Overlapping community detection in rating-based social networks through analyzing topics, ratings and links

Ali Reihanian , Mohammad-Reza Feizi-Derakhshi ,
Hadi S. Aghdasi

PII: S0031-3203(18)30148-1
DOI: 10.1016/j.patcog.2018.04.013
Reference: PR 6527

To appear in: *Pattern Recognition*

Received date: 24 August 2017
Revised date: 14 November 2017
Accepted date: 11 April 2018

Please cite this article as: Ali Reihanian , Mohammad-Reza Feizi-Derakhshi , Hadi S. Aghdasi , Overlapping community detection in rating-based social networks through analyzing topics, ratings and links, *Pattern Recognition* (2018), doi: 10.1016/j.patcog.2018.04.013

Highlights

- A generic framework is proposed for community detection in social networks with special focus on rating-based social networks.
- The framework finds the overlapping communities in which the members are interested in the same topic, and the strengths of their relationships are based on the rate of their viewpoints' unity.
- A novel weighting strategy for rating-based social networks is proposed which performs based on value of ratings.
- Quantitative evaluations show that the proposed framework has better performance than 3 other relevant frameworks.

Overlapping community detection in rating-based social networks through analyzing topics, ratings and links

Ali Reihanian[a], Mohammad-Reza Feizi-Derakhshi[a], and Hadi S. Aghdasi[a]

[a] *Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran*

Corresponding Author: Ali Reihanian

- E-mail address: ali.reihanian@gmail.com
- Tel: +989111116119
- Address: Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran.

Abstract

Owing to advances in information technology, online communications between people living in different parts of the world have considerably increased. The subsequent emergence of social networks helped this kind of communications to be further organized. One of the most important issues considered when analyzing these kinds of networks is community detection, in which a majority of studies tend to detect disjoint communities through analyzing linkages of networks. What this paper aims to achieve is to obtain overlapping communities in which the members have the same topics of interest, and where the strengths of connections between them are the consequence of their communications' content analysis. Consequently, we have hereby proposed a generic framework for overlapping community detection in social networks with special focus on rating-based social networks. This framework considers the information shared by the users (ratings), as well as their topics of interest, for the sake of finding meaningful communities. This will lead us to topical communities in which members are interested in the same topics, and the strengths of their relationships are directly based on the rate of their viewpoints' unity. Quantitative evaluations also reveal that the framework presented in this study achieves favorable results which are quite superior to the results of 3 other relevant frameworks in the literature.

Key Words - Overlapping community detection, Content analysis, Topical community, Semantic network, Rating-based social networks

1. Introduction

With the advance of information technology, online communications between people have increased significantly. This kind of communications has become more organized subsequent to the emergence of social networks. In recent days, social network services strongly affect our daily life, the industry and the academy [1]. Finding meaningful communities in social networks is an interesting research area, and has attracted the attentions of many researchers [2]. Social networks can be considered as a kind of complex networks. The community structures of complex networks reveal both their organization and hidden relations among their constituents [3]. Thus, community detection is truly considered as an important issue in analyzing complex networks [4]. A Community (also sometimes referred to as a module or cluster [5]) is a dense subnetwork within a larger network, such as a close-knit group of friends in a social network or a group of interlinked web pages on the World Wide Web

[6]. The role of community detection is searching a network for groups of nodes with more interactions amongst their members than those between their members and the remainder of the network [5]. In other words, community detection involves grouping the nodes of a network into communities of densely connected nodes [7]. As the members of a community are more likely to have common hobbies, social functions, etc., the identified communities can be used in collaborative recommendation, information spreading, knowledge sharing, and other applications that are beneficent for us [8].

Most of the studies in the field of community detection mainly focus on the topological structure of networks. The results of these studies is the communities which may incorporate different topics since they only consider the strengths of connections between individuals, and no content analysis is done in their process for finding communities. Nowadays, real-world networks, like Facebook and Twitter, are containing a vast range of information including shared objects, comments, etc. Thus, it is unreasonable for a community to be explained by a single entity because the community members are generally interacting with each other via a large number of distinguishable ways in various domains [2, 9]. In very recent studies, extracting semantic information from the networks is started to enhance the performance of community detection algorithms. However, the amount of covert information* which is extracted from a network is very limited, and it yet remained a school of thought. On the other hand, most of these studies are based on an assumption that the communities of complex networks are disjoint or separated, namely, every node just belongs to exactly one non-overlapping community. However, the communities in many real-world networks often overlap to a certain extent. In another word, some nodes in the networks may belong to multiple communities because of their diverse role in the network systems [10]. For example, in a social network, a person may have different roles, and usually be connected to several social groups like family, friends, colleagues, etc [11].

There is plenty of room to study on community detection problem in real complex networks which contain covert information with different natures. One possible solution, which is addressed in this paper, is using a generic framework that considers both the topics of interest and communications' content analysis, simultaneously, in order to find overlapping communities.

Rating-based social networks are social networking sites, in which the users express their feelings toward different objects (like movies) by the means of rating. In spite of their interesting organization, to our best knowledge, the overlapping community structures of rating-based social networks have not been well studied in the literature of community detection. Thus, the proposed generic framework of this paper has a special focus on the detection of overlapping communities in rating-based social networks. Uncovering the community structures of rating-based social networks with the proposed framework of this paper can have many potential applications, such as improving the efficiency of collaborative recommendations, in this kind of social networks. Compared to the existing studies, the proposed framework detects overlapping communities with considering the information of topics, communications' contents (ratings), and topological structure of a rating-based social network, altogether.

