Extracting social networks from e-government by sentiment analysis of users' comments

Rasim M. Alguliyev, Ramiz M. Aliguliyev* and Gunay Y. Niftaliyeva

Institute of Information Technology,
Azerbaijan National Academy of Sciences,
9A, B. Vahabzade Street,
AZ1141, Baku, Azerbaijan
Email: r.alguliev@gmail.com
Email: r.aliguliyev@gmail.com
Email: gunayniftali@gmail.com
*Corresponding author

Abstract: Nowadays, the improvement of governance, ensurance the security and the timely detection of propaganda against the government are major problems of e-government. Extraction of hidden social networks is one of the most actual problems in the term of government security. The extraction of hidden social networks operating against the state in e-government is one of the key factors to ensure the security in e-government. In the paper, a method has been proposed for extracting hidden social networks to improve management in e-government, prevent promotion against the government and ensure the security. In this approach, hidden social networks are extracted through the analysis of user's comments via opinion and text mining technologies.

Keywords: e-government; social network extraction; SONEX; social network analysis; text mining; opinion mining; sentiment analysis.

Reference to this paper should be made as follows: Alguliyev, R.M., Aliguliyev, R.M. and Niftaliyeva, G.Y. (2019) 'Extracting social networks from e-government by sentiment analysis of users' comments', *Electronic Government*, Vol. 15, No. 1, pp.91–106.

Biographical notes: Rasim M. Alguliyev is the Director of the Institute of Information Technology of Azerbaijan National Academy of Sciences (ANAS) and Academician-Secretary of ANAS. He is a full member of ANAS and a Full Professor. He received BSc and MSc in Electronic Computing Machines from Azerbaijan Technical University in 1979. He received his PhD and Doctor of Sciences (higher degree after PhD) in Computer Science in 1995 and 2003, respectively. His research interests include information security, information society, e-government; online social network, cloud computing, evolutionary and swarm computation, data mining, big data analytics, and scientometrics. He has authored more than 580 papers, four monographs, four patents, and several books.

Ramiz M. Aliguliyev is the Head of Department at the Institute of Information Technology of ANAS. He received BSc and MSc in Applied Mathematics from Baku State University (Azerbaijan) in 1983. He received his PhD (2002) in Mathematics and Doctor of Sciences (higher degree after PhD) in Computer Science (2011). His research interests include text mining, big data analytics, evolutionary and swarm computation, e-government, online social networks and scientometrics. He has authored more than 170 papers and four books.

Gunay Y. Niftaliyeva received his BSc and MSc in Applied Mathematics from Baku State University (Azerbaijan) in 2012 and 2014, respectively. She is currently a PhD student at the Institute of Information Technology of ANAS. Her research interests include e-government, online social networks and text mining. She has authored 11 papers and one book.

1 Introduction

The government is a political organisation, management process of the citizens living within certain borders. Today there are more than 200 countries in the world. These countries differ from each other according to the administration form, internal structure, and economic prosperity, and so on. The main goal of the government is to provide citizens' satisfaction and to improve their living standards.

In modern era the humanity is under the influence of rapid development of information and communication technologies (ICT). ICT is one of the main factors that affect to the structure, work principle, activity, and changes of organisations. The rapid development of ICT and the internet opens new possibilities for the relationship among citizens, governments, and institutions. The government has passed to the e-government era using the ICT in order to fulfil its functions more advanced and ensure transparency. The main difference between the classic government and e-government is the services provided faster, be more efficient in terms of time and finance.

The rapid development of the internet at the same time contributes the wide spread of social networks in the web. The wide use of social media tools among citizens makes the government agencies to think about how to take an advantage by joining to these sites. Several studies related with e-government services are studying the relation networks between the people, businesses and government. E-government sites are considered as public sites and citizens who are able to access this sites, directly take part in the decision making process of government.

Nowadays, the improvement of governance, ensurance the security and the timely detection of propaganda against the government are major problems of e-government. The extraction of hidden social networks operating against the state in e-government is one of the key factors to ensure the security in e-government. It is known that the users are able to respond to any information via their comments in e-government. These comments should be analysed to determine whether is there any criminal groups and propaganda behind these comments against the government or not. At present, text mining is one of the most advanced and efficient technologies in the texts analysis.

As seen, the extraction of hidden social networks is one of the most actual problems in the term of government security. Based on the urgency of the problem, this study proposes an approach for extracting of hidden social networks from the e-government. The paper is structured as follows: the role of social networks in the management of e-government is given in Section 2. Section 3 discusses social network extraction (SONEX) methods in different environments and in the Section 4 an approach to extract the hidden social networks from e-government environment is proposed. A conclusion and future work are provided in the Section 5.

