I.J. Intelligent Systems and Applications, **2016, 2, 20-27**
Published Online February 2016 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2016.02.03

Extraction of Hidden Social Networks from Wiki-Environment Involved in Information Conflict

Rasim M. Alguliyev
Institute of Information Technology of Azerbaijan National Academy of Sciences
9, B. Vahabzade str., Baku, AZ1141, Azerbaijan
E-mail: r.alguliev@gmail.com

Ramiz M. Aliguliyev and Irada Y. Alakbarova
Institute of Information Technology of Azerbaijan National Academy of Sciences
9, B. Vahabzade str., Baku, AZ1141, Azerbaijan
E-mail: {r.aliguliyev, airada.09} @gmail.com

***Abstract*—**Social network analysis is a widely used technique to analyze relationships among wiki-users in Wikipedia. In this paper the method to identify hidden social networks participating in information conflicts in wiki-environment is proposed. In particular, we describe how text clustering techniques can be used for extraction of hidden social networks of wiki-users caused information conflict. By clustering unstructured text articles caused information conflict we create social network of wiki-users. For clustering of the conflict articles a hybrid weighted fuzzy-c-means method is proposed.

Index Terms—Wiki-technology; wiki-page; conflict articles; information conflict; social network; hybrid weighted fuzzy c-means.

I. Introduction

Wikipedia developing since 2001, contributed to the appearance of new environment on the Internet – Wiki environment. When referring to the biggest search engines of the Internet among the websites proposed by the system the virtual encyclopedia Wikipedia is one of the first places. There are discussions and expressed opinions in the various fields in Wikipedia – political, social, scientific and cultural.

The escalation of information conflicts in wiki-environment and their long-term continuing have a bad impact on the development of Wikipedia projects (wikidictionary, wikicitation, wikibooks, wikisource, wikinews, wikiversity etc.), quality of encyclopedic articles, and respect between users, which is one of the basic principles of Wikipedia philosophy. For elimination of conflicts between users, protection of encyclopedic articles from vandalism, disinformation and propaganda administrators and active users worked out different rules [6].

All contents in Wikipedia projects like text, image, audio- and video- files are added into database of Wikipedia. From this point of view it is possible to control the behavior of wiki-users (discussing of pages with users, editing of particular articles on particular subjects). For defining the quality of encyclopedic articles in wiki-environment and studying the information conflicts problems the identifying of hidden social networks is important issue. According to many specialists dealing with wikimetrics research, events in society, problems and relationships reflect on wiki-environment. The behavior of wiki-users reveals their purposes and the degree of using of information war technologies in their activity [7, 8].

II. Related Work

The analysis of social network in wiki-environment and the reputation of users, also the influence of 0relationships between them on content were studied by different researchers [9, 10]. The model proposed by the professor Sara Javanmardi for identifying social relationships between wiki-users and their influence gives the opportunity for identifying vandals and inexperienced users in wiki-environment [11].

Another approach for measuring the social relationships, conflicts and cooperation in wiki-environment were proposed by Hagit Mesha-Tal and Edna Tal-Alhsid [12]. 3 criteria are used in this method:

1. Number of wiki-users;
2. Interactivity, i.e. the number of edits by wiki-users in particular time interval;
3. Intensity, i.e. the number of changes in wiki-pages as a result of new edits.

Edit means the changes making in wiki-pages, i.e. removing of information, replacing by other information and adding new information. On the basis of the method proposed by Mesha-Tal and Tal-Alhsid the edit warring,

activity of wiki-users and the quality of their edits are defined [13]. During the edit warring, the changes in wiki-pages are not accepted by another group of users. As a result a conflict arises between one or more groups. The information added by one group are removed by another and this process continues some period of time. The existing situation is so: the number of users editing the page is huge, but the changes of the size and the quality of the encyclopedic article are small.

In some wikimetrics researches the influence of wiki-users and the social groups are identified mainly by their behavior. The idea of identifying the vandalism and the analysis of article quality by measuring the influence was firstly proposed by de Adler, Alfaro and Pye [14]. They proposed so-called method "reputation-based system" (Wiki-Trust) for analysis of the quality of encyclopedic articles in wiki-environment and the influence of users [15]. This proposed approach seems very simple from the first view and in some cases looks like the algorithm proposed by Javanmardi: the influence of users depends on the degree of acceptation of their edits [11]. Wiki-Trust was also focused to identify vandalized encyclopedic articles. Here the acts of vandalism are identified by the activity of anonyms (users acting without registration) and new users, who are just registered [16, 17].