As a brief, the main goal of the present paper is to obtain communities which have three main characteristics: first, their members have the same topics of interest. Second, the strengths of relationships between their members are directly based on the rate of their viewpoints' unity (they are closely related to each other based on their viewpoints), which is concluded according to their communications' contents (ratings). Third, the strengths of connections of intra communities are

* Covert information is the latent semantic information which can be extracted by analyzing the contents of a network in order to enhance the performance of community detection. The information which can be inferred by analyzing the contents of communications between individuals in social networks is an example of the covert information.

much more than those of inter communities. The reason of this paper for finding the like-minded and strongly connected groups of individuals in a rating-based social network is that it can have many potential applications, such as improving the efficiency of collaborative recommendations, in this kind of social networks.

Moreover, the main contributions of this paper are:

1- In this paper, the related works are studied and divided into three different groups according to their contents. The first group considers the graph structure of a network for finding communities, while no content analysis is used in the process of their proposed approaches. The second group tends to partition networks into different groups of nodes, in which every node has the same topic of interest. And the third group considers both the contents that are interchanged in networks and the topological structures of the networks, in order to find meaningful communities. Also, after introduction, the main strengths and weaknesses of these groups are described.

2- This paper proposes a generic framework for community detection in social networks with special focus on rating-based social networks. This framework is able to find the overlapping communities in which the members are interested in the same topic, and the strengths of their relationships are directly based on the rate of their viewpoints' unity.

3- The proposed framework uses a novel weighting strategy for rating-based social networks which performs based on the value of ratings. The weighting strategy calculates the weight of relations in a rating-based social network. This weight quantifies the degree of alignment between members on their sentiments toward social objects that represent the topic in question.

4- In order to evaluate the performance of the proposed framework of this paper, experiments were conducted on 3 real-life data sets. Also, the approaches of 3 relevant frameworks were implemented in order to compare them with the proposed one.

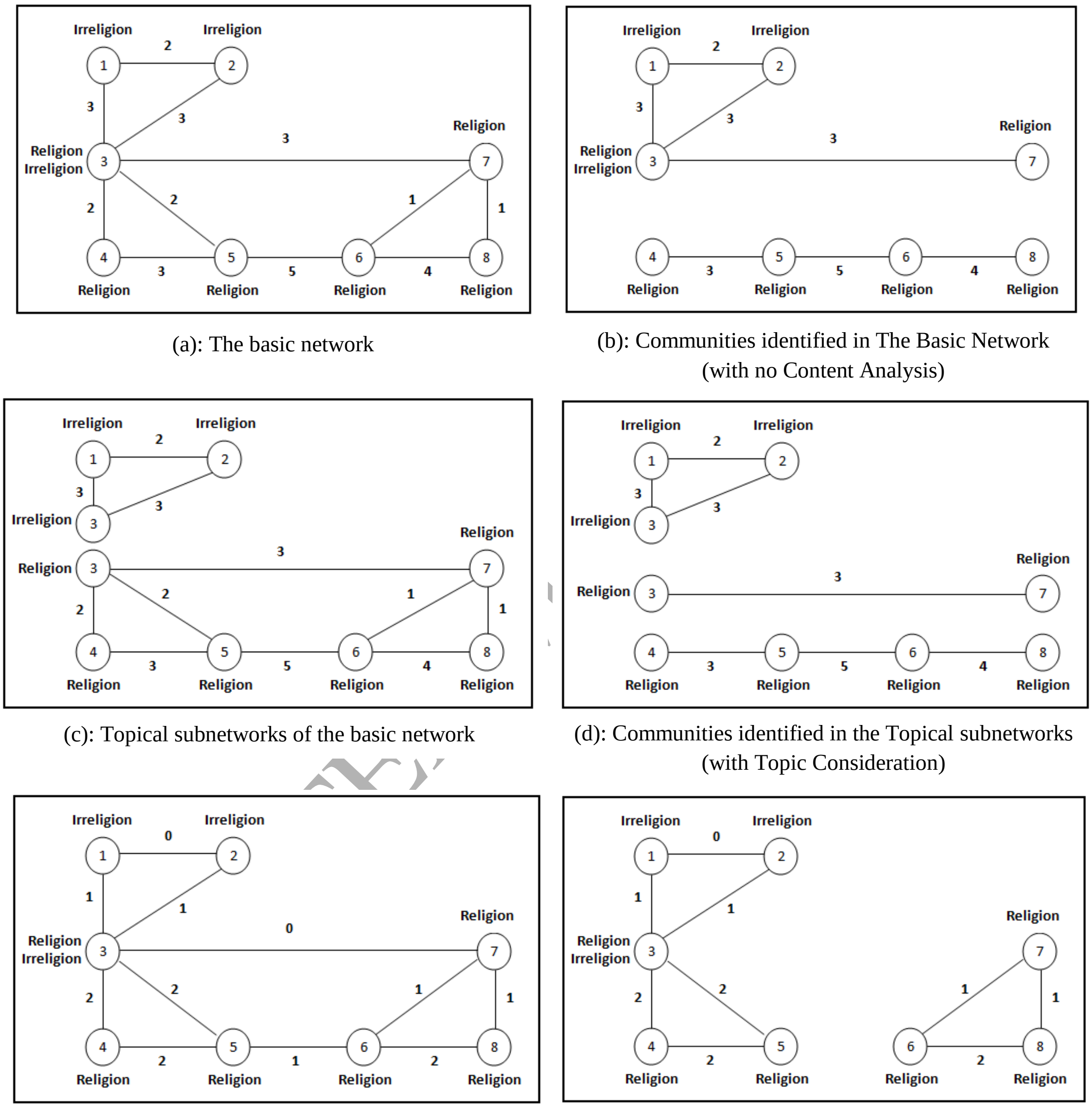
The remainder of this paper is outlined as follows. In Section 2, the motivation of our research is explained. Section 3 reviews the related works. In Section 4, our proposed framework is presented. In order to verify the proposed framework, extensive experiments are conducted on real-life data sets. The descriptions of these data sets, the experimental results and their analysis are given in Section 5. Finally, Section 6 concludes the paper.

