An improved attributed graph clustering method for discovering expert role in real-world communities

Isma Hamid
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.
ms_isma222@hotmail.com

Yu Wu
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
wuyu@cqupt.edu.cn

Qamar Nawaz
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
mqamarnawaz@hotmail.com

Runqian Zhao
College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
workzrq@163.com

ABSTRACT

Question-answer communities are expert centric large-scale information sharing platforms where experts can be incorporated directly or discovered from the communities to provide support to the users who are looking for expert advice. Discovering an expert is a complex task that requires interpretive or structural analysis of the community. The interpretive analysis incorporates techniques such as content analysis, surveys, and ethnography to capture the behaviours and interactions within groups. The structural analysis uses formal methods like structure analysis and clustering to identify the important roles in the community. Structural analysis is mostly used for analysing online communities. Most of the existing expert discovery methods use structural analysis without graph attributes. In this paper, we proposed a structural analysis approach to discover expert role in a support-needed-community by utilising graph attributes. The proposed method is developed specifically for exploration and to accomplish visualisation requirements. We discovered the expert role by using threaded question-answer relationships among people of different occupations. Experimental results demonstrate that the proposed method is faster and effectively find an expert

MOBIMEDIA 2017, July 13-14, Chongqing, People's Republic of China

CCS CONCEPTS

• **Information systems applications→ Collaborative and social computing systems and tools**; Social networking sites;

KEYWORDS

Social Networking Sites, Visualization, Expert Role

1 INTRODUCTION

In today's connected world, the Social Networking Sites (SNS's) are successful phenomena [4]. Because of the constant growth of contents on social media, it is sometimes challenging to find out the appropriate information to fulfil the specific material necessity. The required information may be hidden in Question and Answer (Q&A) portals or encapsulated in the web pages [21]. In Q&A portals, users exchange information in the form of questions and answers. Q&A is a type of social media platform, also called communities, where people from different backgrounds interact with each other on different issues. Anxiety and Panic (A&P) Communities, a type of Q&A community, are built to provide social support to patients. The support can be provided either by other users or by experts who try to reduce the patient's panic or anxiety by realising them that they are not alone. These communities can have users of diverse backgrounds, experiences, expertise and social values. Users in a community can interact and behave with each other in different ways. The dominance of one type of behaviour within the community can affect its growth negatively. For example, users can leave a community if there are technical issues, no expert available to answers questions, or mismanagement of contents. Currently, there is inadequate understanding of how these online communities function and what principles are necessary to keep a community's growth healthy. The growth of online communities also depends on some active users. The users who participate more actively, to answer the questions, asked by other users, are called experts [2, 24]. Experts can either be incorporated directly or can be discovered from within the community based on users' behaviour, their social aspects, and their connections to the other community members.

Discovering experts from a Q&A community is a complex task as the expertise varies from field to field and user to user. Interpretive Analysis and Structural Analysis are two different approaches used to identify experts or expert roles. Golder and Donath [15] method is an example of interpretive analysis approach. Interpretive analysis based methods identify roles using techniques such as content analysis, surveys, and ethnography. To understand the structure of the groups and social systems, these methods are commonly used by sociologist and anthropologists. Although, interpretive analysis based methods are popular but it is difficult to identify roles, reproduce contents, and make inter-community comparisons when using these methods. Structural analysis based approaches [3, 7, 10, 17, 22, 34, 35] use formal methods such as structure analysis or clustering to identify the important roles within the community. A set of behavioural features can be identified from social network graph using structural analysis based methods to identify expert roles. In literature, some researchers proposed role identification methods based on the structural analysis procedure to discover expert role by using social network structure without using attributes [2, 9, 13, 26]. Contrarily, some research studies [6, 33] suggest that augmenting network structure with user attributes can deliver fine-grained information of social network structure. In many real-world networks, the graph topological structure and the vertex properties are equally important. For example, in a social network, vertex properties describe the social roles of a user and topological structure characterises the relations among a group of users [12, 16]. Clustering method of the attributed graph, groups the objects based on both structure and attributes to balance the structure and similar properties of the vertices [25]. Currently, experts finding methods have not been thoroughly studied by using social network structure with attributes.