However, discussed above methods are not effective in identifying of hidden social networks involved in information conflicts. Thus, proposed methods can be used only if there is action of anonyms and new users in acts of vandalism and information conflicts. But the studies show that not only anonyms and new users involve in information conflicts [15, 18]. Wiki-pages causing wars and conflicts take an attention also of experienced and active users. Taking this fact into account, new approach on identifying the hidden social groups in wiki-environment was proposed. This approach is effective if there is a grouping of wiki-pages by content and the analysis of activity of users involved in creation and editing of these pages.

III. The Model Extraction of Hidden Social Networks

To identify the hidden social networks involved in information conflicts in wiki-environment the phased solution is more advisable (Fig. 1).

To reach the aim first of all articles (wiki-pages) causing information conflicts are defined and grouped by content. Then the users involved in editing of these articles are identified and their activities are analyzed.

The criteria used in this research are following:

- volume of encyclopedic articles (in bytes);
- volume of discussion pages of articles (in bytes);
- first paragraph of the article (in our opinion the first paragraph defines the topic of the article);
- number of users involved in creation and edition of the article;
- number of edits in article;
- number of revertings in article;

"Reverting" is a process removing of edits after some defined time interval [18].

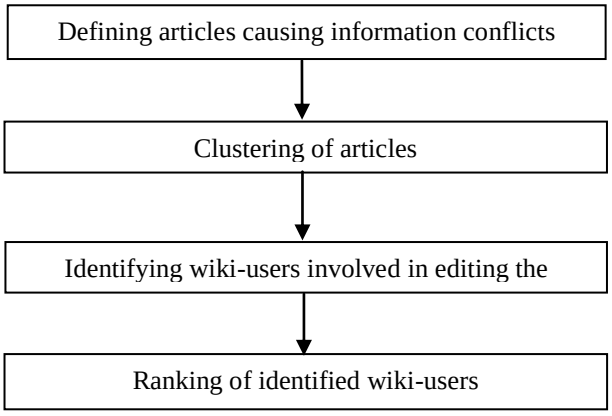

Fig.1. The model of identifying hidden social network in wiki-environment

A. Defining the Articles Causing Information Conflicts

To identify the articles which cause information conflicts in wiki-environment the revert operations in each article, the volume of the article and its discussion page must be taken into account. Thus, an article causes information conflicts, if:

1. The volume of discussion page (twp_n) of encyclopedic article is bigger than the volume of the article (wp_n) itself , $twp_n > wp_n$;
2. The revert operations in article more than three, $rwp_n > 3$, it is considered as anomaly. According to the common rules of Wikipedia the activity of the user blocked after 3 or more revert actions in one day (registration nickname is blocked) [13, 19].

If one of the listed above conditions is met ($twp_n > wp_n$ or $rwp_n > 3$), then the article must be included into the list of the articles which caused information conflict.

B. Clustering of Articles

Clustering is one of the data mining techniques which try to identify groupings of the text documents. Generally speaking, clustering methods attempt to segregate the documents into groups where each group represents some topic that is different than those topics represented by the other groups [20]. It is possible to separate wiki-pages into groups under various conditions using cluster algorithms [21, 22, 23]. In this section, to group the conflict articles we propose a hybrid weighted fuzzy clustering method.

Let's we have the n number of conflict articles and we need to group them by content. For this purpose we use the first paragraphs (stubs) of these articles, which it's considered that first paragraphs define the content of these articles. Let $P = \{P_1, ..., P_n\}$ be the set of first

paragraphs of conflict articles, where P_i denotes the first paragraph of i th conflict article in the collection P, n is the number of conflict articles in collection. Let $T=\{t_1,...,t_m\}$ represents all the distinct terms (words) occurring in P, where m is the number of terms.

Most document clustering methods are relied on the Vector Space Model (VSM). It is a widely used text representation for clustering [24]. According to VSM the document P_i represented as a weighting vector of the terms, $P_i=[w_{i1},...,w_{id}]$, where w_{ij} is the weight of the term t_j in the document P_i. The idea term weighting is to assign a weight to represent the importance of a term. Different weighting schemes are available. The common and popular one is the Term Frequency–Inverse Document Frequency (TF-IDF) weighting scheme.

