right|\right) \tag{14.6}$$

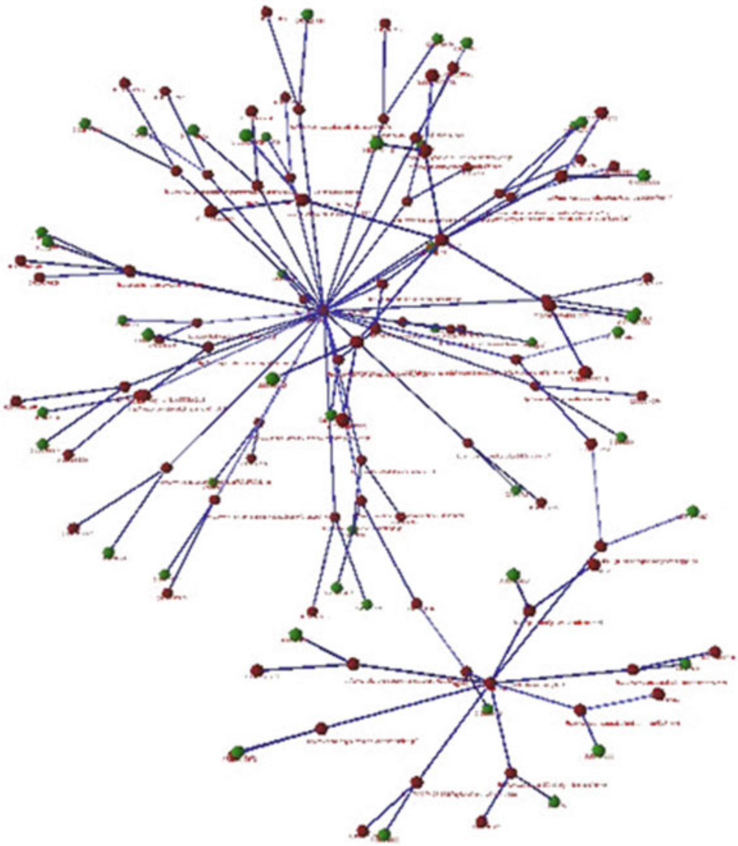

Fig. 14.21 Visualization of third scenario rule 3 – 3D mode

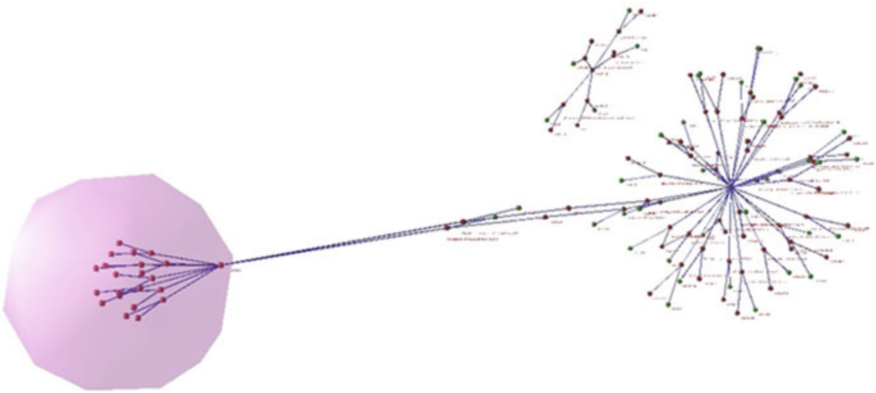

Fig. 14.22 Visualization of third scenario rule 3 – cluster mode

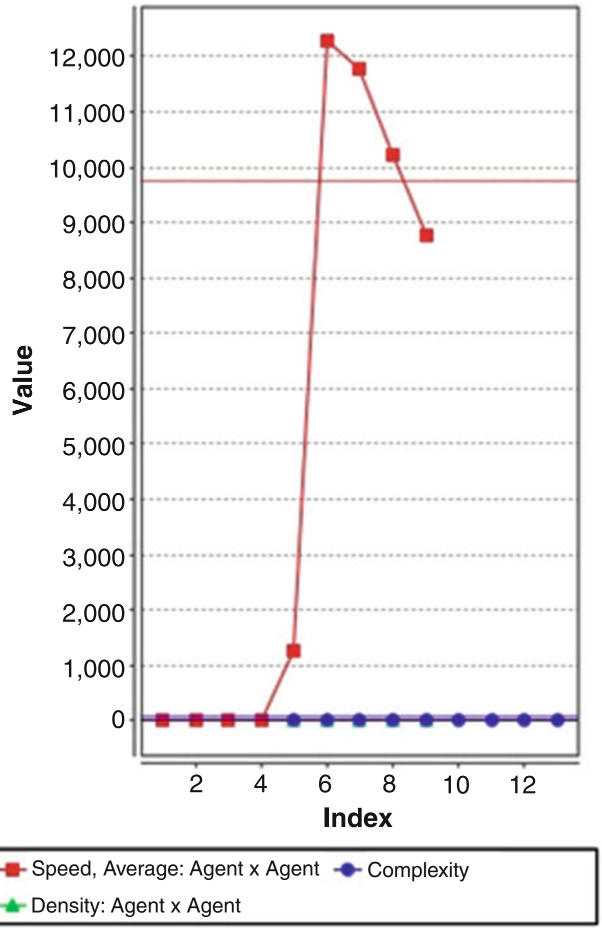

Fig. 14.23 Over time measurement – change detection

(g) Total-degree Centrality

The total degree centrality of a node is the normalized sum of its row and column degrees. Individuals or organizations who are "in the know" are those who are linked to many others and so, by virtue of their position have access to the ideas, thoughts, beliefs of many others. Individuals who are "in the know" are identified by degree centrality in the relevant social network. Those who are ranked high on this metrics have more connections to others in the same network. The scientific name of this measure is total degree centrality and it is calculated on the agent by agent matrices. Equation 14.7 is a total between the in-degree and out-degree formula to get a calculation of the total-degree centrality:

$$\left(C_{d^t}^{\mathrm{DEG}}(v, G) \equiv C_{d+}^{\mathrm{DEG}}(v, G) + C_{d-}^{\mathrm{DEG}}(v, G)\right) \tag{14.7}$$

ORA tool provides an automatic calculation of centrality measurement in their system. We used the ORA facilities to analyze our data.

Discussion and Conclusion

This chapter proposes a new method and reveals hidden information of two different Web log dataset from two different servers using SNA. First scenario shows that the behavior of dataset is centralized in several positions and looks like they are making their own cluster. Second scenario is centralized in one position and makes network resembling a nucleus of an atom. This means that the data in the second scenario is