In this paper, an improved attributed graph clustering method is proposed to discover active users or experts in the online community. It does not comprise of any random number nor has the need for any predetermined number of communities. It is developed specifically for community exploration and to accomplish visualisation requirements. The proposed method works by trimming attributed graph to a single weighted graph using Self-Organizing Maps (SOM). Then, the edge weights are modified according to generalised Minkowski distance, and connected subgraphs are calculated for performing community detection procedure. We discovered the expert role by using threaded question-answer relationships among people of different occupations. Our method can detect the number of clusters automatically on the basis of node degree distribution of a given graph. It is particularly beneficial for visualisation and its running time is very less as compared to other relevant algorithms. Experimental results demonstrate that the proposed method is faster and able to effectively find experts in real-world online communities as compared to existing popular methods.

The rest of the paper is organized as follows. Section 2 presents related work and preliminary concepts. Section 3 describes the proposed attributed graph clustering methodology. Section 4 shows experimental results and Section 5 concludes the paper.

2 BACKGROUND STUDY

The social network is not a new research area to study. Since the beginning of the first human society, social networks have been studied to investigate individual and collective human behaviours. In the academic world, research on social networks can be traced back to the first decade of the twentieth century. The most influential early work on social network analysis is the seminal paper Contacts and Influence [28], written in 1950's. As with the development of social networking sites, most of the researchers focused on developing methods for analysing the important people in the social networks. Schwartz and Wood [27] located people by observing communication patterns in e-mail logs. A set of heuristic graph algorithms was used to cluster people by shared interests. Methods of discovering knowledgeable people, about a particular topic, were identified as one of the potential research areas. Campbell et al. [8] analysed the link structure, defined by senders and receivers of emails, by using a modified version of the HITS algorithm [19]. It was used to identify authorities by creating expertise graph that used e-mail headers and from/to fields. Expertise graph contained people as nodes and e-mail messages as edges. Their work established that utilising only the authority scores, from HITS for candidate ranking, resulted in better precision but lower recall. Other sources for constructing social networks includes chat logs [14], community-based question answering systems [1], or co-authorship information from bibliographic databases [32].

Zhang et al. [31] analysed a large, highly specialised help-seeking community. They tried to recognise users with high levels of expertise. The social graph was built from post-reply user interactions with edges directed from questions to answers to reward answering activity. Three measures were compared, including answers/questions ratio. Noll et al. [23] proposed a method which assumed that an expert should be one who had a tendency to identify useful resources before other users discover them. They also applied an HITS-like algorithm to exploit mutual strengthening relationship between documents and users. They distinguished between followers and discoverers who were meant to be experts in this case. Weng et al. [30] proposed TwitterRank, an extension of PageRank algorithm, which was supposed to measure the influence of users on Twitter. The difference of TwitterRank from PageRank is that the random surfer performs a topic-specific random walk via friendship connections, i.e., the transition probability from one Twitterer to another is topic-specific. So, it discovers not only the areas of expertise of Twitter users but also finds experts in these areas. Aslay et al. [2] proposed the competition based expertise network, a unique structure that constructed the expertise network by creating ties between the best answerer and another answerer, combined with graph centrality metrics to identify the experts. Davoodi et al. [13] presented a hybrid method for an expert system that incorporates the features of content-based recommendation algorithms into a social network based collaborative filtering system. Rowe et al. [26] proposed a method to analyse the communities constructed on their role compositions. They proposed a behaviour ontology that captured the user behaviour within the context of time and community. Their method tuned the roles for a given community platform using statistical clustering. Cao et al. [9] used the clustering technique to re-rank experts. Persons were clustered according to their co-occurrences with topics and other persons.