2 The role of social networks in the management of e-government

E-government is defined in the literature as "the communication of the citizens and institutions with the government through the various electronic mass media (fax, smart-cards, e-mail/internet, etc.)" (Akbaş, 2012). Depending on whom the services are offered the direction of e-government services are gathered into several groups. Such groups are included the citizens, government agencies, public and political organisations, institutions, non-profit organisations.

The use of ICT significantly influences on people's lifestyle and the communication between them. Such a significant influence reveals not only on the work and activities of the people, but also in the functions of the government itself. Social media is one of the main factors affecting the government. The extensive use of social media sites and tools by the private sector, public organisations and people makes the government to think about how to take an advantage from this to rebuild the relationship with their citizens and increase the level of citizens' participation in decision making process. This is very important in the adoption of e-government websites as a social sites and increasing number of citizens who use it. Social networks are the sites where the people interact with each other and such sites can bring the citizens closer.

Social media provides transparency in the activity of government agencies by presenting an opportunity to citizens with better service and information access, opening them an active communication channel, and finally, involving them to the management process (Alguliev et al., 2011). Social media helps to provide an open, transparent and interactive relation between the government and citizens. In order to create such a relationship with the citizens the government have to think carefully about how to manage social media sites, and how to use this channel effectively. It is possible to manage social media sites using different approaches. For example, the 'four R' approach is based on four main idea (Alguliev et al., 2012a):

1 to know what people say online, and to focus on monitoring

2 to be sure of the social media site being an active communication channel

3 to send rich and high quality content

4 to benefit from the similar situation directing the people with similar problems to each other.

If the government uses the social media to provide the citizens with the active information, then the government' existence in the social media sites is very important, in order to gain the confidence of the citizens by showing that the government not only who listens and controls, but also who responds. Social media sites also provides an opportunity to take advantage of more creative and innovative ways and to use effectively valuable resources that are available on these sites in the communication with the citizens.

The literature introduced three types of 'digital citizenship' in e-government (Alguliev et al., 2011, 2012a). The first type includes the citizens, who uses the e-mail instead of paper mail, benefits from government's appearance on social media sites in order to communicate and access real-time information. This type of citizenship is called personally responsible digital citizenship. The second type includes participatory digital citizenship that benefits from online discussion forums through social media sites. The

third type involves the digital citizenship which gives preference to participate in online discussions through social media sites and focuses on justice.

One of government's important tasks is to prevent the spread of propaganda against the government, to detect the hidden social networks and improve the management on the e-government environment. As mentioned above, social networks play an important role in the management of e-government. There are various social networks on internet. Some of these social networks may be goodwill but others may be not. Goodwill social networks are the groups that detect the current problems and carry out a number of studies on it by following the e-government system. Non-goodwill social networks are the groups focused on the propaganda against the government, the weakening of it and the growth of discontent among the people and so on. In addition, social networks can operate obviously and non-obviously. In this regard, social network analysis is very important to provide the security of the government.

At present, the importance of social networks are growing, they are used in the detection of criminal groups and in the studying of hidden connections and the details between them, along with law enforcement and intelligence agencies. In e-government environment the social network analysis tools are useful to describe the view of the relations (Alguliev et al., 2012b; Alguliyev et al., 2016). Social network analysis is the set of powerful methods which can be used to identify groups and hidden structures in social network. At the same time, social network analysis suggests arrangements which can be used as the primary source for the investigation of the concepts of reputation, authority and confidence (Almarabeh and AbuAli, 2010). As a result we can say that, in all cases social networks play an important role in the management of e-government.

3 SONEX methods in different environments

Today web has become one of the popular tools in the dissemination of information. As the result of the rapid growth, diversity, dynamics of data on the web, the information users may encounter a number of problems. The problems include the finding of relevant information, the extraction of knowledge from the available information on the social networking sites, and learning about individuals users' behaviour (Ansari and Jalali, 2011).

The extraction of social networks on the web is one of the most important issues of the social network analysis. Social networks can be extracted from a variety of sources (Catanese et al., 2010). These sources may include the web pages, e-mail, blogs, and other online social networking sites. Depending on the source of information different SONEX methods can be grouped as follows:

- extraction of social networks from the web (Chen, 2003; Cotterill and King, 2006; Culotta et al., 2004)
- extraction of social networks from the e-mail environment (Dehkharghani et al., 2015; Hu et al., 2017)
- extraction of social networks from the blogs (Javanmardi et al., 2011; Kaushik et al., 2013)
- extraction of social networks from the online social networking sites (Kautz et al., 1997; Kazienko et al., 2011; Khasawneh and Abu-Shanab, 2013)

- extraction of social networks from the Wiki environment (Kosala and Blockeel, 2002; Luo and Huang, 2009; Matsuo et al., 2007)
- extraction of social networks from the multi-source data (Meishar-Tal and Tal-Elhasid, 2008; Merhav et al., 2012; Opsahl et al., 2010; Pang and Lee, 2008; Richards, 2010; Song et al., 2010; Stanforth, 2006).