2. Motivation

In this section, the motivation of our research is explained with the example illustrated in Figure 1. Figure 1(a) shows a network of 8 nodes and 11 edges. We refer to this network as the basic network. Each node represents an individual, and each edge shows the social relation of interactions or communications between two individuals in the basic network. The weight of each edge indicates the number of communications between the related nodes. For example, if node i finishes five communications with node j, the assigned weight of their related edge will be 5. Consider that the topics of interest for each node are assigned to them manually. These topics represent the domain of interest for each individual in the basic network. According to Figure 1(a), each node in the basic network can be interested in discussions related to religion, irreligion or both of them.

Figure 1(b) shows the identified communities after applying a topological community detection algorithm to the basic network without performing any content analysis. The members of each identified communities are connected, but the community that is located at the top of Figure 1(b) incorporates different topics. 2 members in this community are interested in the discussions which are related to religion, while 3 members are interested in the discussions which are related to irreligion.

Figure 1(c) shows the partition of the basic network that has two topical subnetworks. Each topical subnetwork contains the nodes of the basic network which have the same topic of interest. For example, in the topical subnetwork that is located at the bottom of the figure, all of the members are interested in the discussions which are related to religion. We applied a topological community detection algorithm to each topical subnetwork of Figure 1(c) for finding communities in which members have the same topics of interest and strong connections. Figure 1(d) shows the identified communities.



(a): The basic network

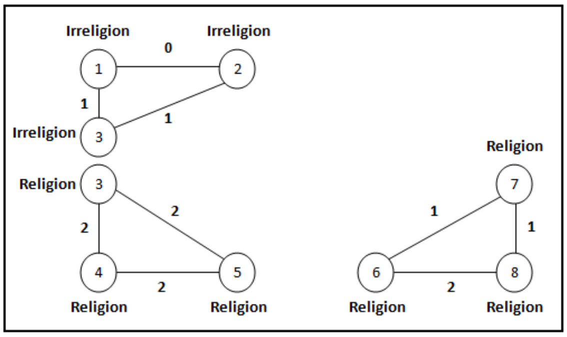
(b): Communities identified in The Basic Network (with no Content Analysis)

(c): Topical subnetworks of the basic network

(d): Communities identified in the Topical subnetworks (with Topic Consideration)

(e): The Semantic Network generated from the Results of Communications' Content Analysis

(f): Communities identified in The Semantic Network (with Communications' Content Analysis but no Topic Consideration)



(g): The expected communities which can be identified by combining both Communications' Content Analysis and Topic Consideration

Fig. 1. An example illustrating the motivation of the research

Now, if we analyze the contents of communications between individuals in the basic network in order to discover the rate of their viewpoints' unity concerning the subject of their communications, we may be encountered with a new situation. Assume that, according to this communications' content analysis, the strengths of connections between individuals are changed, because the weights of these connections are directly calculated based on the results of their corresponding communications' content analysis. Consider the situation, in which node i has 5 communications with node j. In the basic network, the strength of connection between these two nodes is 5, and is attached to them by the weight of their related edge. However, after analyzing the contents of these 5 communications, and extracting the rate of these nodes' viewpoints' unity, we may encounter with this fact that: actually in none of their communications did they have the same opinion about certain subjects. Thus, if we consider this fact, we can't place a strong connection between them, because they are not strongly connected. They have different views about the subjects of their communications.

After performing the mentioned communications' content analysis, we will face a new modified network that we call it the semantic network. In this new network, the weight of each edge represents the strength of the semantic relationship between the two corresponding individuals. This network is shown in Figure 1(e). As an example, the edge between node 3 and node 7 in the basic network has the weight of 3 (see Figure 1(a)), but after applying the results of communications' content analysis, this weight is modified to 0. The reason is that, these two nodes have totally different views about the subjects of their communications. Now, if we apply a topological community detection algorithm to the semantic network (network of Figure 1(e)), the identified communities will somewhat be different from the communities which were detected in the basic network. These new identified communities are shown in Figure 1(f). Although the strengths of connections between nodes of the communities in Figure 1(f) are based on the result of communications' content analysis, and not just the number of communications, if we look at these communities, we can infer that the community which is located at the left side of the figure incorporates different topics of interest. This also happened in Figure 1(b) before. In both situations, no topical subnetworks were generated.