Zhou et al. [33] proposed the incremental clustering algorithm (Inc-Cluster) that incrementally updated the random walk distance matrix in the context of graph clustering with structural and attribute similarities. The problem of this method is its computational time due to overhead of incremental approach. The runtime increases dramatically as the number of clusters increases. Dang and Viennet [12] studied the homophily concept that is the relationship between attribute similarity of users and the topology of social networks. They proposed two approaches that includes SAC1 and SAC2, to extract the communities in an attributed graph. SAC1 approach was based on the modification of Newman's modularity function. Newman's modularity didn't include the attribute similarity, however SAC1 included the attribute similarity. SAC1 used the random walk method to select the communities that leads to its higher running time which is $O(n^2)$. Another approach, SAC2, was divided into two phases. Firstly, it constructed a k-nearest neighbouring graph G_k and tried to find the structural communities in G_k to obtain final clustering. In the second phase, Louvain method [5] was used to find the communities. Cruz et al. [11] proposed a method that combined the information of network topology and network semantic information. According to them, semantic information can be divided into subsets of information called a point-of-view. Their community detection process was divided into two phases. During the first phase, the point-of-view clustered by using Kohonen maps [20] and then fast unfolding algorithm proposed by Blondel et al.[5] was used. The first problem in the Blondel et al. algorithm [5] is that the solution for selecting the nodes as initial community members. Starting with this type of node selection is relatively poor criteria. Secondly, the output of the algorithm depends on the order of the nodes. The ordering of the nodes can also influence the computational time of the algorithm. On sparse social networks, the overall complexity of Cruz method is $O(|Fv*|^3.|v|)$ where $|Fv*|$ is the number of features in the point-of-view.

Our proposed method is different from the above methods because our research aim is not to rank the experts but to find out the important individuals or experts within the online community by using attributed graph clustering.

2.1 Preliminary Concepts

Following definitions are important to understand proposed methodology of attributed graph clustering.

Definition 1: Attribute Augmented Graph. Let $G(V,E,\wedge,F)$ be an undirected or directed graph, where $V=\{v_1,v_2,.....v_n\}$ is a set of N vertices, $E=\{v_i,v_j\}$ is a set of edges, $\wedge=\{a_1,a_2,.....a_m\}$ is a set of m categorical attributes, and $F=\{f_1,f_2,....f_t\}$ is a set of function map of each element in the attribute.

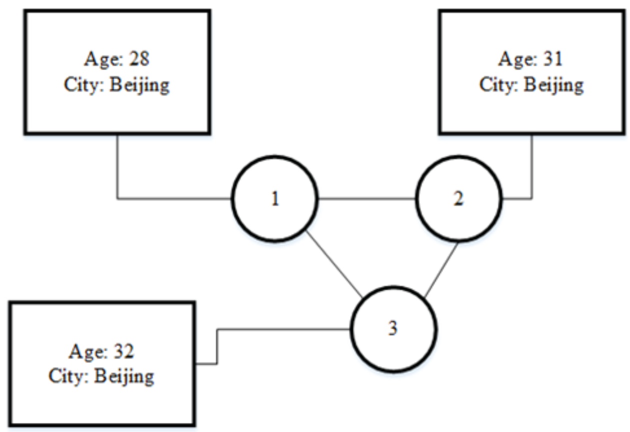
Definition 2: Attributed Graph Clustering. The aim of attributed graph clustering is to partition the vertex set V of an attributed graph G into k disjoint subgraphs $(v_1,v_2,....,v_k)$, where $\bigcup_{i=1}^{k} v_i$ and $v_i \cap v_j = \phi$ for any $i \neq j$. An ideal attributed graph clustering is to generate clusters that have organised intra-cluster structure with homogeneous vertex properties by balancing the structural and attribute similarity.

