Synthetic Social Network Based on Competency-Based Description of Human Resources

Štěpán Kuchař[1], Jan Martinovič[2], Pavla Dráždilová[2], and Kateřina Slaninová[2]

[1] VŠB - Technical University of Ostrava
IT4Innovations
Ostrava, Czech Republic
stepan.kuchar@vsb.cz

[2] VŠB - Technical University of Ostrava
Department of Computer Science,
Ostrava, Czech Republic
{jan.martinovic,pavla.drazdilova,katerina.slaninova}@vsb.cz

Abstract. The approach presented in this paper is based on the field of human resource management with the aim to extend the analysis of human resources by a graph theory perspective with an output representation by synthetic social networks. Further analysis of human resources is focused on their division into communities with similar competencies and skills. We used betweenness concept of centrality for finding important persons in the network that share their skills and competencies with workers in other communities and can therefore serve as contact persons between communities with different skills. This method can also be used for suggesting worker team composition based on similarity of workers' skills for different roles.

Keywords: Synthetic social network, Complex network, Human resource management, Competency model.

1 Introduction

A typical social network is as a set of people, or groups of people, who socially interact among themselves [1]. In these networks the relations are usually defined by one of the types of interaction between the actors, e.g. personal knowledge of one another, friendship, membership, etc. However, in the area of synthetic social networks, we can explore the extended definition of social networks. This can be done by exploring social network as a set of people, or groups of people who have similar patterns of contacts of interactions, or generally with similar attributes [2].

This approach can then be extended to the analysis of complex networks. Complex networks, especially in the web sphere and internet areas, are often called synthetic, or derived, social networks [3]. This type of social network differs from natural social networks due to the relationship between the nodes. They are generated on the basis of the common attributes of the nodes [4]. These attributes do not necessarily represent the physical communication or the interaction among objects like in the natural social networks [5], but other attributes representing the personal similarity. The approach

K. Saeed et al. (Eds.): CISIM 2013, LNCS 8104, pp. 313–324, 2013.

presented in this paper is based on these types of networks and the relations between workers based on similar skills and competencies.

The approach presented in the paper is based on the field of human resource management with the aim to extend the analysis and simulation of human resources in business processes by a graph theory perspective with an output representation by synthetic social networks (because real social relations and interactions between individual workers are not known). The relation to this special type of social networks can lead to further analysis of human resources focused on their thorough division into communities based on similar skills and competencies. This step is not possible without usage of the social network approach. Due to this reason, the issues of social network area are described and defined in this paper, including social network evaluation and community detection field. On the basis of performed analyses, the experiments which detect latent ties between particular resources are presented. The relations between human resources are defined not only by their similarity, but also by their membership in communities with similar behaviour.

Due to the fact, that the proposed approach consists of two different application areas, the structure of the paper is following. In Section *Social Network*, the area of social networks with evaluation and community detection is described. Section *Competency-based description of human resources* focuses on human resources and their competency description. This section ends with the description of connection between vector model for human resources and graph theory approach of synthetic social networks. Afterwards, Section *Experiments* with experiments is presented, which describes usage of social network analysis focused on human resource management. Social network evaluation is used for gaining new information about relations between human resources and for analysis of resources with interesting properties and behaviour.

2 Social Network

Social networking is a complex, large and expanding sector of the information economy. Researchers' interest in this field is growing rapidly. It has been studied extensively since the beginning of the 20th century. The first normative contributions in this area were proposed in 1970s by sociologist Mark Granovetter and mathematician Linton C. Freeman. The basic theory "The Strength of Weak Ties" was mentioned in 1973 [6]. Granovetter argued that within a social network, weak ties are more powerful than strong ties. Another significant principle was published in 1979 by Linton C. Freeman [7]. In his work was presented definition of centrality, which is one node's relationship to other nodes in the network. He defined basic metrics like degree, control and independence, from which reason researchers proceed in their present works.

Social network is a set of people or groups of people with similar patterns of contacts or interactions such as friendship, co-working, or information exchange [8]. The World Wide Web, citation networks, human activity on the internet, physical and biochemical networks are some examples of social networks. Social networks are usually represented by using graphs, where nodes represent individuals or groups and lines represent contacts among them. The configuration of relations among network members identifies a specific network structure, and this structure can vary from isolated structures where no members are connected to saturated structures in which everyone is interconnected.