Figure 1(g) shows the overlapping communities in which the members in each community have the same topic of interest, and the strengths of connections between these members are the consequence of their communications' content analysis and represent the rate of their semantic relationships. This paper aims to achieve these communities by proposing a generic framework with special focus on rating-based social networks since the overlapping community structures of this kind of social networks have not been well studied in the literature of community detection. On the other hand, finding the groups of individuals, who are like-minded and strongly connected, in a rating-based social network can have many potential applications, such as improving the efficiency of collaborative recommendations, in this kind of social networks. For example, suppose that a system is implemented for a movie rating social network to recommend movies to users based on the pattern of the users' previous ratings. If we could find communities of this network in which users are interested in the movies with the same genres (share the same topics of interest), and have dense connections (as the consequence of analyzing the values of their previous ratings), we can recommend suitable movies to a user. Thus, for a random user, named "User X", we can find the user's community, and the movies which have been rated by those members of the community who have a link to "User X" can be recommended to him/her. The advantage of this process is that the recommended movies have been rated by the users who have the same topic of interest as "User X" (users who are interested in the movies with the same genre as the movies which are favored by "User X"), and their dense connections to "User X" indicate the suitable rates of their viewpoints' unity.

3. Literature review

Many studies have been made in the area of community detection. Most of these studies mainly focus on the topological structures or linkage patterns of networks for finding communities [8]. Considering the community detection strategies which are employed in these studies, their proposed methods can be classified into disjoint community detection methods and overlapping community detection methods.

One of the most important studies in the literature was a research done by Newman and Girvan [12], in which they introduced Modularity. Since its introduction, many studies have been conducted to optimize Modularity, such as the methods which were proposed in [13-16]. This function has been influential in the literature of community detection, and has gained success in many applications. Modularity is used to evaluate the quality of a particular division of a network into communities [8]. Le Martelot and Hankin investigated stability, a measure for partition quality, as an optimization criterion that exploits a Markov process view of networks to enable multi-scale community detection [17]. Since community detection can be viewed as a clustering optimization problem, evolutionary computation and swarm-intelligence-based algorithms have a chance to be used for the community detection problem. Compared with traditional algorithms, intelligent optimization algorithms can effectively find a proper, high-quality solution within a reasonable period of time [18]. For this reason, Many single objective evolutionary algorithms, such as [19-23], along with multi-objective ones, such as [18, 24-33], have been proposed in recent years to solve the problem of community detection in complex networks. Most of the single-objective evolutionary algorithms aimed to find the community structure of a network with the largest value of Modularity while the multi-objective ones were proposed to optimize two conflicting objective functions of the community detection problem, such as Modularity and Normalized Mutual Information (NMI) [4]. Shang et al. proposed a community integration strategy for large-scale networks, based on a novel improved modularity density increment [34]. The Experiments conducted in their research showed that their proposed algorithm can efficiently utilize the network node information. In another research, Shang et al. proposed a method to identify communities in a large network [35]. Their proposed approach includes three stages: 1) preprocessing and preliminary labeling, 2) merging sub-communities, and 3) modifying the misclassified nodes. Their method efficiently makes a good balance between accuracy, stability and computation time for finding communities in a large network.

The majority of the algorithms, which were described in the previous paragraph, only find disjoint communities. However, communities often overlap to some extent in many real-world networks [10]. Evans and Lambiotte used a partition of the links of a network in order to uncover its community structure [36]. This approach allows for communities to overlap at nodes, so that nodes may be in more than one community. In another research, Evans and Lambiotte developed the idea to partition the edges of a weighted graph in order to uncover overlapping communities of its nodes [37]. Ahn et al. introduced communities as groups of links rather than nodes, and showed that this unorthodox approach successfully reconciles the antagonistic organizing principles of overlapping communities and hierarchy [38]. Liu addressed the fuzzy clustering problem for networks with gradient methods [39]. In her research, the hard clustering concept, in which a node belongs to only one cluster, has been extended to the fuzzy clustering concept, in which each node may belong to different clusters with nonzero probability. Li et al. proposed an improved multi-objective quantum-behaved particle swarm optimization (IMOQPSO) based on spectral-clustering to detect the overlapping community structures in complex networks [10]. Zhou et al. proposed an ant colony based overlapping community detection algorithm which mainly includes ants' location initialization, ants' movement and post processing phases [11]. Shang et al. proposed an algorithm, based on node membership grade and sub-communities integration, to detect community structure in networks [40]. Their algorithm can accurately obtain both non-overlapping communities and overlapping communities. Furthermore, their proposed algorithm employs a framework resembling label propagation, which has low time complexity and is suitable for detecting communities in large-scale networks. Some methods in the literature are proposed for overlapping community detection by considering the dynamics of a network [41, 42]. Wu et al. proposed a method for community detection via the clustering dynamics of a network [41]. In their proposed method, first, the initial phases of the nodes in the network are given randomly. Then, these initial phases evolve according to a set of dedicatedly designed differential equations. After a period of evolution, the phases of the nodes are naturally separated into several clusters (communities). For the networks with overlapping communities, the phases of the overlapping nodes evolve to the interspace of the two communities. Wu and Jiao proposed a discrete-time clustering model for modular networks, which acts as a bridge between the structure and the dynamics of the networks [42]. Based on their proposed model, they introduced DTD algorithm for community detection.

Even though the studies, which were described in the two previous paragraphs, gained success in some applications, they ignored the contents that are interchanged in networks since they mainly focus on the topological structures of the networks; As a result, the relationships between the members in these studies are mainly based on the total number of communications. On the other hand, these studies identify communities which often contain members interested in different topics, which mislead or mix the meanings of the community [8].