3.1 Extraction of social networks from the web

In several studies, search engine has been used for SONEX on websites. In Chen (2003), the first automated, interactive tool has been developed for SONEX. This interactive tool is called Altavista. Here, Altavista is used for SONEX using the co-occurrence of names in any document (for example, personal homepages, co-authors lists in articles, citations, etc.). In Cotterill and King (2006) have been proposed a system for SONEX using the attributes of the conference participants, such as the names, e-mail, affiliation and etc. Here, the relationship between any two participants is determined based on web information gathered by the use of Jaccard measure, and sending of the query to the search mechanism as in Chen (2003). In Culotta et al. (2004) have been described network system considered for the academic community to facilitate communication and cooperation based on the extraction of social networks from the web. This system is called polyphonet, and was created from the combination of 'polyphony' and 'network' terms.

3.2 Extraction of social networks from the e-mail environment

E-mail is one of the essential ways that people use to communicate and obtain an access to social networking sites. Due to the specific characteristics of social networks, e-mail is considered a priority research area. According to several advantages of e-mail data, the social networks analysis have become a powerful information source. These advantages include the ability to use e-mail everywhere, longevity, the e-mail usage being mutual and others. Besides the advantages, there are some problems in the use of e-mail. Here, the social networks extraction is not an easy issue. So, the issues like the same person possessing several e-mail addresses; spams; classification of social relations according to e-mails contents and, etc., create a number of problems.

In Dehkharghani et al. (2015), EmailNet system has been proposed for the SONEX. In EmailNet system are extracted the letters both from individual e-mail users, as well as the corporate mail servers. Then, the filters spam are used to extract the spams. Here, text clustering technology has been used also to divide e-mails into several types of social connections. E-mail and the Internet, social networking and relation data extraction have been discussed on Hu et al. (2017). In this article, the system has been proposed for the extraction of user's social network, and contact information of their members based on the user's e-mail. This system is implemented recursively and is built based on 'friend of a friend' principle of larger network.

3.3 Extraction of social networks from the blogs

The blogosphere is the network of the social media sites where people express their opinions, discuss the public events and facts. This environment is one of the richest

sources of social information. In recent years, the dramatic increase of blogs' number, diversity, and popularity, make it mature field to extract the networks automatically.

In Javanmardi et al. (2011), a system, called SONEX, has been proposed to extract the social networks by determining the names of the entities and relationship between them in the blogosphere. The system works by clustering and extracting a pair of entities in the blogosphere. In clustering step the entities with similar context are grouped together, each cluster is analysed and after verifying the contents of pairs, they are included to the same class.

Blogs have become an important tool in the diffusion of information, and hidden social structures have a significant influence on the level and scope of information flow. In Kaushik et al. (2013), a new approach has been proposed to measure the influence of the social structure on the information flow. Experiments in this article show that the social structures have influence on the scope of the diffusion networks of 'interest' topics and the influence of social networks on information diffusion is related to the characteristic of the information itself.

3.4 Extraction of social networks from the online social networking sites

Online social networks have gained popularity in recent years, and the majority of personal data are available on these sites. Online social networking sites contain huge number of user profiles with semi-structured personal data. Several studies have been carried out to extract the social networks out of the dynamic data (information) available on websites. In Kautz et al. (1997), the availability of relevant information extraction about the relationship between Facebook users was discussed. There are some researches that analysed relationships within the message topics in online social networking services. In Kazienko et al. (2011), social networks are detected by the extraction of hidden social relationships within the message topics. Here, they try to detect hidden relationships within the message topics by determining the most co-occurred set of users who write a message on the same subject. In Khasawneh and Abu-Shanab (2013) have been proposed an approach to construct the social networks from Facebook photo albums with the help of unsupervised face recognition methods.

3.5 Extraction of social networks from the Wiki environment

Relations between the users and social network analysis in Wiki environment have been studied by various researchers. In Kosala and Blockeel (2002) has been proposed the model to detect social relationships between Wiki users and to determine vandals, inexperienced users in Wiki environment. In Luo and Huang (2009), another approached has been proposed to measure social relations and impacts in Wiki environment. In the proposed method are used three criteria:

1 number of Wiki users

2 interactivity, it means, the number of revisions in particular time interval by the Wiki users

3 intensity, the number of changes in Wiki pages as a result of new revisions.

In Matsuo et al. (2007), a method has been proposed to extract the social networks which caused to the information conflict. In this article social networks of Wiki users are

extracted by clustering unstructured articles that leads to the information conflict. Here, weighted hybrid fuzzy c-means method is used for the clustering of conflicted articles.