A relationship between the actors in a social network can be very complex, often making them multidimensional. This fact leads to the formation of various types of social networks. Amongst others, we can mention multi-layered social networks (with homogeneous nodes, but with multiple relations), bipartite social networks (with two types of nodes), multi-modal social networks (with many types of nodes), temporal social networks (which reflect the network evolution), or multidimensional social networks, in which are combined a hierarchy of relations with a group hierarchy of nodes, and a time dimension [9].

Social network analysis was defined by Barry Wellman as "work at describing underlying patterns of social structure, explaining the impact of such patterns on behavior and attitudes" [10]. Therefore, researchers are not interested only on describing the different social structures, but they emphasize on investigating the consequences of this variation on the member's behaviors.

2.1 Evaluation of Social Networks

For a description of social networks defined in 1979, see Linton Freeman [7] various types of *centrality*, where individual network nodes are directly evaluated, or where the average value of selected centrality in a graph may be an item of interest.

A primary use of graph theory in social network analysis is to identify the important or prominent actors at both the individual and group levels of analysis. *Centrality* and *prestige* concepts and measures seek to quantify graph theoretic ideas about an actor's prominence within a complete network by summarizing the structural relations among all nodes. Centrality means that a prominent actor has high involvement in many relations, regardless of whether sending or receiving ties. Prestige is when a prominent actor initiates few relations but receives many directed ties. Knoke and Yang defined the above mentioned terms in [11].

Degree centrality requires the usage of matrix algebra notation. Unlike actor degree centrality, group degree centralization measures the extent to which the actors in a social network differ from one another extent to which the actors in a social network differ from one another in their individual degree centralities.

Closeness centrality was developed to reflect how near a node is to the other nodes in a social network [12]. Closeness and distance refer to how quickly an actor can interact with others, for example, by communicating directly or through very few intermediaries. An actor's closeness centrality is a function of its geodesic distance (length of the shortest path connecting the two nodes) to all other nodes.

Betweenness concept of centrality concerns how other actors control or mediate the relations between two nodes that are not directly connected. Actor betweenness centrality measures the extent to which other actors lie on the geodesic path between pairs of actors in the network.

To understand networks and their participants, we provide the location of actors in the network. Measuring the network location is finding the centrality of a node. These measures determine the various roles and groupings in a network – who are the connectors, specialists, leaders, bridges, isolates, where are the clusters and who is in them, who is in the core of the network, and who is on the periphery.

2.2 Community Detection

The discovery and analysis of community structure in networks is a topic of considerable recent interest in sociology, physics, biology and other fields. Networks are very useful as a foundation for the mathematical representation of a variety of complex systems such as biological and social systems, the Internet, the world wide web, and many others [13]. A common feature of many networks is âĂIJcommunity structureâĂİ, the tendency for vertices to divide into groups, with dense connections within groups and only sparser connections between them [14].

Newman and Girvan [15] proposed algorithms for finding and evaluating community structure in network. They used a âĂIJdivisiveâĂİ technique which iteratively removes edges from the network, thereby breaking it up in communities. The edges to be removed are identified by using one of a set of edge betweenness measures, of which the simplest is a generalization to edges of the standard shortest-path betweenness of Freeman. Than, their algorithms include a recalculation step in which betweenness scores are re-evaluated after the removal of every edge.

To detect communities, graph partitioning methods or hierarchical clustering has been applied. Originally, graph partitioning methods, based on edge removal [16], divide the vertices of a network into a given number of (non-overlapping) groups of a given size, while the number of edges between groups is minimal.

2.3 Spectral Clustering

Spectral clustering is one of the divisive clustering algorithms which can be applied in the graph theory. The spectral clustering algorithm uses eigenvalues and eigenvectors of a similarity matrix derived from the data set to find the clusters. In this section, there is described the type of spectral clustering based on the second smallest eigen vector of the Laplacian matrix.