This scheme combines the definitions of term frequency and inverse sentence frequency, to produce a composite weight for each term in each sentence. This weighting scheme assigns to term a weight in document given by

$$w_{ij}=IF_{ij}\times IDF_j;\ i=1,...,n,\ j=1,...,m, \tag{1}$$

where IF_{ij} is the term frequency and IDF_j is the inverse document frequency of term t_j over the collection P.

The term frequency is calculated as the ratio of number of times the term ccurs in the document to the total number of terms in the documnet. It measures the importance of a term within a document:

$$IF_{ij}=\frac{m_{ij}}{m_i};\ i=1,...,n,\ j=1,...,m, \tag{2}$$

where m_{ij} is the number of occurrence of term t_j in document P_i, m_i is the number of terms in document P_i. This formula assigns a higher weight to terms that occur often in a document.

The IDF measures the importance of a term within the document collection. It is obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient:

$$IDF_j=\log(n/n_j), \tag{3}$$

where n is the number of documents in the collection P and n_j denotes the number of documents in which term t_j appaers.

The IDF factor accounts for the global weighting of term t_j. The IDF factor has been introduced to improve the discriminating power of terms in the traditional information retrieval. A term that occurs in every document of the collection gets a lower IDF value. This reflects the fact that it is not as significant for the distinction between documents as the terms that rarely occur throughout the document collection.

Once each document vector is represented, in this study the Euclidean distance is used to calculate the distance between pair of documents. Then the Euclidean distance between vectors $P_i=[w_{i1},...,w_{im}]$ and $P_l=[w_{l1},...,w_{lm}]$ is calculated as:

$$d_{il}=\|P_i-P_l\|=\sqrt{\sum_{j=1}^{m}(w_{ij}-w_{lj})^2},\ i,\ l=1,...,n. \tag{4}$$

The fuzzy clustering algorithm allows one piece of data to belong to more one cluster according to a membership function. The value of a fuzzy membership function belongs to any number between 0 and 1, and is meant to be a mathematical characterization of a "set" which may not be precisely defined [25].

Let $U=\left[u_{ik}\right]_{n\times c}$ be a partition matrix where u_{ik} is the membership value of P_i belonging to class C_k, and $V=\{V_1,...,V_c\}$ is a set of cluster centers, where $V_k\in R^m$, $V_k=(v_{k1},...,v_{km})$.

The goal of clustering is to assign data points P_i ($i=1,...,n$) into c partitions. Assume that the c centers are $V_1,...,V_c$ and in cluster C_k there exist N_k points. So we can calculate its center by averaging its members:

$$V_k=\frac{1}{N_k}\sum_{i=1}^{N_k}S_i,\ k=1,...,c. \tag{5}$$

In this paper, a novel hybrid weighted fuzzy c-means (HWFCM) clustering method is proposed which integrates the properties of the density weighted fuzzy c-means and the cluster-dependent fuzzy c-means clustering methods [26, 27, 28, 29]:

$$F_{HWFCM}(U,V)=\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_{ik}u_{ik}^{\mu}d_{ik}^2+ \\ +\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_i u_{ik}^{\mu}d_{ik}^2+\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_k u_{ik}^{\mu}d_{ik}^2\rightarrow\min \tag{6}$$

Subject to

$$0\le u_{ik}\le 1,\ i=1,...,n;\ k=1,...,c, \tag{7}$$

$$\sum_{k=1}^{c}u_{ik}=1,\ i=1,...,n, \tag{8}$$

$$0<\sum_{i=1}^{n}u_{ik}<n,\ k=1,...,c, \tag{9}$$

where $\mu>1$ is the degree of fuzziness associated with the partition matrix. If we consider μ to be one, the soft clustering will be changed into hard one. Let u_{ik} satisfy the above conditions Eqs.(7)-(9) represented by a $n\times c$

matrix $U=[u_{ik}]$. The proposed clustering method aims to determine cluster centers V_k and fuzzy partition matrix U by minimizing the objective function $F_{HWFCM}(U,V)$. The method provides the fuzzy membership matrix U and the fuzzy cluster center vector $V=[V_k]$.

d_{ik} is the Euclidean distance from document vector $P_i=[w_{i1},...,w_{im}]$ to the cluster center $V_k=[v_{k1},...,v_{km}]$

$$d_{ik}=\|P_i-V_k\|=\sqrt{\sum_{j=1}^{m}(w_{ij}-v_{kj})^2}\ ,\ i=1,...,n;\ k=1,...,c. \quad (10)$$