Another group of studies tends to partition the networks into different groups of nodes, in which every node has the same topic of interest. In other words, these studies focus on topic modeling through analyzing the contents of social objects. It should be considered that social objects refer to the objects like e-mails, which people communicate with each other through them. Several topic models have been proposed, such as LSA [43], pLSA [44], LDA [45], etc. Latent semantic analysis (LSA) is a widely adopted approach to map the high dimensional co-occurrence matrix into a lower dimensional representation as latent semantic space to reveal semantic relations between entities. Hofmann made a significant leap forward to LSA by proposing the probabilistic LSA (pLSA) where the detected clusters are more topic-oriented [44]. Blei et al. proposed the Latent Dirichlet Allocation (LDA) [45], a three-level hierarchical Bayesian model that models words and documents over an underlying set of topics, to avoid the pLSA's serious problems of over-fitting [46].

The studies, which were described in the previous paragraph, aim to find communities, in which all members have the same topic of interest, while they ignore the relationships between members; As a result, the communities detected by these studies tend to contain topologically-diverse subcommunities within each community [46].

In recent years, several studies have proposed approaches which consider both the contents that are interchanged in networks and the topological structures of the networks, in order to find more meaningful communities. Zhao et al. proposed a topic-oriented community detection approach based on social objects' clustering and link analysis [8]. In their proposed method, first they used a clustering algorithm to group all the social objects into topics. After that, they divided the members that are involved in those social objects into topical clusters. Each of the generated topical clusters was related to a distinct topic. Finally, they identified the communities by performing a link analysis on each topical cluster. Their proposed approach could identify the communities which reflect the topics and strengths of connections, simultaneously. Zhao and Ma proposed a framework to apply a semantically structured approach to the Web service community modeling and discovery [47]. Xia and Bu constructed a semantic network from semantic information extracted from user-comment contents, and then implemented a community-detection algorithm on the giant component of the constructed semantic network in order to find communities [48]. Bu et al. proposed a sock puppet detection algorithm which combines authorship-identification techniques and link analysis [49]. Zhu et al. combined classic ideas in topic modeling with a variant of mixed-membership block model, which is developed in the statistical physics community [50]. Bu et al. constructed interest networks using given social network data sets, mined the semantic information in these data sets, and updated the interest networks using the attitude consistence value [51]. In order to discover the communities in the updated interest networks, they proposed a new Modularity optimization algorithm. Grabowicz et al. used the identity and bond theory to define a set of features to classify groups into topical or social categories performing their experiments on a data set from Flickr [52]. The common identity and common bond theory states that people join groups based on identity (i.e., interest in the topics discussed) or bond attachment (i.e., social relationships). Yang et al. proposed an algorithm, named CESNA, for detecting overlapping communities in networks with node attributes [53]. CESNA statistically models the interaction between the network structure and the node attributes. Tchuente et al. proposed a community-based algorithm that is applied to a part of a user's social network (egocentric network), and derives the user social profile that can be reused for any purpose (e.g., personalization, recommendation) [54]. Wang et al. proposed a semantic method to analyze the topical community "fingerprint" in a social network [55]. Smith et al. proposed an information-theoretic method that identifies modules by compressing descriptions of information flow on a network [56]. Reihanian et al. evaluated the effect of topic consideration for finding meaningful communities in rating-based social networks without considering the value of ratings in the process of community detection [2]. With conducting experiments on real life data sets, they came to this conclusion that the results of community detection in rating-based social networks will be improved when the topic of interest is considered. Atzmueller et al. introduced a description-oriented community detection algorithm which aims to identify communities according to standard community quality measures, while providing characteristic descriptions of these communities at the same time [57]. Reihanian et al. proposed a multi-objective discrete Biogeography Based Optimization algorithm to find communities of social networks with node attributes [4]. Their proposed method tends to reach to a trade-off between similarity of nodes' attributes and density of connections in identified communities.

Each of the studies, which were described in the previous paragraph, analyzed the contents of a network from different aspects, and used the results of this analysis along with considering the topological structures of the network in order to identify communities. For example, some of these methods extracted the topics of interest in the networks while some of the other methods used the results of content analysis to build modified networks; As a result, one of the two following situations is occurred: in the first situation, the identified communities contain more than one topic of interest shared by their members. In the second one, the strengths of relationships between members in the networks are not based on the contents interchanged between them, and thus the identified communities contain members who usually don't have strong semantic relationships. Also, most of the mentioned methods can find disjoint communities, and are not able to find overlapping ones. In order to overcome the specified problems, we propose a generic framework with the special focus on rating-based social networks that considers topics, semantic relationships, and topological analyses in its process of community detection. This framework can automatically detect overlapping communities.

4. Semantic Network-based Topical Overlapping Community Detection (SNTOCD) framework

In this section, we present the Semantic Network-based Topical Overlapping Community Detection (SNTOCD) framework with special focus on rating-based social networks. In the rest of the paper, we refer to this framework as the SNTOCD framework. The SNTOCD framework identifies topical overlapping communities whose members have unique topics of interest, and the strengths of relationships between these members are directly based on their communications' content analysis. This section consists of two subsections. In Subsection 4.1, the steps of the SNTOCD framework are described in details, and Subsection 4.2 gives the summarization of the SNTOCD framework.

4.1. Framework steps

The SNTOCD framework is implemented in 4 steps. Each step of the SNTOCD framework is described in detail in the consecutive Subsections of 4.1.1 to 4.1.4. Generally, the SNTZOCD framework can be applied to a social network data set, in which the contents of communications between individuals are included.

We use a test rating-based data set for better clarifying the performance of the SNTOCD framework. We are going to check the output of applying each step of the SNTOCD framework to the test data set. This data set, which is a movie rating data set, can be seen in Table 1.