Given a set of data points $\{x_1, \dots x_n\} \in \mathbb{R}^m$ and similarity (cosine measure) $a_{ij} \geq 0$ between all pairs of the data points x_i and x_j. Let $G = (V, E)$ be an undirected graph with vertex set $V = \{v_1, \dots, v_n\}$. Each vertex v_i in this graph represents the data point x_i. Two vertices are connected, if the similarity a_{ij} between the corresponding data points x_i and x_j is positive, and the edge is weighted by a_{ij}. The weighted adjacency matrix of the graph is the matrix $A = (a_{ij})\ i, j = 1, \dots, n$. If $a_{ij} = 0$ than $(v_i, v_j) \notin E(G)$. It governs that A is symmetric for the undirected graph. The degree of a vertex $v_i \in V$ is defined as $d_i = \sum_{j=1}^{n} a_{ij}$. The degree matrix D is defined as the diagonal matrix with the degrees $d_1, \dots, d_n$ on the diagonal. The unnormalized graph of Laplacian matrix is defined as $L = D - A$. In[17], Fiedler defines the second smallest eigenvalue $a(G)$ of the of Laplacian matrix $L(G)$ as algebraic connectivity of the graph G. In his honor, the corresponding eigenvector is called *Fiedler vector*. The Spectral Partitioning Algorithm which uses Fiedler vector is summarized in [16]. We used algorithm for spectral clustering (Left-Right algorithm) which is described in article [18].

3 Competency-Based Description of Human Resources

The description of the employees' skills in the process is a human resources management area of expertise where the competency models [19–21] and skills frameworks (e.g. Skills Framework for the Information Age [22]) are used. Competency models define various competencies which are important for the company and its processes. Competencies are defined as sets of knowledge, abilities, skills and behaviour that contribute to successful job performance and the achievement of organizational results [21]. Skills frameworks have the same purpose, but they describe skills particular for one domain rather than general competencies. But in fact skills are just a special type of competencies.

Competency models and skills frameworks also describe how to measure and evaluate individual competencies. In most cases competencies are measured by a number of advancing stages where higher levels of competency include everything from their lower levels. The first competency model had five stages [19] and later models used the same system, but they did not keep the number of stages. There is no standard for how many stages should a competency model have and every model defines its own set of stages.

Therefore, competencies of a specific human resource can be described by the competency level acquired by the resource. This also means that this resource has mastered this given level and all lower levels of the competency. This way it is not important how many levels does the competency model have because the computing model can assume, that the highest acquired level of the best resource is also the highest level of the competency model.

Let's have a small example of one Developer working in a software development company. His competencies in a 10-level model could look as follows:

– Java development - 7. level,
– C# development - 2. level,
– UML knowledge - 4. level,
– communication - 2. level,
– customer knowledge of VSB-TUO - 4. level,
– customer knowledge of MyCompany - 0. level.

Domain specific skills (development, UML knowledge), general competencies (communication) and knowledge of the environment (customer knowledge) are contained in this example. It is clear that competencies in the model have to be based on the company requirements and professional domain.

3.1 Competency-Based Description of Process Activity Requirements

All activities in the process also have competency-based requirements that describe what competencies should the worker performing the activity know. Therefore, each activity will be defined by the set of competency levels for each required resource type entering the activity specifying that only workers with given or higher level will do the activity as planned. Resources with lower competencies are able to finish the activity, but it will take additional time to learn how to perform the activity and their work is prone to contain more errors. A simple example of requirements for the activity of developing customer specific code in the software development company follows:

- Development - 6. level,
- UML knowledge - 3. level,
- communication - 3. level,
- customer knowledge - 4. level.

If we compare this example with the worker example from previous chapter, one can notice the generalization of some requirements (development and customer knowledge). When assessing the employee's competencies, it is better to define the competency levels in specific parts of the domain so that the resources are assessed as precisely as possible. On the other hand, the activity requirements should only define a level for the whole competency category, and relevant part of the domain will be specified by actual process case. In other words, if the development company tackles with a case where they have to develop a Java code for the company VSB-TUO, then the requirements in this case will be refined as Java development and customer knowledge of VSB-TUO.