α_{ik} is a weighting of P_i belonging to cluster C_k. It is the cluster-dependent weight, which can change and be updated during the clustering process. The weights α_i (α_i is a density measurement of input data) are independent of a particular cluster and are constants during the clustering process. α_k is the weight which depends on cluster cardinality. Contrary to α_i, the weights α_k can change during the clustering process. For computing the weights α_{ik}, α_i and α_k, we utilize the following equations [27, 30, 31, 32]:

$$\alpha_{ik}=\left(\frac{1}{d_{ik}}\right)^{\frac{2}{\eta-1}},\ i=1,...,n;\ k=1,...,c, \quad (11)$$

$$\alpha_i=\left(\frac{d_i}{\sum_{p=1}^{n}d_p}\right)^{\frac{2}{\beta-1}},\ i=1,...,n, \quad (12)$$

$$\alpha_k=\left(\frac{1}{n_k}\right)^{\frac{1}{\gamma-1}},\ k=1,...,c. \quad (13)$$

α_{ik} is a weight indicating the importance of distance between the data P_i and cluster center V_k. The difference between α_{ik} and u_{ik} obtained by FCM algorithm is that there is no limitation to "$\sum_{i=1}^{n}\alpha_{ik}=1\ \forall\ k$ and $\sum_{k=1}^{c}\alpha_{ik}=1\ \forall\ i$." So, the α_{ik} has more representative than u_{ik} in reflecting correlation between data and clusters. Especially, the noises do not need to satisfy the limitation. Hence, the influence of noises is reduced by α_{ik} [26].

Here $\eta>1$ is a parameter depending on the variation of outliers, $\beta>1$ and $\gamma>1$ are the user-defined parameters, $\overline{V}$ is the center of input data and d_i is the distance between data point S_i and the center $\overline{V}$, $d_i=\|S_i-\overline{V}\|$, where $\overline{V}=\frac{1}{n}\sum_{i=1}^{n}S_i$.

Using the Lagrange multiplier method, the problem is equivalent to minimizing the following equation satisfying the constraint Eq.(8):

$$L(U,V,\lambda)=\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_{ik}u_{ik}^{\mu}d_{ik}^2+\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_i u_{ik}^{\mu}d_{ik}^2+ +\sum_{k=1}^{c}\sum_{i=1}^{n}\alpha_k u_{ik}^{\mu}d_{ik}^2+\sum_{i=1}^{n}\lambda_i\left(1-\sum_{k=1}^{c}u_{ik}\right) \quad (14)$$

For the sake of simplicity in computations, we use an assumption that $\partial\alpha_{ik}/\partial v_k=0$, $\partial\alpha_i/\partial v_k=0$ and $\partial\alpha_k/\partial v_k=0$. Therefore, setting $\partial L/\partial u_{ik}=0$, we will obtain the following equation for u_{ik}:

$$\frac{\partial L(U,V,\lambda)}{\partial u_{ik}}=\mu\cdot\alpha_{ik}u_{ik}^{\mu-1}d_{ik}^2+\mu\cdot\alpha_i u_{ik}^{\mu-1}d_{ik}^2+ +\mu\cdot\alpha_k u_{ik}^{\mu-1}d_{ik}^2-\lambda_i=0\Rightarrow u_{ik}^{\mu-1}\mu(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2= =\lambda_i\Rightarrow$$

$$u_{ik}=\left(\frac{\lambda_i}{\mu(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}. \quad (15)$$

Replacing u_{ik} in Eq. (8) we obtain:

$$\sum_{k=1}^{c}\left(\frac{\lambda_i}{\mu(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}=1\ \Rightarrow$$

$$\left(\frac{\lambda_i}{\mu}\right)^{1/(\mu-1)}=\frac{1}{\sum_{k=1}^{c}\left(\frac{1}{(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}}. \quad (16)$$

Further replacing Eq.(16) in Eq.(15), u_{ik} can be rewritten as follows:

$$u_{ik}=\left(\frac{\lambda_i}{\mu(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}=\left(\frac{\lambda_i}{\mu}\right)^{1/(\mu-1)}\cdot \cdot\left(\frac{1}{(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}=\frac{1}{\sum_{q=1}^{c}\left(\frac{1}{(\alpha_{iq}+\alpha_i+\alpha_q)d_{iq}^2}\right)^{1/(\mu-1)}}\cdot \cdot\left(\frac{1}{(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}\right)^{1/(\mu-1)}=\frac{1}{\sum_{q=1}^{c}\left(\frac{(\alpha_{ik}+\alpha_i+\alpha_k)d_{ik}^2}{(\alpha_{iq}+\alpha_i+\alpha_q)d_{iq}^2}\right)^{1/(\mu-1)}} \quad (17)$$