mostly referred by other nodes inside the network. According to the measurement using the standard network analysis, the speed of second scenario drops much faster than the first scenario due to the complexity of network on dataset two which is higher than the first scenario. Alternatively, the third scenario also reveals interesting results; the first rule characteristic is centralized in one node, while in second rule it has been localized at two major locations. The third rule shows a different view than the previous rule. It is illustrated in two connected locations and one separated position. The different between expected and obtained results between each rule can probably a consequence of the exact rules for the precise dataset, as claimed in section “Experimental Result and Analysis”.

This experiment reveals the capability of SNA in revealing the hidden behavior and information of Web cache data. If we use either the accurate proposed Rough FFPSO rules or trial and error queries for the right Web log dataset, we will find that a network position will be centralized. The centralization in one location demonstrates that one of the dataset is potential to be another data. This means we need to handle this data in the future in order to avoid overload access to one of the dataset.

Acknowledgments The first author would like to thank the Ministry of Higher Education (MOHE) for her scholarship. The authors also would like to thank Research Management Centre (RMC), Human Capital Development Unit and Soft Computing Research Group (SCRG), K-Economy Research Alliance, Universiti Teknologi Malaysia (UTM) for their overwhelming cooperation. We are sincerely thankful to Ahmad Hoirul Basori for providing helpful and excellent support to realize this research.