3.2 Competency Models and Synthetic Social Network

To create a synthetic social network based on the competencies of human resources, similarities between these resources had to be evaluated. This evaluation was performed using vector space model that is very often used in document searches [23]. To use this model for the competencies, a way to describe the resource competencies as vectors had to be found. This was solved by devising fragmented vector representation of the competency levels for given resource. This representation and its different properties and validation was described in our previous work (see [24]).

On the basis of created vector model, similarity matrix $M^{H \times H}$ can be constructed for the set of resources R. The matrix contains similarities of particular resources from range of values $< 0, 1 >$. It is suitable to filter vertices between resources, which are of lower importance, for construction of synthetic social network and for further finding of communities. For this purpose, the threshold λ is defined. Afterwards, it is possible to construct graph $G(R,E)$, where E represents a strength of vertices between particular resources, while weights $w \in E$ meet the constraint $w \geq \lambda$. After construction of graph G, we can find community resources with similar attributes. In the proposed approach, we use Left-Right Algorithm for community detection, described in Section *Spectral Clustering*. The output set of communities C is used in further experiments.

4 Experiments

For the experiment, we created a synthetic social network for the workers involved in a software process of a local middle-sized software development company. 8 roles were identified in the process and their possible competency level intervals were specified based on the process requirements and several selected worker profiles (competency profiles were not available for all workers in the process). Then, 143 competency profiles were created based on these constraints, each containing 19 competencies important in the software process, and each on a 10-level scale. 41 basic activities were analysed in the process and their requirements were specified for the same 19 competencies to ensure their compatibility.

The network in this experiment was created by using the similarities between competency profiles specifying individual human resources in the process. The threshold λ for creating the network was set to 0.7 to filter insignificant connections that cluttered the network.

4.1 Visualization of Detected Communities

The first added value that the created network brought was the possibility to detect communities of workers in the process based on their competencies. Communities shown in Fig. 1 were detected by the Left-Right Algorithm described in Section *Spectral Clustering*.