Furthermore, by setting $\partial L/\partial V_k=0$, we obtain the following updating equation for the centroids:

$$\frac{\partial L(U,V,\lambda)}{\partial V_k}=0\Rightarrow-2\sum_{i=1}^{n}\alpha_{ik}u_{ik}^{\mu}(S_i-V_k)-2\sum_{i=1}^{n}\alpha_i u_{ik}^{\mu}(S_i-V_k)- -2\sum_{i=1}^{n}\alpha_k u_{ik}^{\mu}(S_i-V_k)=0\ \Rightarrow$$

$$\Rightarrow V_k\sum_{i=1}^{n}(\alpha_{ik}+\alpha_i+\alpha_k)u_{ik}^{\mu}=\sum_{i=1}^{n}(\alpha_{ik}+\alpha_i+\alpha_k)u_{ik}^{\mu}S_i\ \Rightarrow$$

$$V_k = \frac{\sum_{i=1}^{n}(\alpha_{ik}+\alpha_i+\alpha_k)u_{ik}^{\mu}S_i}{\sum_{i=1}^{n}(\alpha_{ik}+\alpha_i+\alpha_k)u_{ik}^{\mu}}. \quad (18)$$

Determination of the optimal number of clusters in a data set is a difficult issue and depends on the adopted validation and chosen similarity measure, as well as on data representation. For clustering of documents, customers can't predict the latent topic number in the document collection, so it's impossible to offer the number c of clusters effectively. The strategy that we used to determine the optimal number of clusters (the number of topics in a document collection) is based on the distribution of words in the documents [33]:

$$c = n\frac{\left|\bigcup_{i=1}^{n}P_i\right|}{\sum_{i=1}^{n}|P_i|}, \quad (19)$$

where $|P|$ denotes the number of terms in the document P.

Based on Eqs.(4)-(19) we describe the main steps of the proposed HWFCM algorithm as follows:

Step 1: By using Eq.(19) define the number c of clusters. HWFCM then randomly initializes the centroids. Let $\beta = 2$ Compute α_i $(i = 1,...,n)$ according to Eq.(12). Choose a threshold value ε. Let $\mu = 2$. Initialize the fuzzy partition matrix U by generating $c \times n$ random numbers in the interval $[0,1]$.

Step 2: Let $\eta = \gamma = 2$. Compute α_{ik} and α_k ($i = 1,...,n$; $k = 1,...,c$) by using Eqs.(11) and (13), respectively.

Step 3: Compute V_k according to Eq.(18).

Step 4: Compute all d_{ik} according to (10) and then all u_{ik} $(i = 1,...,n; k = 1,...,c)$ according to (17). Thus update the fuzzy partition matrix U by the new computed u_{ik}.

Step 5: Compute the objective function F_{HWFCM} by using Eq.(6). If it converges or the difference between two adjacent computed values of objective function F_{HWFCM} is less than the given threshold ε then stop. Otherwise go to step 2.

C. Identifying and Ranking of Wiki-Users Involved in Editing the Articles

After identifying of articles which cause information conflicts and their separation into groups the problem of identifying of social networks involved in the creation and editing of these articles must be solved.

The stage after clusterisation is identifying and ranking of users of social networks involved in editing the articles. For this purpose we identify users involved in the editing of one article, then users involved in the editing the articles within each cluster, and finally the users involved in the editing the articles collected in all clusters. In this task the theory of graphs is used.

Relations in cluster can be regarded as a social network, and as it be shown in Fig.2 each article has been described as a graph: *S(U, E)*, where *U* – set of vertices, *E* – set of edges that describe the interaction of agents. Let's imagine the article as an agent of the user group. The relations between agents are provided by users. Let,

- $A^k = \{A_1^k, A_2^k, ...A_{m_k}^k\}$ – the set of articles in cluster C_k, where m_k – the number of articles in cluster C_k, $k = 1,\ldots,c$;
- $U\left(A_j^k\right) = \left\{U_{j,1}^k, U_{j,2}^k, ..., U_{j,n_k}^k\right\}$ – the group of users who edited the article A_j^k in cluster C_k; $j = 1,...,m_k$; $k = 1,\ldots,c$;
- $U\left(C_k\right) = \left\{U_1, U_2, ..., U_{n_k}\right\}$ – the group of users who edited articles in cluster C_k; n_k – the number of users in cluster C_k. It is clear that, $U(C_k) = \bigcup_{j=1}^{m_k} U(A_j^k)$.