Definition 3: Node Attributed Graph. The aim of node attributed graph clustering is to detect the groups of nodes that share the common characteristics regarding their attributes and their position in the graph. Formally, node attributed graph can be expressed as a triplet $G=(V,E,F)$, where each node of V is associated with a set of attributes, denoted as a feature vector $[f_1(v),....f_a(v)]$.

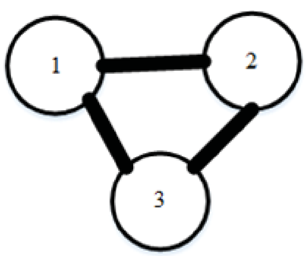
3 PROPOSED ATTRIBUTED GRAPH CLUSTERING METHOD

3.1 Problem and Methodology

Clustering algorithms are commonly used to reduce the visual complexity for exploring a large graph and detecting densely connected subgraphs. Most of the current attributed graph clustering methods do not focus on the visualisation requirements and lack in exploration capability. To explore the behaviour patterns of the experts in a community, we proposed an attributed graph-clustering method. The proposed method is developed for exploration purpose and aims to fulfil visualisation requirements. It trims the node attributed graph to a single weighted graph, as shown in Figure 1(b), where weights represent the attribute similarity. We use SOM algorithm [20] to convert or combine the relevant attributes. Mostly, SOM is used to find the latent information i.e. to create the similarity between nodes. The resultant weighted graph is finally clustered by using improved clustering algorithm. In Figure 2, the general architecture of proposed attributed graph clustering methodology is presented.



(a) A simple node attributed graph



(b) Attribute similarities are stored in edge weights

Figure 1: Simple node attributed graph to attribute similarity Graph

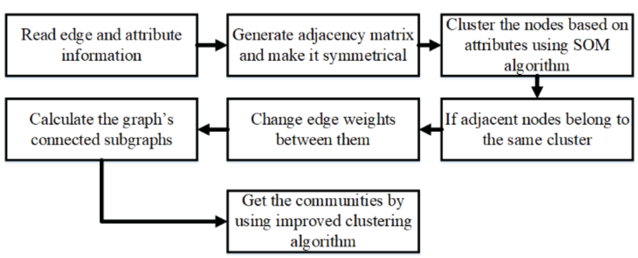


Figure 2: General architecture of proposed attributed graph clustering methodology

Before the execution of clustering method, edges weights are changed according to SOM. For each pair of neighbor vertices $(v_i, v_j \forall i \neq j \in V)$ the weight of the edge $e(v_i, v_j)$ is modified using generalized Minkowski distance. The graph is converted to weighted graph to calculate the modularity by summing the weights. This weighted graph also indicates the nodes that are semantically close to each other therefore, having a stronger connection between them.

Our graph clustering algorithm efficiently determines the number of clusters. In our algorithm, Initial seed nodes are selected having degrees greater than the sum of the average degree and standard deviation of a graph. There may be some nodes remaining in the same cluster after initial seed node selection procedure. To refine the seed nodes, common neighbours are calculated for each pair of seed nodes. If the sum of the degree of these common neighbours is greater than half of the corresponding seed node's degree, then that node is removed from that pair. This procedure is summarised as follows:

- Choose the seed nodes (each seed node represent a cluster)
- Refine the seed nodes.
- Rest of the nodes put into clusters where its neighbour belongs to. Calculate ΔQ using equation (1) and then add it to the cluster where ΔQ is minimized. This step is repeated until each rest of the nodes joins a cluster.

$$\Delta Q = \left[\frac{\Sigma_{bet} + (k_i - k_{i,in})}{\Sigma_{in} + k_{i,in}} - \frac{\Sigma_{bet}}{\Sigma_{in}} \right] \tag{1}$$

where k_i is the sum of the weights of links incident to the node, Σ_{bet} is the sum of the weights of between links, Σ_{in} is the sum of the weights of links inside C and $k_{i,in}$ is the sum of the weights of the links from i to nodes in C.