3.6 Extraction of social networks from the multi-source data

The web which is considered as the world's biggest database has been used in various studies to extract social networks (Meishar-Tal and Tal-Elhasid, 2008). In many studies, single source has been used to extract social networks. Very few studies have been used several sources for extraction of social networks. In Merhav et al. (2012) and Opsahl et al. (2010) has been proposed a method for determination of researchers' network based on information obtained from various sources on the web. In Pang and Lee (2008), an approach has been proposed for the synthesis of scientific organisations' social networks based on information sources available on the web. In Richards (2010), a system has been proposed to extract social networks from instant messenger and e-mail. Here, two major components are used: offline data collection and online data processing. Firstly, related communication data is collected from e-mail and instant messenger/chat. The data extracted in such way is filtered and stored in the database. These processes is carried out by offline collection module. Then this data is used by online processing module for the extraction and visualisation of social network. Song et al. (2010) proposes a system to extract social networks using an average time spent on each page by every user on the web server daily log files. Stanforth (2006) proposes general model for the extraction of multi-dimensional social networks. In this model, information is gathered about dynamic behaviour of users. The proposed model uses three dimensions, i.e., relations, groups, and time for social networks analysis. Under the relationship is intended, the whole relations between the system users (e-mail and phone, or indirect). The group dimension covers all social groups, which can be accessed in clustering process. Time dimension is based on relationship existing at a certain time or extracted at a certain period of time, i.e., a human activity within a certain time-window.

4 An approach to extract of the hidden social networks from e-government environment

Extracting of the hidden social networks from e-government is one of the main factors to provide the government security and improvement of management. Today, various criminal organisations use some social networking sites in order to connect and enlist, and such organisations keep their relationship a secret. It is possible to trace the hidden relationships by analysis of social network data in e-government environment. The social network analysis can be used to detect criminal networks in this environment.

Note that many studies have been carried out on the social network analysis. It was determined that the majority of users communicate in the virtual world by use of e-mail, blogs, online chat and etc. (Kazienko et al., 2011). The user' comments written on any government agency, institution, banks and others in e-government environment can play a key role in determination of social relationships. In this article, an approach for the extraction of hidden social networks from the e-government environment has been proposed.

It is known that users are able to give their opinion on any information by writing comments in e-government environment. So, before adoption of any bill it is possible to learn e-citizens' opinions and to determine the positive and negative aspects of law by placing it on the e-government system. Ineffectiveness of the adopted bill can be identified constructing feedback mechanism through the analysis of citizens' opinions. The relationships between users can be modelled through the comments of users written to each other. Here, can be seen different approaches, for example, the response of users to the comment written by any other user, the number of users commented to this statement, and so on.

The proposed approach in this article consists of the following stages:

1 data collection and preprocessing

2 classification

3 SONEX

4 social network analysis.

Each stage of the proposed approach is explained in detail below.

4.1 Data collection and preprocessing

At this stage, initially the comments written to any information (web page, information) are collected. Suppose, n number information T_i, $i = 1, 2, \ldots, n$ are placed in e-government environment. The comments written to the i^{th} information are denoted as follows:

$$C_i = \left\{c_i^j\right\} i = 1, 2, ..., n,\ j = 1, 2, ..., m \tag{1}$$

where c_i^j is the set of comments written by j^{th} user to the i^{th} information, m is the number of users.

After collecting comments, the preprocessing process is carried out on them. Punctuation marks and spaces are eliminated during the preprocessing process.

4.2 Classification

At the next stage, the set of comments written to each information are grouped into three classes: positive, negative and neutral. Positive class is denoted as C_i^+, negative C_i^-, and neutral C_i^0 :

$$C_i = C_i^+ \cup C_i^- \cup C_i^0, \quad i = 1, ..., n \tag{2}$$

In this article, sentiment analysis is proposed to group the written comments into three classes.

Sentiment analysis is one of the most advanced technologies for the analysis of texts. It is the process of automatic detection of the emotion in textual contents (document, comment, e-mails, etc.) by utilisation of computer. This work is also named as opinion mining in literature. The emotion which is being detected can contain author's mood and his/her ideas about subject, the effect wanted to create and etc. The sentiment analysis

studies can be classified into three main groups: 'dictionary-based approaches', 'machine learning algorithms' and 'rules-based approaches' (Tang et al., 2009; Tasleem et al., 2014).

Sentiment analysis can be carried out at different levels:

- *Word level:* in this level sentiment polarity of words and multi-words (the positive or negative expression of these words) is defined.
- *Phrase level:* sentence may consist of various phrases and each phrase can express different meanings (for example, I liked it, but not everyone did). Therefore, each phrase in this level is considered separately.
- *Sentence level:* the sentiment analysis of sentences is based on word and phrase level. If the sentence has a mixed polarity (both positive and negative due to multiple aspects or phrases), one can assign an overall polarity based on relative sentiment strengths of the components/phrases inside the sentence.
- *Aspect level:* each sentence in the same domain may consist of several aspects and each aspect may have different polarity. For example, the lighting is good, but I did not like the sound effects.
- *Document level:* In this level the overall polarity of document is defined. Often document polarity is aggregated from the estimated polarity of the constituent words or sentences. Researches show that sentence-level analysis is effective and initial and last sentences may have higher influence on document polarity, compared to sentences in the middle.