Then,

- $Q_{jl}^k = U(A_j^k) \cap U(A_l^k)$ – the group users who edit articles $A_j^k \in C_k$ and $A_l^k \in C_k$; $l \neq j = 1,..., m_k$; $k = 1,\ldots,c$;
- $R_{kp} = U(C_k) \cap U(C_p)$ – the group of users who edited the articles in cluster C_k and C_p.; $k \neq p = 1,..., c$;
- $P_{jl}^{k,p} = U(A_j^k) \cap U(A_l^p)$ – the group of users who edited the articles $A_j^k \in C_k$ and $A_l^p \in C_p$; $p \neq k = 1,..., c$, $j = 1,..., m_k$, $l = 1,..., m_p$;

If $Q_{jl}^k \neq \varnothing$, then there is would be edge (internal) between the majority of elements (users) Q_{jl}^k. If $P_{jl}^{k,p} \neq \varnothing$, then there is would be edge (external) between the majority of elements (users) $P_{jl}^{k,p}$.

As can be seen there are two types of contact between the elements of network: internal and external. To set up internal (topic) network internal edges are used. The edge weight in the topic network is assigned by the numbers of edges between the tops [34].

The following formula is used to rank users in the topic network:

$$\omega_i^k = \sum_{\substack{s=1 \\ s\neq i}}^{n_k} \omega_{is}^k . \quad (20)$$

Here,

$$\omega_{is}^k = \sum_{j=1}^{m_k-1} \sum_{l=j+1}^{m_k} I(U_i \in Q_{jl}^k \,\&\, U_s \in Q_{jl}^k) \quad (21)$$

Where

$$I(x) = \begin{cases} 1, \ if \ \ x \ true \\ 0, \ otherwise \end{cases} \quad (22)$$

Users in the topic network are ranked by the value of ω_i^k . General social network can be gotten by the synthesis of topic networks. This network is synthesised by the following way:

- to assign the group of users involved in the all topic networks. For this the majority, $U(C) = \bigcap_{k=1}^{c} U(C_k)$ is found and edges between them is calculated. Two tipes of this edges can be: internal and external. Edges weight between the two tops (users) are gotten as the sum of internal and external edges;
- after that the weight of each vertex in the network is found. The weights of the vertexes (users) are calculated as the sum of the weights of edges connecting to other tops. The users are ranked according to their weight.

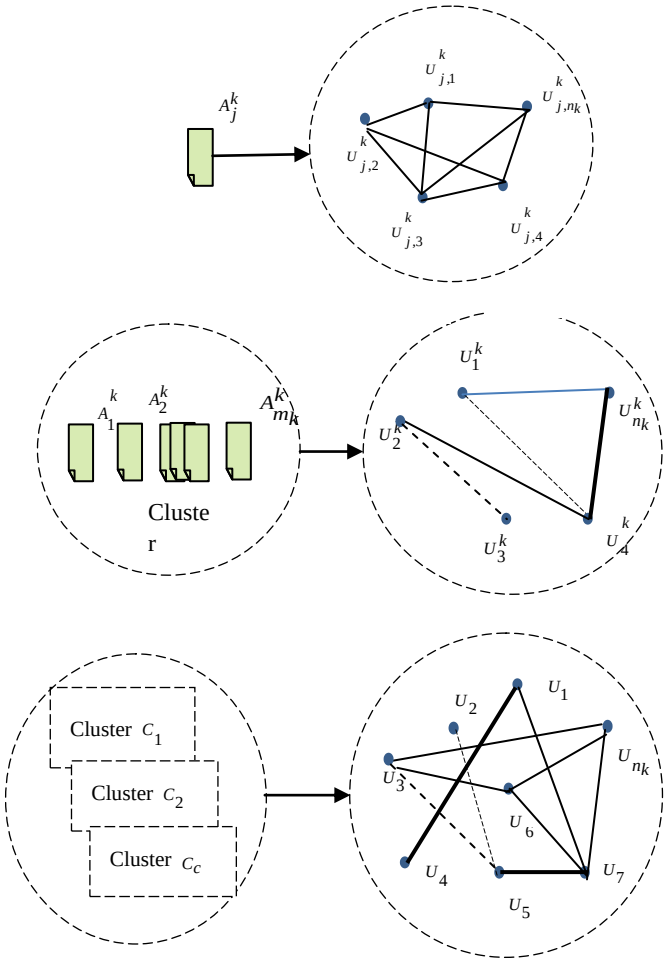

Fig.2. Social networks of wiki-users involved in editing the article, cluster and all articles

IV. Conclusion

The proposed approach for identification of the hidden social network controling encyclopedic articles causing conflicts, can be helpfull in preventing information conflicts in virtual encyclopedia Wikipedia and in solution of the problem of information security in wiki-environment. By identifying articles, which are used in information conflicts, and groups, which propagand some ideology, it is possible to increase the quality of encyclopedic articles, provide the Wikipedia with correct and independ information.