The overall time complexity of proposed attributed graph clustering method is described in equation (2), that is T_{SOM} + $T_{similarity}$ + $T_{clustering}$:

$$T_{complexity} = O(|F_v|^2 \times n) + O(|E|) + O(|v| \log |v|) \tag{2}$$

3.2 Clustering Algorithm

Input $G = (V, E)$ is an adjacency matrix of a connected graph, $Conn$ = A vector whose elements specifies the connected subgraph membership of each node.

Output: C = An indicator vector whose elements identify the cluster membership of each node.

Steps: For each connected subgraph G' of G

1. Calculate degree of each node in G';
2. Calculate average degree and its standard deviation;
3. Generate initial seed node sets N';
4. Refine seed set N' as the initial cluster member;
5. Calculate the neighbours of each node Neighbour i;
6. for each vertex vi in V', if vi ∈ N', then set Ci ⟵ i, else set Ci ⟵ -1;
7. for each i ∈ V'-N', remove node i from the cluster ci that it belongs to, and move it to its each neighbour node j' cluster cj if cj is not equal to -1, then gain a ΔQcj ;

8. min ΔQ is the minimum value of all ΔQcj , move node i to minΔQ's corresponding cluster. If there is no ΔQcj get in Step 7, then move node i back to ci;

9. Repeat Step 7 and Step 8 until each node belongs to a cluster and is no longer change its cluster.

4 EXPERIMENTS AND RESULTS

In the first part of this experiments section, we have examined the expert's role in social support A&P community. Anxiety and Panic (A&P) community is one of the mental health community of WebMD [29]. As introduced above, this A&P community is a combination of discussion group, Q&A sessions, and has a large and active population. We have manually crawled one-year data from A&P community, comprises of 938 nodes and 1,320 edges. The collected data only covers those people who have received at least one reply or replied to at least one message during the study. This data includes the date when the message was sent, address of the sender, address of replier, subject of the post, and user's profile information e.g., gender, occupation, place lived, age etc. For experiments we divide the user's occupation into eleven broad categories, including: Academics (Professors, Teacher, etc.), Cultural (Actors, Musicians, Philosopher, Writers), Medical experts (physicians, psychologist, etc.), Industry (Finance, Architecture, etc.), Public servants (Firefighters, judges, military officers, etc.), Transport (Air traffic controller, Aircraft Pilots, etc.), Scientists (Astronomers, Biologists, chemist, etc.), Technology and Techniques (Programmers, Web Developers, etc.), Management related (Administration, Business Directors, etc.), Engineers (Electrical, Mechanical, etc.), Other (Students , House Wives, etc.).

Experiments are conducted to explore who are important individuals within the A&P community and how to recognize present experts in this community who are giving social support to patients.

4.1 Clustering and Visualisation

To explore the expert's role, we discover the potential patterns of user's connections among user's occupation by using proposed attributed graph clustering method. By changing the attributes, the proposed methodology can be used to explore the user's potential patterns based on other user's profile attributes such as age, etc. The visualisation results of proposed and Cruz method [11] by using one-year dataset of A&P community present in Figure 3 and Figure 4 respectively, where the categories of occupation are encoded in different colours. ForceAtlas2 layout algorithm proposed by Jacomy [18] is used for getting the visualisation results of proposed and Cruz method. As Figure 3 shown; the graph contains many small and disconnected components. However, a significant number of people are connected with the medical expert community (within the black circle in Figure 3), which is the large connected component. Another large component is public servant community but it is smaller than the medical expert community. This online community is a good example of sparse real-world graph while in many cases real-world social network graphs are sparse. The results show that there exists an expert role in A&P community. In figure 3, the biggest community is "Medical Expert" community.