Appendices

Appendix 1 Scenario One (Top Scoring Nodes Side by Side for Selected Measure)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
1	http://elearning1.utm.my/07082/07082/course/view.php?id=440	http://elearning1.utm.my/07082/07082/course/view.php?id=440	http://elearning1.utm.my/07082/07082/course/view.php?id=440	http://elearning1.utm.my/07082/07082/course/view.php?id=440	http://elearning1.utm.my/07082/07082/course/view.php?id=440	http://elearning1.utm.my/07082/07082/course/view.php?id=440
2	0.00413256	0.00413256	0.00413256	0.00413256	0.00413256	0.00413256
3	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91854/07082/file.php/6114/week_3/2-Intro-UTM-.ppt
4	0.0639999	0.0639999	0.0639999	0.0639999	0.0639999	0.0639999
5	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/course/category.php?id=26&perpage=20&page=3/07082/course/view.php?id=6114
6	0.00324554	0.00324554	0.00324554	0.00324554	0.00324554	0.00324554
7	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=91855/07082/file.php/6114/week_3/3-Intro-UTM.ppt

(continued)

Appendix 1 (continued)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
8	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114	http://elearning1.utm.my/07082/file.php/6114/200708/Tutorial-basic.doc/07082/course/view.php?id=6114
9	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93456/07082/file.php/6114/200708/4-stereochem-UTM.ppt
10	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt	http://elearning1.utm.my/07082/mod/resource/view.php?id=93458/07082/file.php/6114/200708/5-Bondin-VBT-UTM.ppt

Appendix 2 Scenario Two (Top Scoring Nodes Side by Side for Selected Measure)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
1	http://elearning.utm.my/07082	http://elearning.utm.my/07082	http://elearning.utm.my/07082	http://elearning.utm.my/07082	http://elearning.utm.my/07082	http://elearning.utm.my/07082
2	0.0000186767	0.0000186767	0.0000186767	0.0000186767	0.0000186767	0.0000186767
3	http://elearning.utm.my/07082/	http://elearning.utm.my/07082/	http://elearning.utm.my/07082/	http://elearning.utm.my/07082/	http://elearning.utm.my/07082/	http://elearning.utm.my/07082/
4	0.00180896	0.00180896	0.00180896	0.00180896	0.00180896	0.00180896
5	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/standard/styles.php?lang=en_utf8
6	0.00269226	0.00269226	0.00269226	0.00269226	0.00269226	0.00269226
7	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8	http://elearning1.utm.my/07082/07082/theme/utm/styles.php?lang=en_utf8
8	0.000631347	0.000631347	0.000631347	0.000631347	0.000631347	0.000631347
9	http://elearning1.utm.my/07082/07082/lib/javascript-static.js	http://elearning1.utm.my/07082/07082/lib/javascript-static.js	http://elearning1.utm.my/07082/07082/lib/javascript-static.js	http://elearning1.utm.my/07082/07082/lib/javascript-static.js	http://elearning1.utm.my/07082/07082/lib/javascript-static.js	http://elearning1.utm.my/07082/07082/lib/javascript-static.js
10	0.000404541	0.000404541	0.000404541	0.000404541	0.000404541	0.000404541

Appendix 3 Scenario Three: First Rule (Top Scoring Nodes Side by Side for Selected Measure)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
1	http://wings.buffalo.edu/world	http://wings.buffalo.edu/world	http://wings.buffalo.edu/world	http://wings.buffalo.edu/world	http://wings.buffalo.edu/world	http://wings.buffalo.edu/world
2	0.0000342883	0.0000342883	0.0000342883	0.0000342883	0.0000342883	0.0000342883
3	0.000300571	0.000300571	0.000300571	0.000300571	0.000300571	0.000300571
4	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif	http://mistral.enst.fr/~pioch/louvre/corp/eb/jump.gif
5	0.0000338095	0.0000338095	0.0000338095	0.0000338095	0.0000338095	0.0000338095
6	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/redball.gif
7	0.0000303987	0.0000303987	0.0000303987	0.0000303987	0.0000303987	0.0000303987
8	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif	http://cs-www.bu.edu/staff/TA/nafnaf/www/icons/yellowball.gif
9	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif	http://zcias3.ziff.com/~pcweek/eamonn/icons/redball.gif
10	0.0000318348	0.0000318348	0.0000318348	0.0000318348	0.0000318348	0.0000318348