Table 1 A movie rating data set (a test data set) containing ratings from 8 users to 19 movies

Movies	User1	User2	User3	User4	User5	User6	User7	User8
M1	Pos	Neu	Pos	-	-	-	-	-
M2	Neg	Neu	Neu	-	-	-	-	-
M3	Neu	-	Neg	-	-	-	-	-
M4	-	Pos	Neu	-	-	-	-	-
M5	-	-	Pos	Pos	Pos	-	-	-
M6	-	-	Pos	Pos	Pos	-	-	-
M7	-	-	-	Neu	Neg	-	-	-
M8	-	-	Neg	-	-	-	Pos	-
M9	-	-	Neg	-	-	-	Neg	-
M10	-	-	Neg	-	-	-	Neu	-
M11	-	-	-	-	-	Neu	Neu	Neu
M12	-	-	-	-	-	Pos	-	Neu
M13	-	-	-	-	-	Neu	-	Neu
M14	-	-	-	-	-	Neu	-	Neg
M15	-	-	-	-	Pos	Pos	-	-
M16	-	-	-	-	Neg	Pos	-	-
M17	-	-	-	-	Neu	Neu	-	-
M18	-	-	-	-	Pos	Neg	-	-

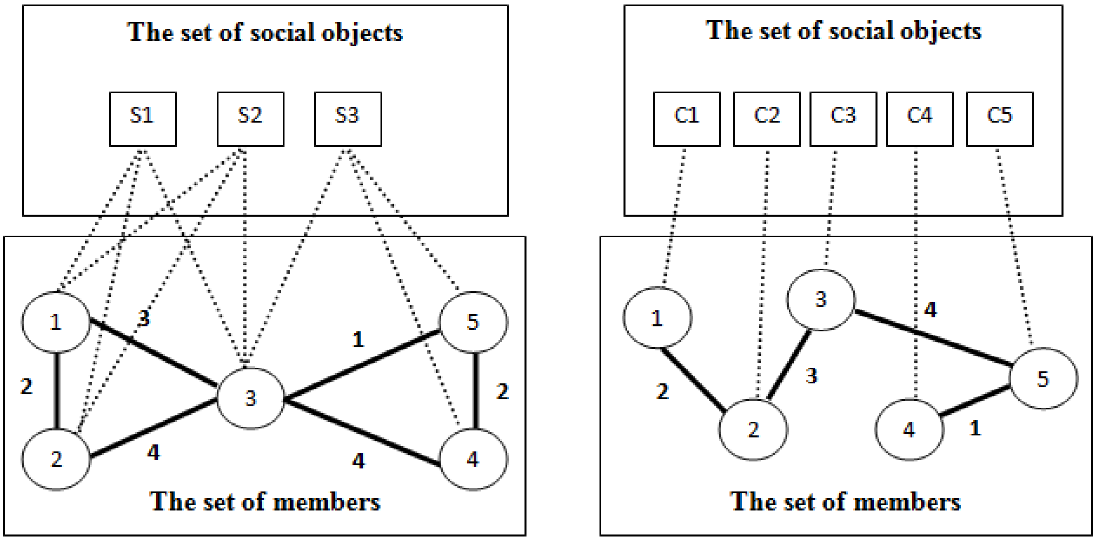
M19	-	-	-	-	Neg	Neg	-	-

According to Table 1, the movie rating data set contains ratings from 8 users to 19 movies. The ratings are categorized into three groups of negative (Neg), positive (Pos), and neutral (Neu). For example, user 1 gives ratings to 3 movies (M1, M2, and M3), and its rating to the first movie (M1) is positive (Pos).

4.1.1. Step 1: Preprocessing

Generally, People communicate with each other through social objects. These objects often imply the topics which people are interested in. A social object can be classified into one of the two following categories [8]: 1) attached to multi-members, 2) attached to only one member.

When social objects of a social network are from the first category, the edges between members of the network are built because of these social objects (please refer to part (a) of Figure 2). An example of this situation can be happened in a movie rating social network. In this network, edges between members are built when they rate the same movie. As a matter of fact, in this network, each movie (social object) is attached to multi members. The members of the movie rating social network are connected to each other due to the rating of the same movie [2].



(a) The social objects which are attached to multi-members

(b) The social objects which are attached to one member

Fig. 2. Two different kinds of social objects [2]

When social objects of a social network are from the second category, each social object is attached to only one member (please refer to part (b) of Figure 2). In this situation, the social objects are considered to be the attributes of the members of the network. An example of this situation can be happened in a paper citation network. In this network, papers (members) cite each other. Also, each paper contains a text content (the title of the paper), which is a social object, that can be considered as the attribute of the corresponding paper [2].

With considering the above explanations, in this step, a data set is preprocessed, and will be ready to use. In this process, the social objects of the data set are recognized. Afterwards, the topics of each social object are retrieved. Subsequently, each social object is labeled by its corresponding topic. In some cases the topics of each social object can be retrieved manually, or there are corresponding tags which represent the topics for each social object. But in cases where a social object is represented by text, and its labels cannot easily be retrieved, a method has been introduced in [8] which can identify the topic label of each social object. It should be considered that the data sets which are used in this paper (including the test data set) contain social objects with labeled topics.

Now, we want to check the output of applying the first step of the SNTOCD framework to the test data set of Table 1:

- First, the social objects of the data set are retrieved. Since communication between users (individuals) in this data set is indirectly performed by means of rating movies, these movies are considered as social objects.
- Afterwards, the topics of each movie in the data set are retrieved. As you know, the genre of a movie represents the general topic, in which a movie is made about. Thus, for the data set, the genres of the movies are extracted.
- Subsequently, each movie is labeled by its corresponding topic (genre). The first 4 movies (M1-M4) are in genre "G1" while the others are in genre "G2".