Comments are usually consists of several sentences, but the thought is expressed in the sentence level. Considering this, firstly, the polarity of each sentence and then the overall polarity of comment is determined (Ting et al., 2009; Tomobe et al., 2003).

Suppose, V polarity table of words has been built through the sentiment analysis approaches. In this approach to determine which groups (positive, negative and neutral) the comments are belonging, firstly comments divided into sentences. After the split of comments, each sentence is being cleaned from the stop words. After cleaning, the polarity degree of each sentence is determined. The polarity is calculated through the polarity degree of words in the same sentence. To calculate the polarity degree of each comment the following formula is offered:

$$score\left(c_i^j\right) = sign\left(\sum_{s \in c_i^j} score(s)\right) \tag{3}$$

where $score(c_i^j)$ – is the polarity degree of comment written by the j^{th} user to the i^{th} information, $score(s)$ – is the polarity degree of sentence in this comment.

$sign(x)$ is the mark function and is defined as follows:

$$sign(x) = \begin{cases} 1 & if \quad x > 0 \\ 0 & if \quad x = 0 \\ -1 & if \quad x < 0 \end{cases} \tag{4}$$

If the overall polarity of sentences in the comment greater than zero, then the comment can be included to positive class, if equals to zero, then included to neutral class else negative class.

The following formula is used to calculate the polarity degree of sentence s:

$$score(s) = \sum_{t \in s} score(t) \tag{5}$$

To calculate the polarity degree of each word in the sentence, the pointwise mutual information (PMI) is used (Van Alstyne and Zhang, 2003). The polarity degree of each word is calculated as follows:

$$score(t) = \sum_{x \in V} PMI(t, x)\left(score^{+}(x) - score^{-}(x)\right) \tag{6}$$

$$PMI(t, x) = \log_2\left(\frac{n_{tx}}{n_t . n_x}\right) \tag{7}$$

where $score(t)$ is the polarity degree of term t in the sentence s, $score^{+}(x)$ is the positive and $score^{-}(x)$ is the negative score of term x in the table V, and n_t, n_x, n_{tx} are the number of observations of t, x, t and x occurring in a corpus, respectively.

So, comments written to the i^{th} information are grouped into three classes positive, neutral and negative.

$$\begin{aligned} C_i^{-} &= \left\{c_i^j \middle| score\left(c_i^j\right) = -1\right\} \\ C_i^{+} &= \left\{c_i^j \middle| score\left(c_i^j\right) = 1\right\} \\ C_i^{0} &= \left\{c_i^j \middle| score\left(c_i^j\right) = 0\right\} \end{aligned} \tag{8}$$

4.3 Social network extraction

At this stage, the social network actors and the relationships between them are determined. Firstly, users gathering around the negative class are defined. We consider that each user can be identified (either by registering, or by IP address). Users writing negative comments at least to one information are defined as follows:

$$U_{\Sigma}^{-} = \bigcup_{i=1}^{n} U_i^{-} \tag{9}$$

where U_i^{-} are the users writing negative comments to the i^{th} information.

Users writing negative comments to all information are defined as follows:

$$U_{\Pi}^{-} = \bigcap_{i=1}^{n} U_i^{-} \tag{10}$$

U_{Π}^{-} is the core (key actors) of social network.

Two types of approaches are used to determine the relationship between the social network actors.

In the first approach, the relationships between social network actors are determined through the number of information written negative comments by the users. Therefore, the following formula is proposed:

$$w_1^{j_1 j_2} = \frac{n^{j_1 j_2}}{n^{j_1} + n^{j_2}} \tag{11}$$

where $n^{j_1 j_2}$ is the number of information written negative comments by the j_1 and j_2th users, n^{j_1} is the number of information written negative comments by j_1th user, n^{j_2} is the number of information written negative comments by j_2th user.

$n^{j_1 j_2}$, n^{j_1} and n^{j_2} are determined by the following formulas:

$$n^{j_1 j_2} = \sum_{i=1}^{n} I\left(c_i^{j_1}\right) \cdot I\left(c_i^{j_2}\right) \tag{12}$$

$$n^{j_1} = \sum_{i=1}^{n} I\left(c_i^{j_1}\right) \tag{13}$$

$$n^{j_2} = \sum_{i=1}^{n} I\left(c_i^{j_2}\right) \tag{14}$$

where $I(c_i^j)$ is defined as follows:

$$I\left(c_i^j\right) = \begin{cases} 1, & if\ c_i^j \neq \emptyset \\ 0, & otherwise \end{cases} \tag{15}$$

If the jth user comments at least one time to the ith information then $I(c_i^j)$ function is defined as 1, else 0.