Appendix 4 Scenario Three: Second Rule (Top Scoring Nodes Side by Side for Selected Measure)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
1	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html	http://www-penninfo.upenn.edu:1962/penninfo-srv.upenn.edu/9000/11584.html
2	0.0000136435	0.0000136435	0.0000136435	0.0000136435	0.0000136435	0.0000136435
3	0.000901713	0.000901713	0.000901713	0.000901713	0.000901713	0.000901713
4	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html	http://cs-www.bu.edu/faculty/heddaya/CS103/syllabus.html
5	0.0000172937	0.0000172937	0.0000172937	0.0000172937	0.0000172937	0.0000172937
6	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue	gopher://gopher.enews.com:2100/11/newsletter/bw/Current%20Issue
7	0.0000187897	0.0000187897	0.0000187897	0.0000187897	0.0000187897	0.0000187897
8	0.000601142	0.000601142	0.000601142	0.000601142	0.000601142	0.000601142
9	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue	gopher://gopher.enews.com:2100/11/magazines/alphabetic/gl/games/Current%20Issue
10	0.0000147206	0.0000147206	0.0000147206	0.0000147206	0.0000147206	0.0000147206

Appendix 5 Scenario Three: Third Rule (Top Scoring Nodes Side by Side for Selected Measure)

Rank	Betweenness centrality	Closeness centrality	Eigenvector centrality	In-degree centrality	Out-degree centrality	Total degree centrality
1	http://multivac.ludd.luth.se/pub/sounds/songs/bach/	http://multivac.ludd.luth.se/pub/sounds/songs/bach/	http://multivac.ludd.luth.se/pub/sounds/songs/bach/	http://multivac.ludd.luth.se/pub/sounds/songs/bach/	http://multivac.ludd.luth.se/pub/sounds/songs/bach/	http://multivac.ludd.luth.se/pub/sounds/songs/bach/
2	0.00006726	0.00006726	0.00006726	0.00006726	0.00006726	0.00006726
3	0.000601142	0.000601142	0.000601142	0.000601142	0.000601142	0.000601142
4	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html	http://www.ccs.neu.edu/home/thigpen/ROM/ROM.html
5	0.0000619941	0.0000619941	0.0000619941	0.0000619941	0.0000619941	0.0000619941
6	0.00120228	0.00120228	0.00120228	0.00120228	0.00120228	0.00120228
7	http://metaverse.com/	http://metaverse.com/	http://metaverse.com/	http://metaverse.com/	http://metaverse.com/	http://metaverse.com/
8	0.0000667215	0.0000667215	0.0000667215	0.0000667215	0.0000667215	0.0000667215
9	http://www.ibm.com/images/solutions_icon1.gif	http://www.ibm.com/images/solutions_icon1.gif	http://www.ibm.com/images/solutions_icon1.gif	http://www.ibm.com/images/solutions_icon1.gif	http://www.ibm.com/images/solutions_icon1.gif	http://www.ibm.com/images/solutions_icon1.gif
10	0.0000647468	0.0000647468	0.0000647468	0.0000647468	0.0000647468	0.0000647468