Table 2 shows the output of applying the first step of the SNTOCD framework to the test data set of Table 1.

Table 2 The output of applying the first step of the SNTOCD framework to the test data set of Table 1

Movies	Genres	User1	User2	User3	User4	User5	User6	User7	User8
M1	**G1**	Pos	Neu	Pos	-	-	-	-	-
M2	**G1**	Neg	Neu	Neu	-	-	-	-	-
M3	**G1**	Neu	-	Neg	-	-	-	-	-
M4	**G1**	-	Pos	Neu	-	-	-	-	-
M5	**G2**	-	-	Pos	Pos	Pos	-	-	-
M6	**G2**	-	-	Pos	Pos	Pos	-	-	-
M7	**G2**	-	-	-	Neu	Neg	-	-	-
M8	**G2**	-	-	Neg	-	-	-	Pos	-
M9	**G2**	-	-	Neg	-	-	-	Neg	-
M10	**G2**	-	-	Neg	-	-	-	Neu	-
M11	**G2**	-	-	-	-	-	Neu	Neu	Neu
M12	**G2**	-	-	-	-	-	Pos	-	Neu
M13	**G2**	-	-	-	-	-	Neu	-	Neu
M14	**G2**	-	-	-	-	-	Neu	-	Neg
M15	**G2**	-	-	-	-	Pos	Pos	-	-
M16	**G2**	-	-	-	-	Neg	Pos	-	-
M17	**G2**	-	-	-	-	Neu	Neu	-	-
M18	**G2**	-	-	-	-	Pos	Neg	-	-
M19	**G2**	-	-	-	-	Neg	Neg	-	-

4.1.2. Step 2: Creating topical subgroups

In the previous step, each social object has been annotated with a topic label. In this step, members are partitioned into different subgroups with considering the topic labels of the social objects they are involved in. In other words, in this step, we construct topical subgroups, in which every member has the same topic of interest. Thus, the total number of topical subgroups is equal to the number of topics of interest in the network. A user can be a member of several topical subgroups since it is common for a user to be interested in several topics.

Now, we check the output of applying the second step of the SNTOCD framework to the test data set. Since, in the previous step, we recognize two genres ("G1" and "G2") for the movies, we construct two topical subgroups for each genre in this step. For example, if the movie (social object) which was rated by user 1 has the topic (genre) "G1", in this step, the user 1 will join topical subgroup "G1".

Table 3 shows the rating information of the users who are the members of subgroup “G1”. Also, Table 4 shows the rating information of the users who are the members of subgroup “G2”. According to Table 3 and Table 4, there are 3 members in subgroup “G1”, 6 members in subgroup “G2”, and User 3 is the member of both subgroups since it rated movies in both genres.

Table 3 The first output of applying the second step of the SNTOCD framework to the test data set of Table 1: The members of topical subgroup “G1” along with their rating information

Movies	Genres	User1	User2	User3
M1	**G1**	Pos	Neu	Pos
M2	**G1**	Neg	Neu	Neu
M3	**G1**	Neu	-	Neg

Table 4 The second output of applying the second step of the SNTOCD framework to the test data set of Table 1: The members of topical subgroup “G2” along with their rating information

Movies	Genres	User3	User4	User5	User6	User7	User8
M5	**G2**	Pos	Pos	Pos	-	-	-
M6	**G2**	Pos	Pos	Pos	-	-	-
M7	**G2**	-	Neu	Neg	-	-	-
M8	**G2**	Neg	-	-	-	Pos	-
M9	**G2**	Neg	-	-	-	Neg	-
M10	**G2**	Neg	-	-	-	Neu	-
M11	**G2**	-	-	-	Neu	Neu	Neu
M12	**G2**	-	-	-	Pos	-	Neu
M13	**G2**	-	-	-	Neu	-	Neu
M14	**G2**	-	-	-	Neu	-	Neg
M15	**G2**	-	-	Pos	Pos	-	-
M16	**G2**	-	-	Neg	Pos	-	-
M17	**G2**	-	-	Neu	Neu	-	-
M18	**G2**	-	-	Pos	Neg	-	-
M19	**G2**	-	-	Neg	Neg	-	-

4.1.3. Step 3: Generating semantic topical subnetworks

In the first part of this step, we introduce a weighting strategy in order to compute the weight of the semantic relationships in a rating-based social network. With computing these weights, we describe a graph model, in the second part of the step, for transforming the topical subgroups of the previous step into semantic topical subnetworks.

4.1.3.1. Part 1: Computing the weight of each semantic relationship

Communications’ content is an important source for estimating the depth of users’ connections. Thus, in this part, a weighting strategy for rating-based social networks is introduced. This weighting strategy performs based on the value of ratings. In the following paragraphs, we are going to explain this weighting strategy.