Here the number of negative comments written to the same information by users can be considered too. In this case, the weight of relationship between the users can be defined by the following formula:

$$\tilde{w}_1^{j_1 j_2} = \frac{\sum_{i=1}^{n} \left(m_i^{j_1} + m_i^{j_2}\right) * I\left(c_i^{j_1}\right) \cdot I\left(c_i^{j_2}\right)}{M^{j_1} + M^{j_2}} \tag{16}$$

$$M^j = \sum_{i=1}^{n} m_i^j \tag{17}$$

where $m_i^{j_1}$ is the number of comments written by the j_1th user to the ith information, $m_i^{j_2}$ is the number of comments written by the j_2th user to the ith information. $\sum_{i=1}^{n} (m_i^{j_1} + m_i^{j_2}) * I\left(c_i^{j_1}\right) \cdot I\left(c_i^{j_2}\right)$ is the overall number of comments written by the j_1 and j_2th users to the same information. M^j is the overall number of comments written by the jth user.

In the second approach, the relationships between the social network actors are determined through the semantic similarity between comments written by users in the negative class. Obviously, similarity measures have an important role in the analysis of social networks. It is possible to judge about the strength of the relationship between the users through the similarity measures. In this article, Jaccard measure is used to calculate the similarity between comments:

$$w_2^{j_1 j_2} = sim\left(c^{j_1}, c^{j_2}\right) = \frac{\left|c^{j_1} \cap c^{j_2}\right|}{\left|c^{j_1} \cup c^{j_2}\right|} \tag{18}$$

where $sim(c^{j_1}, c^{j_2})$ is the semantic similarity between comments written by the j_1 and j_2th users.

So, the relationships between the actors of hidden social network are determined through the linear combination of above-proposed weights (16) and (18):

$$w^{j_1,j_2} = \alpha \cdot \tilde{w}_1^{j_1,j_2} + (1-\alpha) \cdot w_2^{j_1,j_2} \tag{19}$$

where $\alpha (0 \le \alpha \le 1)$ denotes weight coefficients.

To determine relationships between hidden social network's users precisely, addition information sources can be used. It includes e-mail, tweets, instant messaging and etc. It can be increased the adequacy of the social network and the strength of the relationships between users through determining which relations they are connected.

So, hidden social networks between users in e-government were defined through the users' comments.

4.4 Social network analysis

In the next stage, social network analysis is carried out. Analysis indicates determining the key actors in the network, the degree of their importance and so on. To determine the key actors in the social network, it is necessary to show compactness of the core. Therefore, using the number of users and relations between them in the social network is proposed. Firstly, the number of users in the whole social network is defined through the following formula:

$$N_{\bar{\Sigma}} = \left| U_{\bar{\Sigma}} \right| \tag{20}$$

Then the number of users in the core of the social network is defined as follows:

$$N_{\bar{\Pi}} = \left| U_{\bar{\Pi}} \right| \tag{21}$$

After determining the number of users in the social network, the number of relations between them is defined as follows:

$$M_{\bar{\Sigma}} = \sum\nolimits_{j_1, j_2 \in U_{\bar{\Sigma}}} I_1 \left(w^{j_1,j_2} \right) \tag{22}$$

$$M_{\bar{\Pi}} = \sum\nolimits_{j_1, j_2 \in U_{\bar{\Pi}}} I_1 \left(w^{j_1,j_2} \right) \tag{23}$$

where $M_{\bar{\Sigma}}$ is the number of relations between users in the whole social network, $M_{\bar{\Pi}}$ is the number of relations between users in the core of the social network.

The function $I_1(w^{j_1,j_2})$ is defined as follows:

$$I_1(x) = \begin{cases} 1, & if \quad x > 0 \\ 0, & if \quad x = 0 \end{cases} \tag{24}$$

Then the density coefficient of the whole network is determined using the following formula:

$$\sigma_{\bar{\Sigma}} = \frac{M_{\bar{\Sigma}}}{\frac{N_{\bar{\Sigma}} \cdot \left(N_{\bar{\Sigma}} - 1\right)}{2}} \tag{25}$$

where $\frac{N_{\bar{\Sigma}} \cdot \left(N_{\bar{\Sigma}} - 1\right)}{2}$ is the number of possible relations between the social network actors.

Similarly, the density coefficient of the core is determined as follows:

$$\sigma_{\bar{\Pi}} = \frac{M_{\bar{\Pi}}}{\frac{N_{\bar{\Pi}} \cdot \left(N_{\bar{\Pi}} - 1\right)}{2}} \tag{26}$$

where $\frac{N_{\bar{\Pi}} \cdot \left(N_{\bar{\Pi}} - 1\right)}{2}$ is the number of all possible relations between the core's actors.