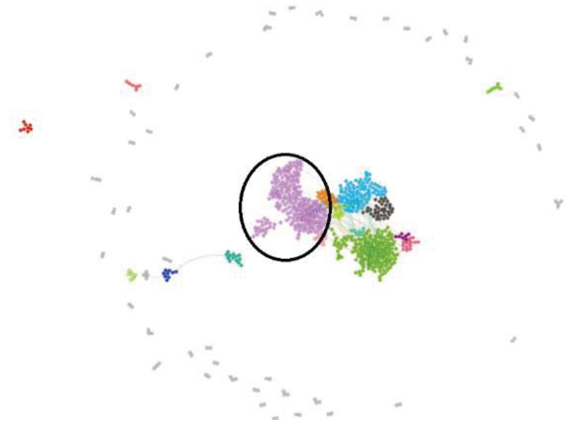


Figure 3: Visualization of A&P community by using attributed graph clustering method

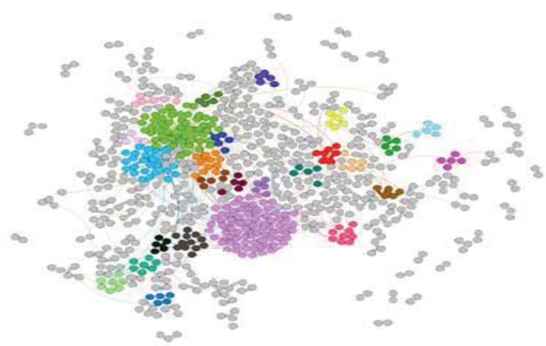


Figure 4: Visualization of A&P community by using Cruz Method

By comparing results presented in Figure 3 and Figure 4, it is clear that our method reduces the visual complexity by detecting connected subgraphs, so it is suitable for exploration of medium and large scale graphs. We also compared the proposed method with Cruz's method by using three quality measure matrices i.e. modularity, density, and entropy. The definition of density that we used is following.

Given a graph $G(V,E)$ where V represents nodes and E represents edges and a partition $C(C_1, C_2, \ldots C_k)$ of G , C represents the communities of the graph. Density is defined as:

$$\text{density}=\delta(C) = \frac{1}{|E|}\sum_{C_i \in c}|E(C_i)| \tag{3}$$

where $E(C_i)$ is the set of edges that start and end within the i^{th} community. Density denotes the ratio of edges lies within the communities. Higher density represents a better clustering. The definition of entropy that we used is the following:

$$\text{entropy}=\mathcal{H}(c)=\frac{1}{|V|}\sum_{C_i\in C}H(C_i) \quad (4)$$

$$H(C_i)=-\sum_{j=1}^{r}p_{ij}\ln p_{ij}+(1-p_{ij})\ln(1-p_{ij})$$

where $H(C_i)$ is the entropy of the i^{th} community, r is the number of attributes and p_{ij} is the proportion of elements in the community, C_i has the same value of the attribute j. The objective of the attribute clustering is to reduce the entropy. According to the quality measure results, shown in Table 1, it is evident that the proposed method performed better than Cruz's method. Specifically, the proposed method showed the better results in terms of the running time and density of the clusters.

Table 1: Comparison of proposed method with Cruz Method

Method	# of Nodes	# of Edges	Running Time(s)	Q	Entropy	Density
Cruz's Method	938	1,320	0.67	0.72	0.10	0.47
Proposed Method	938	1,320	0.36	0.35	0.06	0.74

4.2 Other Real-World Dataset Experiments

For the evaluation of proposed method, it also compared with two state of the art algorithms including SAC1 proposed by Dang and Vinnet [12] and Inc-Cluster method [33] proposed by Zhou et al. Two real-world graph datasets, including Political Blogs and US patent citation are used for comparison of proposed, SAC1 and Inc-Cluster method.

Political Blogs Dataset. The political blog network dataset is a network of 1,490 web blogs on US politics with 19,090 hyperlinks between these web blogs. In the dataset, each blog has an attribute describing its political leaning as either conservative or liberal.