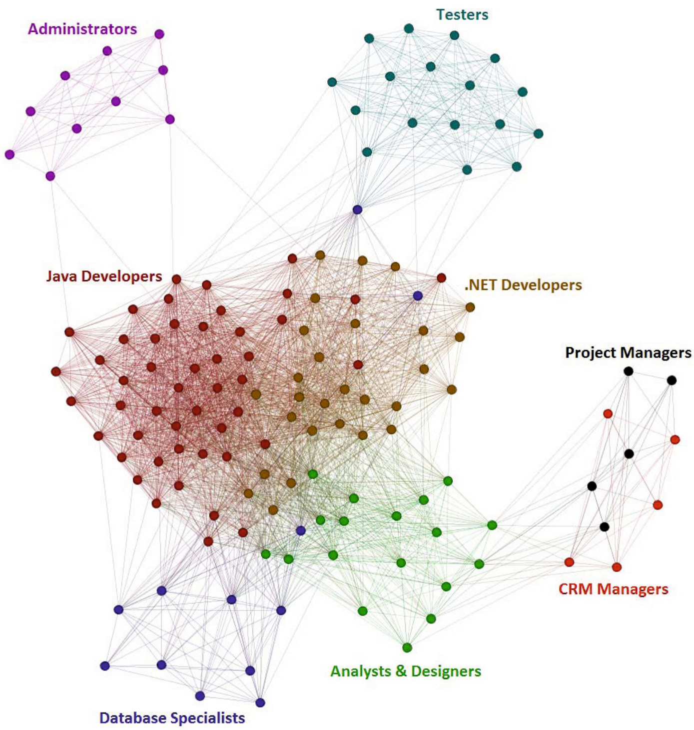

Fig. 1. Communities in the Competency-based Synthetic Social Network

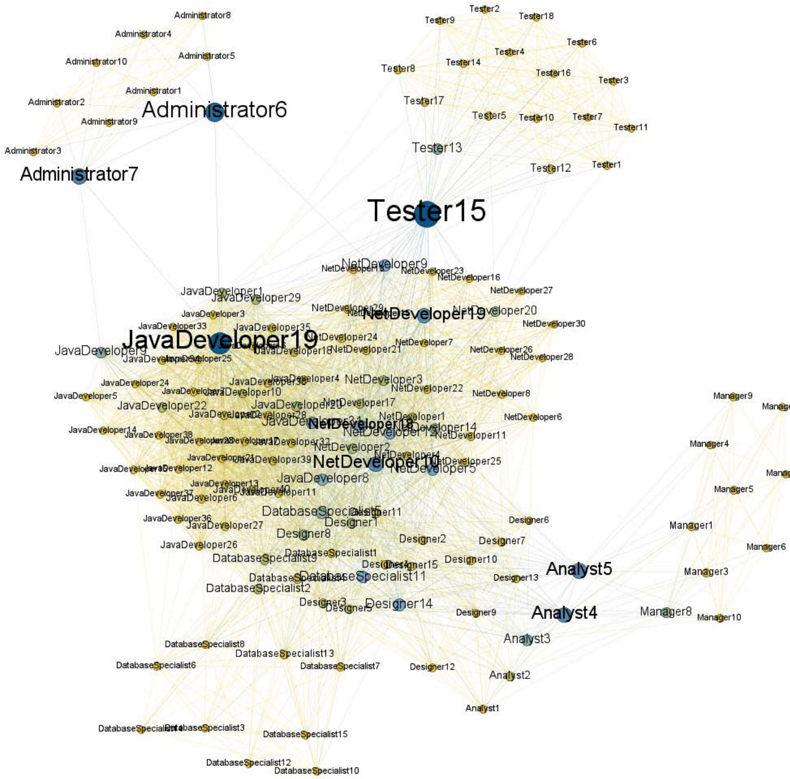

Fig. 2. Betweenness Centrality Evaluation

The algorithm detected eight distinct communities: Administrators, Analysts & Designers, CRM Managers, Database Specialists, Java Developers, .NET Developers, Project Managers and Testers. These communities are very similar to the roles in the process except for one combined community for both Analysts and Designers (showing that these two roles are very similar in their competencies) and two separate communities for two types of managers that were defined as one role in the process. Even though these roles were known prior to the experiment, it showed that this detection algorithm could be effectively used for discerning roles in the environment without predefined roles.

4.2 Evaluation of Betweenness Centrality

One of the interesting evaluation method of the social networks is betweenness centrality (see Section *Evaluation of Social Networks*) that specifies how much a node in the network links other nodes together. When considering the network based on competency similarities, betweenness identifies universal workers that have acquired knowledge from multiple disciplines. These workers can serve as communication bridges among different communities. Betweenness centrality for individual workers is proportionally displayed in Fig. 2.

Table 1. Number of Neighbours for Resources Suitable for the System Architecture Analysis-MSSQL,VSB-TUO Activity

Resource	Own community	Neighbouring communities				
	Analysts & Designers	.NET Developers	CRM Managers	Java Developers	Project Managers	Database Specialists
Similarity threshold = 0.7						
Analyst3	17	5	2	2	2	
Analyst1	17	2	1	1		
Analyst5	17	5	3		4	1
Analyst4	16	4	4	1	3	1
Similarity threshold = 0.75						
Analyst3	17	1	2		1	
Analyst1	9	2				
Analyst5	10	2	2		1	
Analyst4	11	2				
Similarity threshold = 0.8						
Analyst3	5	1				
Analyst1	4					
Analyst5	6	1				
Analyst4	4					
Similarity threshold = 0.85						
Analyst3	4					
Analyst1	1					
Analyst5	3					
Analyst4	1					

Tester15 has the biggest betweenness centrality for obvious reasons because he connects the community of Testers with the Java developers and .NET developers communities. Administrator6 and Administrator7 create a similar link between Administrators and Java and .NET developers. On the other hand, JavaDeveloper19 creates a bridge between a lot of Java developers and Administrators. Analyst4 and Analyst5 connect Managers with Analysts & Designers and other communities.

4.3 Analysis of Connections to Different Communities

All prior analyses considered the network as a whole, but very interesting results can be gained by looking at individual workers in the process. The network connects resources with similar competencies and with similar knowledge. Therefore, connected individuals can understand each other more easily because they share common knowledge and common behaviour. Analysing these connections for a specific worker can lead to finding people in other communities that could make a more effective team or that could be used for easier acquisition of additional knowledge from another part of the process. This information could also be used for choosing more appropriate worker for performing an activity because his hidden knowledge and easier collaboration could help him to understand the domain more quickly.