The weight of the core in the whole social network is defined using the (24) and (25) formulas as follows:

$$\sigma^{-} = \frac{\sigma_{\bar{\Pi}}}{\sigma_{\bar{\Sigma}}} \tag{27}$$

where σ^{-} is the weight of the core in the whole social network. Based on this, compactness of the core is defined.

After identifying the compactness of the core, the importance score of core actors is calculated by using the number and weight of relations between users. For this purpose, the following formula is proposed (Wang et al., 2010):

$$c_j^{w\beta} = k_j^{(1-\beta)} \cdot s_j^{\beta}, \quad 0 \le \beta \le 1 \tag{28}$$

$$k_j = \sum\nolimits_{l \in U_{\bar{\Sigma}}} I_1\left(w^{jl}\right) \tag{29}$$

$$s_j = \sum\nolimits_{l \in U_{\bar{\Sigma}}} w^{jl} \tag{30}$$

where $c_j^{w\beta}$ is the centrality degree, k_j is the total number and s_j is the total weight of relations between the j^{th} actor of the core and other actors in the network, respectively.

So, hidden social networks suspected of anti-government propaganda, the key actors of this network and their importance degree have been defined through the analysing comments of citizens in e-government environment.

5 Conclusions and future work

In this article, the method has been proposed to prevent anti-government propaganda. In this approach social networks were extracted through the analysis of comments of the citizens written to the documents in e-government environment. For this purpose, firstly the comments were analysed; grouped and hidden social networks were detected among the users giving at least once a negative comment. Then the key actors were defined from

this network. This is the one of the approaches to ensure the government security. The proposed approach can also be used for different purposes:

- determination of the interest of the country's regions
- determination of the citizens' attitude to the bills, judgements, and other documents
- determination of the citizens' satisfaction.

The proposed approach is conceptual and the complex issues for its implementation should be solved. First of all, there is a possibility of comments being in different languages. In this case, there is a need to develop the technologies of texts analysis for multi-lingual environment. On the other hand, in the proposed approach the social network is synthesised by the use of only one source. But studies show that such synthesis of social networks does not reflect the real situation. Therefore, there is a need for additional information sources. In this case, appears a new challenge – the determination of reliability degree of information received from various sources. Considering all these limitations, there are a few problems waiting to be solved:

- information being in different languages
- insufficiency of only one information source
- the reliability of this information sources.

At the end we come to the conclusion that the proposed approach is a framework approach, and the specification of its details is a main issue of future studies.

Acknowledgements

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – Grant No. EIF-KETPL-2-2015-1(25)-56/05/1.

References

Akbaş, E. (2012) *Aspect Based Opinion Mining on Turkish Tweets*, Unpublished Master's thesis, Bilkent University, Turkey.

Alguliev, R.M., Aliguliyev, R.M. and Ganjaliyev, F.S. (2011) 'Extracting a heterogeneous social network of academic researchers on the web based on information retrieved from multiple sources', *American Journal of Operations Research*, Vol. 1, No. 2, pp.33–38.

Alguliev, R.M., Aliguliyev, R.M. and Ganjaliyev, F.S. (2012a) 'Aggregating edge weights in social networks on the web extracted from multiple sources with different importance degrees', *Journal of Intelligent Learning Systems and Applications*, Vol. 4, No. 2, pp.154–158.

Alguliev, R.M., Aliguliyev, R.M. and Ganjaliyev, F.S. (2012b) 'Building a social network of research institutes from information available on the web', *International Journal of Networking and Virtual Organizations*, Vol. 11, No. 1, pp.62–76.

Alguliyev, R., Aliguliyev, R. and Alakbarova, I. (2016) 'Extraction of hidden social networks from wiki-environment involved in information conflict', *International Journal of Intelligent Systems and Applications*, Vol. 8, No. 2, pp.20–27.

Almarabeh, T. and AbuAli, A. (2010) 'A general framework for e-government: definition maturity challenges, opportunities, and success', *European Journal of Scientific Research*, Vol. 39, No. 1, pp.29–42.

Ansari, A. and Jalali, M. (2011) 'A system for social network extraction of web complex structures', *International Journal of Computer Science and Information Security*, Vol. 9, No. 8, pp.67–75.

Catanese, S.S., Pasquale De Meo, P., Ferrara, E. and Fiumara, G. (2010) 'Analyzing the facebook friendship graph', in *Proceedings of the 1st International Workshop on Mining the Future Internet-MIFI'10*, Berlin, pp.14–19.

Chen, Y-C. (2003) 'E-government network: the role of information technology in managing networks', in *National Public Management Research Conference*, pp.1–36.

Cotterill, S. and King, S. (2006) 'The role of social networks in the development of English local e-government', in *Political Studies Association Conference*, Reading, pp.1–13.

Culotta, A., Bekkerman, R. and McCallum, A. (2004) 'Extracting social networks and contact information from e-mail and the web', in *Proceedings of Conference on Email and Anti-Spam*, pp.1–9.