References

1. Grissa, D., Guillaume, S., Nguifo, E. M.: Combining clustering techniques and formal concept analysis to characterize interestingness measures. CoRR abs/1008.3629 (2010)
2. Martinez, A., Dimitriadis, Y., Rubia, B., Gomez, E., Fuente, P.D.L.: Combining qualitative evaluation and social network analysis for the study of classroom social interactions. Comput. Educ. **41**, 353–368 (2003)
3. Kudelka, M., Snasel, V., Horak, Z., Hassanien, A.E., Abraham, A.: Web communities defined by web page content. In: Computational Social Networks: Tools, Perspectives and Analysis, pp. 349–370. Springer, London (2010). ISBN 978–1–84882–228–3
4. Erlin, A., Yusof, N., Abdul, R.A.: Overview on agent application to support collaborative learning interaction. J. US-China Educ. Rev. **5**(38), 52–60 (2008). ISSN 1548–6613, USA
5. Shi, X.: Social Network Analysis of Web Search Engine Query Logs. Technical Report, University of Michigan, School of Information (2007)
6. Al-Fayoumi Jr., M., Banerjee, S., Mahanti, P.K.: Analysis of social network using clever ant colony metaphor. World Acad. Sci. Eng. Technol. J. **53**, 970–974 (2009)
7. Sammantha, L., Magsino, R.: Applications of Social Network Analysis for Building Community Disaster Resilience: Workshop Summary. National Academies Press, Washington, DC (2009). ISBN 0–309–14095–1
8. Padmanabhan, V.N., Mogul, J.C.: Using predictive pre-fetching to improve world wide web latency. Comput. Commun. Rev. **26**, 22–36. In: ACM, SIGCOMM'96, July 1996
9. Bestavros, A., Cunha, C.: A pre-fetching protocol using client speculation for the WWW. Technical Report TR-95–011, Boston University, Department of Computer Science, Boston, MA, Apr 1995
10. Pallis, G., Vakali, A., Pokorny, J.: A clustering-based approach for short-term prefetching on a web cache environment. Comput. Electr. Eng. J. **34**(4), 309–323 (2008)
11. Kroeger, T.M., Long, D.E., Mogul, J.C.: Exploring the bounds of web latency reduction from caching and prefetching. In: USENIX Symposium on Internet Technologies and Systems (USITS), Monterey, CA, 8–11 Dec 1997
12. Harding, M., Storz O., Davies, N., Friday, A.: Planning ahead: techniques for simplifying mobile service use. In: Proceedings of the 10th workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, pp. 1–6, 23–24 Feb 2009
13. Ye, F., Li, Q., Chen, E.: Benefit based cache data placement and update for mobile peer to peer networks. World Wide Web J. **14**(3), 243–259 (2008)
14. Sulaiman, S., Shamsuddin, S.M., Forkan, F., Abraham, A., Sulaiman, S.: Intelligent web caching for e-learning log data. In: Third Asia International Conference on Modelling and Simulation, AMS 2009, pp. 136–1410. IEEE Computer Society Press, Washington, DC (2009)
15. Sulaiman, S., Shamsuddin, S.M., Forkan, F., Abraham, A.: Autonomous SPY: intelligent web proxy caching detection using neurocomputing and particle swarm optimization. In: Proceeding of the 6th International Symposium on Mechatronics and its Applications (ISMA09), Sharjah, UAE, pp. 1–6. IEEE Press, Washington, DC, ISBN: 978–1–4244–3480–0 (2009)
16. Sulaiman, S., Shamsuddin, S.M., Abraham, A., Sulaiman, S.: Intelligent mobile web prefetching using XML technology. In: Proceedings of the Sixth International Conference on Next Generation Web Services Practices (NWeSP 2010), India, pp. 129–134. IEEE Press, Washington, DC, ISBN:978–1–4244–7818–7 (2010)
17. Sulaiman, S., Shamsuddin, S.M., Abraham, A.: Rough neuro-PSO web caching and XML prefetching for accessing Facebook from mobile environment. In: Proceedings of the 8th International Conference on Computer Information Systems and Industrial Management (CISIM 2009), pp. 884–889. IEEE Press, Washington, DC, ISBN: 978–1–4244–5612–3 (2009)
18. Sulaiman, S., Shamsuddin, S.M., Abraham, A.: Rough Web Caching, Rough Set Theory: A True Landmark in Data Analysis, Studies in Computational Intelligence, pp. 187–211. Springer, Berlin (2009). ISBN 978–3–540–89920–4

19. Sulaiman, S., Shamsuddin, S.M., Abraham, A.: Data warehousing for rough web caching and prefetching. In: Proceedings of the IEEE International Conference on Granular Computing (IEEE GrC 2010), San Jose, pp. 443–448. IEEE Computer Society, Washington, DC, ISBN 978–0–7695–4161–7 (2010)
20. Borgatti, S.P., Everett, M.G.: A graph-theoretic perspective on centrality. Soc. Netw. **28**(4), 466–484 (2006)
21. Butts, C.T.: Social network analysis with SNA. J. Stat. Softw. **24**(6), 1–51 (2008)