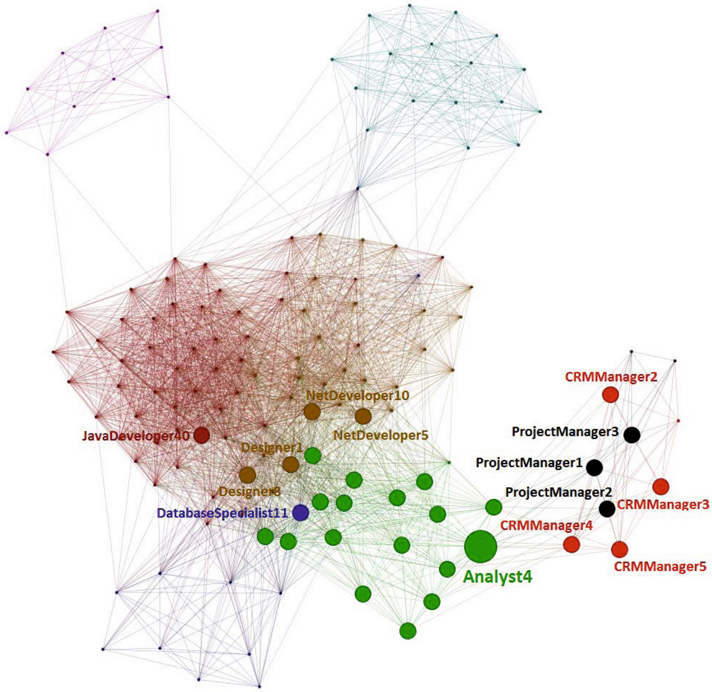

Fig. 3. Neighbours of the Analyst4 Resource

Process activities and their competency requirements play an important role in this analysis because the correct choice of appropriate worker is primarily based on these requirements. Table 1 contains the number of neighbouring resources for each resource that is suitable for the System Architecture Analysis activity for the process case specialized on MSSQL database for VSB-TUO customer.

This table is separated into several sections based on the similarity threshold that was used to look for neighbouring nodes for each resource. The resources are sorted according to their competency suitability to perform the specified activity, Analyst3 being the most suitable and Analyst4 being still able to perform the activity but having to spent more time with the activity. The neighbourhood analysis shows that even though Analyst5 is the third in suitability, his knowledge similar to one of the Database Specialists could help him overcome some specific difficulties concerning an analysis of heavily database-centric system and therefore it could be a better match for such task. On the other hand, Analyst3 could be a better match for a process case concerning Java-related specifics because he could simplify the further work on the design by providing language-related hints to the architecture analysis document.

In considering the team composition, not only number of neighbours is important, but actual neighbouring workers have to be identified. These neighbours share common knowledge and could find a better ground at understanding each other when collaborating even though they have not met before. Fig. 3 shows neighbouring workers for the Analyst4 worker.

5 Conclusion and Future Work

A new method for using the synthetic social networks in the field of allocating human resources and finding suitable representatives for other spheres of human activities was presented in this paper. The experiments proved the hypothesis that selected human resources have relation to other sources, which are oriented not only to queried resource and similar activities, but to other activities as well. Many neighbouring resources are classified into other communities and this information can be used to enhance the communication between different parts of the process. Moreover, the betweenness centrality evaluation detected resources that create bridges between communities obtained by our developed Left-Right algorithm.

In future work, presented results will be used for identifying teams that will be able to collaborate more effectively due to their common knowledge [25–27]. We intend to use extended queries by several fields and obtained knowledge about community overlapping. This means that selected worker may be important not only for his own community but that he may have relations to other communities as well.

Based on presented results, a combination of social network analysis and human resources field can provide added value for human resource allocation, collaboration and decision support processes.

Acknowledgments. This work was partially supported by the European Regional Development Fund in the IT4Innovations Centre of Excellence project (CZ.1.05/1.1.00/02.0070) and by SGS, VSB – Technical University of Ostrava, Czech Republic, under the grant No. SP2013/167 Analysis of the behaviour patterns in complex networks.