Dehkharghani, R., Yanikoglu, B. and Saygin, Y. (2015) 'Sentiment analysis in Turkish: towards a complete framework', *Journal of Natural Language Engineering*, Vol. 1, No. 1, pp.1–25.

Hu, Y-H., Chen, Y-L. and Chou, H-L. (2017) 'Opinion mining from online hotel reviews – a text summarization approach', *Information Processing & Management*, Vol. 53, No. 2, pp.436–449.

Javanmardi, S., McDonald, D.W. and Lopes, C.V. (2011) '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*, ACM, New York**,** pp.82–90.

Kaushik, A., Kaushik, A. and Naithani, S. (2013) 'A study on sentiment analysis: methods and tools', *International Journal of Science and Research*, Vol. 4, No. 12, pp.287–292.

Kautz, H., Selman, B. and Shah, M. (1997) 'The hidden web', *American Association for Artificial Intelligence Magazine*, Vol. 18, No. 2, pp.27–35.

Kazienko, P., Musial, K., Kukla, E., Kajdanowicz, T. and Brodka, P. (2011) 'Multidimensional social network: model and analysis', in *Proceedings of International Conference on Computer and Computational Intelligence*, Bangkok, pp.378–387.

Khasawneh, R.T. and Abu-Shanab, E.A. (2013) 'E-government and social media sites: the role and impact', *World Journal of Computer Application and Technology*, Vol. 1, No. 1, pp.10–17.

Kosala, R. and Blockeel, H. (2002) 'Web mining research: a survey', *SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery and Data Mining*, Vol. 2, No. 1, pp.1–15.

Luo, D. and Huang, H. (2009) 'Link prediction of multimedia social network via unsupervised face recognition', in *Proceedings of MM'09*, Beijing, pp.805–808.

Matsuo, Y., Mori, J. and Hamasaki, M. (2007) 'POLYPHONET: an advanced social network extraction system from the web', in *Proceedings of World Wide Web Conference*, pp.262–278.

Meishar-Tal, H. and Tal-Elhasid, E. (2008) 'Measuring collaboration in educational Wikis – a methodological', *Emerging Technologies in Learning*, Vol. 3, No. 3, pp.46–49.

Merhav, Y., Mesquita, F., Barbosa, D., Yee, W.G. and Frieder, O. (2012) 'Extracting information networks from the blogosphere', *ACM Transaction on the Web*, Article No. 11, Vol. 6, No. 3, pp.1–33.

Opsahl, T., Agneessens, F. and Skvoretz, J. (2010) 'Node centrality in weighted networks: generalizing degree and shortest paths', *Social Networks*, Vol. 32, No. 3, pp.245–251.

Pang, B. and Lee, L. (2008) 'Opinion mining and sentiment analysis', *Foundations and Trends in Information Retrieval*, Vol. 2, Nos. 1–2, pp.1–135.

Richards, R. (2010) 'Digital citizenship and Web 2.0 tools', *MERLOT Journal of Online Learning and Teaching*, Vol. 6, No. 2, pp.516–522.

Song, M., Lee, T. and Kim, J. (2010) 'Extraction and visualization of implicit social relations on social networking services', in *Proceedings of the 24th AAAI Conference on Artificial Intelligence-AAAI'10*, Atlanta, pp.1425–1430.

Stanforth, C. (2006) 'Using actor-network theory to analyze e-government implementation in developing countries', *Information Technologies and International Development*, Vol. 3, No. 3, pp.35–60.

Tang, J., Wang, T. and Wang, J. (2009) 'Measuring the influence of social networks on information diffusion on blogosphere', in *Proceedings of the 8th International Conference on Machine Learning and Cybernetics*, Baoding, pp.3492–3498.

Tasleem, A., Rashid, A. and Asger, M. (2014) 'Social network extraction: a review of automatic techniques', *International Journal of Computer Applications*, Vol. 95, No. 1, pp.16–23.

Ting, I., Wu, H. and Chang, P. (2009) 'Analyzing multi-source social data for extracting and mining social networks', in *Proceedings of 12th International Conference on Computational Science and Engineering*, Vancouver, Canada, pp.815–820.

Tomobe, H., Matsuo, Y. and Hasida, K. (2003) 'Social network extraction of conference participants', *in Proceedings of 12th International Conference on World Wide Web*, Budapest, pp.1–2.

Van Alstyne, M. and Zhang, J. (2003) 'EmailNet: a system for automatically mining social networks from organizational email communication', in *Proceedings of 2003 North American Association for Computational Social and Organizational Science*, pp.1–4.

Wang, K., Ting, I., Wu, H. and Chang, P. (2010) 'A dynamic and task-oriented social network extraction system based on analyzing personal social data', in *Proceedings of 2010 International Conference on Advances in Social Networks Analysis and Mining*, Denmark, pp.464–469.