The approach proposed for identifying the hidden

social networks is not only for wiki-technologies but may be used also for analys many social networks based on web2.0 technology.

REFERENCES

[1] P. Shachaf, N. Hara, "Beyond vandalism: wikipedia trolls". Journal of Information Science, vol. 36, no. 3, 2010, pp. 357-370.

[2] T. Yasseri, R. Sumi, A. Rung, A. Kornai, J. Kertész, "Dynamics of conflicts in wikipedia". PLoS ONE, vol. 7, no. 6, 2012, e38869.

[3] A. G. West, S. Kannan, I. Lee, "STiki: an anti-vandalism tool for wikipedia using spatio-temporal analysis of revision metadata", in Proceedings of the 6th International Symposium on Wikis and Open Collaboration, New York, ACM, 2010, pp. 47-48.

[4] B. Leuf, W. Cunningham, "The wiki way: quick collaboration on the web", Laflin, PA: Addison-Wesley, 2001, 200 pp.

[5] B. Luyt, D. Tan, "Improving wikipedia's credibility: references and citations in a sample of history articles", American Society for Information Science and Technology, Vol. 61, No. 4, 2010, pp. 715-722.

[6] T. Iba, K. Nemoto, B. Peters, P. A. Gloor, "Analyzing the creative editing behavior of wikipedia editors: through dynamic social network analysis", Procedia – Social and Behavioral Sciences, vol. 2, no. 4, 2010, pp. 6441-6456.

[7] T. Holloway, M. Božicevic, K. Börner, "Analyzing and visualizing the semantic coverage of wikipedia and its Authors", Journal Complexity, Vol. 12, No. 3, 2007, pp. 30-40.

[8] N. Hara, P. Shachaf, K. F. Hew, "Cross-cultural analysis of the wikipedia community" American Society for Information Science and Technology, vol. 61, no. 10, 2010, pp.2097-2108.

[9] J. Moskaliuk, J. Kimmerle, U. Cress, "Collaborative knowledge building with wikis: the impact of redundancy and polarity", Computers & Education, 2012, vol. 58, No. 4, pp. 1049-1057.

[10] S. Javanmardi, C. Lopes, P. Baldi, "Modeling user reputation in wikis", Statistical Analysis and Data Mining, vol. 3, no. 2, 2010, pp. 126-139.

[11] S. Javanmardi, D. W. McDonald, C. V. Lopes, "Vandalism detection in wikipedia: a high-performing, feature-rich model and its reduction through lasso", in Proceedings of the 7th International Symposium on Wikis and Open Collaboration, New York, ACM, 2011, pp. 82-90.

[12] H. Meishar-Tal, E. Tal-Elhasid, "Measuring collaboration in educational wikis – a methodological", Emerging Technologies in Learning, vol. 3, 2008, pp. 46–49.

[13] http://en.wikipedia.org/wiki/Wikipedia:Edit_warring

[14] B.T. Adler, L. de Alfaro, I. Pye, "Detecting wikipedia vandalism using wikitrust", Lab Report for PAN at CLEF, 2010, (http://ceur-ws.org/Vol-1176/CLEF2010wn-PAN-AdlerEt2010.pdf).

[15] B. T. Adler, L. de Alfaro, "A content-driven reputation system for the wikipedia", in Proceedings of the 16th International Conference on World Wide Web, New York, ACM, 2007, pp. 261-270.

[16] B. T. Adler, L. de Alfaro, S. M. Mola-Velasco, P. Rosso, A. G. West, "Wikipedia vandalism detection: combining natural language, metadata, and reputation features", in Proceedings of the 12th International Conference on Intelligent Text Processing and Computational Linguistics, Berlin, Springer-Verlag, 2011, pp. 277-288.