US patent citation Dataset. Patent data from the year 1991-1995 is used. The graph contains 10, 0000 nodes, 18,8631 edges and five attributes. Each patient is represented as nodes in graph and citation between patient are represented as edges of the graph.

In Table 1, experimental results by using A&P community dataset have shown, while in Table 2 the experiments have done by using other two real-world datasets. Table 2 results on the real-world datasets prove that proposed method is a balanced method between attribute and structural similarity. In Table 2 PM indicates (Proposed Method) and Inc(Inc-Clustering).

Table 2: Comparison of proposed method with other relevant algorithms

Network	Running Time			Entropy			Density		
	PM	Inc	SAC1	PM	Inc	SAC1	PM	Inc	SAC1
Political	20.54	65.50	45.8	0.01	0.52	0.06	0.99	0.75	0.91
Patents	1200	5400	6500	0.95	3.12	2.30	0.80	0.52	0.65

5 CONCLUSIONS

In this paper, the work of integrating the structural and attributive data has been thoroughly analysed in the context of expert finding. We used improved attributed-graph-based method for finding expert in A&P community. It has also been evaluated on two large-scale real-world data sets by using three quality matrices. The results showed that proposed method achieves flexibility in combining structural and attributes similarity. Proposed method is compared with three related state-of-the-art algorithms. It is evident from the experimental and comparison results that the proposed method is superior specifically regarding running time and density measure of the clusters.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (61572092).

REFERENCES

[1] Ackerman, M.S., Bakshy, E., Adamic, L.A. and Zhang, J. 2008. Knowledge sharing and Yahoo Answers: Everyone knows something. (2008).

[2] Aslay, Ç., O'Hare, N., Aiello, L.M. and Jaimes, A. 2013. Competition-based networks for expert finding. *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (2013), 1033–1036.

[3] Backstrom, L. and Kleinberg, J. 2014. Romantic partnerships and the dispersion of social ties: a network analysis of relationship status on facebook. *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (2014), 831–841.

[4] Bastos, M.T. 2015. Shares, pins, and tweets: News readership from daily papers to social media. *Journalism Studies.* 16, 3 (2015), 305–325.

[5] Blondel, V.D., Guillaume, J.-L., Lambiotte, R. and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment.* 2008, 10 (2008), P10008.

[6] Bothorel, C., Cruz, J.D., Magnani, M. and Micenkova, B. 2015. Clustering attributed graphs: models, measures and methods. *Network Science.* (2015), 1–37.

[7] Buntain, C. and Golbeck, J. 2014. Identifying social roles in reddit using network structure. *Proceedings of the companion publication of the 23rd international conference on World wide web companion* (2014), 615–620.