References

1. Newman, M.E.J.: Networks: An Introduction. Osford University Press (2010)
2. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks (February 2004)
3. Musiałl, K., Kazienko, P.: Social networks on the internet. World Wide Web 1, 1–42 (2012)
4. Costa, L., Rodrigues, F., Travieso, G., Boas, P.: Characterization of complex networks: A survey of measurements. Advances in Physics 56(1), 167–242 (2007)
5. Bisgin, H., Agarwal, N., Xu, X.: Investigating homophily in online social networks. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2010, pp. 533–536. IEEE Computer Society, Los Alamitos (2010)
6. Granovetter, M.S.: The Strength of Weak Ties. American Journal of Sociology 78(6), 1360–1380 (1973)
7. Freeman, L.C.: Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239 (1979)
8. Garton, L., Haythornthwaite, C., Wellman, B.: Studying online social networks. Journal of Computer-Mediated Communication 3(1) (1997)
9. Kazienko, P., Musial, K., Kukla, E.z., Kajdanowicz, T., Bródka, P.: Multidimensional social network: Model and analysis. In: Jędrzejowicz, P., Nguyen, N.T., Hoang, K. (eds.) ICCCI 2011, Part I. LNCS (LNAI), vol. 6922, pp. 378–387. Springer, Heidelberg (2011)

10. Knoke, D., Yang, S.: Social network analysis. Quantitative applications in the social sciences, vol. 154. Sage (2008)
11. Knoke, D., Yang, S.: Social Network Analysis, 2nd edn. Sage Publications, Inc. (2008)
12. Sabidussi, G.: The centrality index of a graph. Psychometrika 31(4), 581–603 (1966)
13. Newman, M., Barabasi, A.L., Watts, D.J.: The Structure and Dynamics of Networks (Princeton Studies in Complexity). Princeton University Press, Princeton (2006)
14. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proceedings of the National Academy of Sciences of the United States of America 99(12), 7821–7826 (2002)
15. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear and Soft Matter Physics 69(2 pt. 2), 16 (2004)
16. Pothen, A., Simon, H.D., Liou, K.P.: Partitioning sparse matrices with wigenvectors of graphs. SIAM J. Matrix Anal. Appl. 11(3), 430–452 (1990)
17. Fiedler, M.: Algebraic connectivity of graphs. Czechoslovak Mathematical Journal 23, 298–305 (1973)
18. Drázdilová, P., Martinovic, J., Slaninová, K.: Spectral clustering: Left-right-oscillate algorithm for detecting communities. In: ADBIS Workshops, pp. 285–294 (2012)
19. Dreyfus, S.E., Dreyfus, H.L.: A five-stage model of the mental activities involved in directed skill acquisition. Technical report, DTIC Document (1980)
20. Ennis, M.R.: Competency models: a review of the literature and the role of the employment and training administration (ETA). US Department of Labor (2008)
21. Sinnott, G., Madison, G., Pataki, G.: Competencies: Report of the competencies workgroup, workforce and succession planning work groups (September 2002)
22. SFIA Foundation: Framework reference SFIA version 4G (2010)
23. Berry, M.W.: Survey of text mining: clustering, classification, and retrieval, vol. 1. Springer-Verlag New York Inc. (2004)
24. Kuchar, S., Martinovic, J.: Human Resource Allocation in Process Simulations Based on Competency Vectors. AISC, vol. 188. Springer, Heidelberg (2013)
25. Fitzpatrick, E.L., Askin, R.G.: Forming effective worker teams with multi-functional skill requirements. Computers & Industrial Engineering 48(3), 593–608 (2005)
26. Karduck, A., Sienou, A.: Forming the optimal team of experts for collaborative work. In: Bramer, M., Devedzic, V. (eds.) Artificial Intelligence Applications and Innovations, IFIP 18th World Computer Congress, TC12 First International Conference on Artificial Intelligence Applications and Innovations (AIAI 2004), August 22-27, pp. 267–278. Kluwer, Toulouse (2004)
27. Wi, H., Oh, S., Mun, J., Jung, M.: A team formation model based on knowledge and collaboration. Expert Systems with Applications 36(5), 9121–9134 (2009)