[17] L. de Alfaro, A. Kulshreshtha, I. Pye, B. T. Adler, "Reputation systems for open collaboration", Communications of the ACM, vol. 54, no. 8, 2011, pp. 81–87.

[18] I. Y. Alakbarova, "Analysis factors influencing on ranking of papers in wiki environment". Problems of Information Society, vol. 2, no. 6, 2012, pp. 27-32. (in Azerbaijani)

[19] http://en.wikipedia.org/wiki/Wikipedia: Reviewing

[20] S. Shedata, F. Karray, M. Kamel, "An efficient-based mining model for enhancing text clustering", IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 10, 2010, pp. 1360-1371.

[21] R. M. Alguliev, R. M. Aliguliyev, I. Y. Alekperova, "Cluster approach to the efficient use of multimedia resources in information warfare in wikimedia", Automatic Control and Computer Sciences, vol. 48, no 2, 2014, pp. 97-108.

[22] A. Skabar, K. Abdalgader, "Clustering sentence-level text using a novel fuzzy relational clustering algorithm", IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 1, 2013, pp. 62-75.

[23] N. Ye, S. M. Emran, Q. Chen, S. Vilbert, "Multivariate statistical analysis of audit trails for host-based intrusion detection", IEEE Transactions on Computers, vol. 51, no. 7, 2002, pp. 810-820.

[24] J. Jayabharathy, S. Kanmani, "Correlated concept based dynamic document clustering algorithms for newsgroups and scientific literature", Decision Analytics, vol. 1, no. 3, 2014, pp. 1-21.

[25] H. L. Shieh, "A hybrid fuzzy clustering method with a robust validity index", Fuzzy Systems, vol. 16, no.1, 2014, pp. 39-45.

[26] J. L. Chen, J. H. Wang, "A new robust clustering algorithm-density-weighted fuzzy c-means", in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, vol. 3, 1999, pp. 90-94.

[27] A. H. Hadjahmadi, M. M. Homayounpour, S. M. Ahadi, "Bilateral Weighted Fuzzy C-Means Clustering", Iranian Journal of Electrical & Electronic Engineering, 2012, Vol. 8, No. 2, pp. 108-121.

[28] M. Nazari, J. Shanbehzadeh, A. Sarrafzadeh, "Fuzzy C-Means Based on Automated Variable Feature Weighting", in Proceedings of the International MultiConference of Engineers and Computer Scientists, 2013, pp.1-5.

[29] N. R. Pal, K. Pal, J. C. Bezdek, "A mixed c-means clustering model", in Proceedings of the IEEE International Conference on Fuzzy Systems, vol. 1, 1997, pp.11-21.

[30] R. M. Aliguliyev, "Performance evaluation of density-based clustering methods", Information Sciences, vol. 179, no. 20, 2009, pp. 3583-3602.

[31] R. M. Aliguliyev, "Clustering of document collection – a weighting approach", Expert Systems with Applications, vol. 36, no. 4, 2009, pp. 7904-7916.

[32] M. El. Agha, W.M. Ashour, "Efficient and fast initialization algorithm for k-means clustering", International Journal of Intelligent Systems and Applications, vol.4, no.1, 2012, pp. 21-31.

[33] R. M. Aliguliyev, "A new sentence similarity measure and sentence based extractive technique for automatic text summarization", Expert Systems with Applications, vol. 36, no. 4, 2009, pp. 7764-7772.

[34] P. Wadhwa, M.P.S. Bhatia. Discovering hidden networks in on-line social networks // International Journal of Intelligent Systems and Applications, vol. 6, no. 5, 2014, pp.44-54.

Authors' Profiles

Rasim M. Alguliyev. He is director of the Institute of Information Technology of Azerbaijan National Academy of Sciences (ANAS) and academician-secretary of ANAS. He is professor and full member of ANAS. His research interests include: Information Security, E-government; Information Society, Social Network Mining and Analysis, Cloud Computing, Evolutionary and Swarm Optimization, Data Mining, Text Mining, Web Mining, Social Network Analysis, Big Data Analytics, Scientometrics and Bibliometrics.

Ramiz M. Aliguliyev. He is head of department at the Institute of Information Technology of ANAS. His research interests include: Data Mining, Text Mining, Web Mining, Social Network Analysis, Evolutionary and Swarm Optimization, Big Data Analytics, and Scientometrics.

Irada Y. Alakbarova. She is head of sector at the Institute of Information Technology of ANAS. Her research interests include: Information War, Wiki-technology, Wikimetrics, Data Mining and Social Networks Analysis.