[8] Campbell, C.S., Maglio, P.P., Cozzi, A. and Dom, B. 2003. Expertise identification using email communications. *Proceedings of the twelfth international conference on Information and knowledge management* (2003), 528–531.
[9] Cao, Y., Liu, J., Bao, S. and Li, H. 2005. Research on Expert Search at Enterprise Track of TREC 2005. *TREC* (2005).
[10] Chan, J., Hayes, C. and Daly, E.M. 2010. Decomposing Discussion Forums and Boards Using User Roles. *ICWSM*. 10, (2010), 215–218.
[11] Cruz, J.D., Bothorel, C. and Poulet, F. 2011. Semantic Clustering of Social Networks using Points of View. *CORIA* (2011), 175–182.
[12] Dang, T.A. and Viennet, E. 2012. Community detection based on structural and attribute similarities. *International Conference on Digital Society (ICDS)* (2012), 7–12.
[13] Davoodi, E., Kianmehr, K. and Afsharchi, M. 2013. A semantic social network-based expert recommender system. *Applied intelligence*. 39, 1 (2013), 1–13.
[14] Ehrlich, K., Lin, C.-Y. and Griffiths-Fisher, V. 2007. Searching for experts in the enterprise: combining text and social network analysis. *Proceedings of the 2007 international ACM conference on Supporting group work* (2007), 117–126.
[15] Golder, S.A. and Donath, J. 2004. Social roles in electronic communities. *Internet Research*. 5, (2004), 19–22.
[16] Gong, N.Z., Xu, W., Huang, L., Mittal, P., Stefanov, E., Sekar, V. and Song, D. 2012. Evolution of social-attribute networks: measurements, modeling, and implications using google+. *Proceedings of the 2012 ACM conference on Internet measurement conference* (2012), 131–144.
[17] Hautz, J., Hutter, K., Füller, J., Matzler, K. and Rieger, M. 2010. How to establish an online innovation community? the role of users and their innovative content. *System Sciences (HICSS), 2010 43rd Hawaii International Conference on* (2010), 1–11.
[18] Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*. 9, 6 (2014), e98679.
[19] Kleinberg, J.M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*. 46, 5 (1999), 604–632.
[20] Kohonen, T., Kaski, S. and Lappalainen, H. 1997. Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural computation*. 9, 6 (1997), 1321–1344.
[21] Li, X., Liu, Y., Zhang, M., Ma, S., Zhu, X. and Sun, J. 2015. Detecting promotion campaigns in community question answering. *Proceedings of the 24th International Conference on Artificial Intelligence* (2015), 2348–2354.
[22] Nolker, R.D. and Zhou, L. 2005. Social computing and weighting to identify member roles in online communities. *Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence* (2005), 87–93.
[23] Noll, M.G., Au Yeung, C., Gibbins, N., Meinel, C. and Shadbolt, N. 2009. Telling experts from spammers: expertise ranking in folksonomies. *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (2009), 612–619.
[24] Pal, A., Harper, F.M. and Konstan, J.A. 2012. Exploring question selection bias to identify experts and potential experts in community question answering. *ACM Transactions on Information Systems (TOIS)*. 30, 2 (2012), 10.
[25] Parimala, M. and Lopez, D. 2015. Graph clustering based on Structural Attribute Neighborhood Similarity (SANS). *Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on* (2015), 1–4.
[26] Rowe, M., Fernandez, M., Angeletou, S. and Alani, H. 2013. Community analysis through semantic rules and role composition derivation. *Web Semantics: Science, Services and Agents on the World Wide Web*. 18, 1 (2013), 31–47.
[27] Schwartz, M.F. and Wood, D. 1993. Discovering shared interests using graph analysis. *Communications of the ACM*. 36, 8 (1993), 78–89.
[28] de Sola Pool, I. and Kochen, M. 1979. Contacts and influence. *Social networks*. 1, 1 (1979), 5–51.
[29] WebMD - Better information. Better health.: 2016. *http://www.webmd.com/default.htm*. Accessed: 2016-05-05.
[30] Weng, J., Lim, E.-P., Jiang, J. and He, Q. 2010. Twitterrank: finding topic-sensitive influential twitterers. *Proceedings of the third ACM international conference on Web search and data mining* (2010), 261–270.
[31] Zhang, J., Ackerman, M.S. and Adamic, L. 2007. Expertise networks in online communities: structure and algorithms. *Proceedings of the 16th international conference on World Wide Web* (2007), 221–230.
[32] Zhang, J., Tang, J. and Li, J. 2007. Expert finding in a social network. *Advances in Databases: Concepts, Systems and Applications*. Springer. 1066–1069.
[33] Zhou, Y., Cheng, H. and Yu, J.X. 2010. Clustering large attributed graphs: An efficient incremental approach. *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (2010), 689–698.
[34] Zhu, T., Wang, B., Wu, B. and Zhu, C. 2011. Role defining using behavior-based clustering in telecommunication network. *Expert Systems with Applications*. 38, 4 (2011), 3902–3908.
[35] Zhu, T., Wu, B. and Wang, B. 2009. Social influence and role analysis based on community structure in social network. *Advanced Data Mining and Applications*. Springer. 788–795.