With considering the studies related to sentiment analysis, first we divide the sentiments of users toward social objects into three categories: positive, negative, and neutral. This division is directly based on the value of users’ ratings. Then, we assign a numeric value to each rating based on its sentiment. Positive ratings are assigned 1, negative ratings are assigned -1, and neutral ratings are assigned 0. For example, after retrieving the rating of a user to a specific social object, 1 is assigned to the rating if the sentiment of it is positive. After assigning a numeric value to each rating, a relationship between each pair of users who rate the same social objects is considered. Each relationship can be the result of several communications. For example, if user1 and user2 rate five similar social objects, their relationship will be the consequence of their communications through rating these five social objects. In a rating-based social network, each communication between each two users is based on their ratings to the same social object. We calculate the weight of each communication as follows:

$$Value_{com_{ij}} = 1 - \left| Value_i - Value_j \right| \quad (1)$$

Where, Value_i and Value_j are the assigned numeric values of user_i's and user_j's ratings to a specified social object, respectively. $Value_{com_{ij}}$ is the weight of the communication between user_i and user_j through rating the mentioned social object. Table 5 represents the $\text{Value}_{\text{com}}$ of all possible communications between two users through different values of ratings to a special social object with considering the sentiment of each rating.

Table 5 The $\text{Value}_{\text{com}}$ of all possible communications between two users through different values of ratings to a special social object with considering the sentiment of each rating

	Negative (Neg)	**Neutral (Neu)**	**Positive (Pos)**
Negative (Neg)	+1	0	-1
Neutral (Neu)	0	1	0
Positive (Pos)	-1	0	+1

According to Table 5, if user_i gives a negative rate, and user_j gives a neutral rate to a specified social object, According to Table 5, their communication's weight will be 0. But the weight of a relationship between two users of each topical subgroup is the sum of the weights of their communications through the social objects which both of them rate. For example, if user_i and user_j rate five same social objects related to topical subgroup "A", and the $\text{Value}_{\text{com}}$ of these five communications be calculated as 1, 0, -1, 0, 1, the weight of their relationship, considering that topical subgroup, will be the sum of these five values, and will be equal to 1.

In case of positive relation weights, the value of semantic relationship between two users shows that the strength of their overall agreement toward different social objects is more than the strength of their disagreement. In case of negative relation weights, the value of semantic relationship between two users shows that the strength of their overall disagreement toward different social objects is more than the strength of their agreement. In case of zero relation weights, the value of semantic relationship between two users shows that the strength of their overall disagreement toward different social objects is equal to the strength of their agreement.

Since the goal of this step is to estimate the rate of the agreement of each pair of users or the rate of their viewpoints' unity, the weight of the relationship between each pair whose rate of disagreement is more than or equal to the rate of their agreement is not important for us. Thus, after calculating the weight of the semantic relationships between each pair of users, the negative values will be changed to zero. Each calculated relation weight in this step, quantifies the degree of alignment between the related members on their sentiments toward social objects which have the topic in question.

Now, we check the output of applying the first part of the third step of the SNTOCD framework to the test data set. As mentioned before, a communication between two users is considered when they rate

Table 6 All communications between each two users of topical subgroup "G1" of Table 3

the same social object. Also, it was mentioned before that a relationship between two users can be the result of their several mutual ratings (communications). Thus, With considering Table 3 and Table 4, Table 6 and Table 7 show the relationship between each two users of topical subgroup "G1" and topical subgroup "G2", respectively.

	User1	**User2**	**User3**
User1	-	Pos-Neu Neg-Neu	Pos-Pos Neg-Neu Neu-Neg
User2	Pos-Neu Neg-Neu	-	Neu-Pos Neu-Neu Pos-Neu
User3	Pos-Pos Neg-Neu	Neu-Pos Neu-Neu	-

Table 7 All communications between each two users of topical subgroup "G2" of Table 4

	User3	**User4**	**User5**	**User6**	**User7**	**User8**
User3	-	Pos-Pos Pos-Pos	Pos-Pos Pos-Pos	-	Neg-Pos Neg-Neg Neg-Neu	-
User4	Pos-Pos Pos-Pos	-	Pos-Pos Pos-Pos Neu-Neg	-	-	-
User5	Pos-Pos Pos-Pos	Pos-Pos Pos-Pos Neu-Neg	-	Pos-Pos Neg-Pos Neu-Neu Pos-Neg Neg-Neg	-	-
User6	-	-	Pos-Pos Neg-Pos Neu-Neu Pos-Neg Neg-Neg	-	Neu-Neu	Neu-Neu Pos-Neu Neu-Neu Neu-Neg
User7	Neg-Pos Neg-Neg Neg-Neu	-	-	Neu-Neu	-	Neu-Neu
User8		-	-	Neu-Neu Pos-Neu Neu-Neu Neu-Neg	Neu-Neu	-

For example, according to Table 6, user 1 has a relationship with user 2 by means of the two communications which are Pos-Neu and Neg-Neu. Pos-Neu represents the positive and neutral ratings of the two users to the same social object (movie) while Neg-Neu represents the negative and neutral ratings of the two users to the same social object (movie).

Now, with considering Table 5, the weight of the relationships between each two users of topical subgroup "G1" and topical subgroup "G2" are shown in Table 8 and Table 9, respectively.

Table 8 The first output of applying the first part of the third step of the SNTOCD framework to the test data set of Table 1: the weight of the relationships between each two users of topical subgroup "G1" of Table 3 with considering Table 5

	User1	User2	User3
User1	-	0	1
User2	0	-	1
User3	1	1	-

For example, if we consider the relationship between user 3 and user 7, we can find that according to Table 7, they have 3 communications which are Neg-Pos, Neg-Neg, and Neg-Neu. The weights of these communications, according to Table 5, are -1, +1, and 0, respectively. The sum of these communications' weights is equal to 0 which is the weight of the semantic relationship between user 3 and user 7 of topical subgroup "G2". Thus, the weight of the relationship between user 3 and user 7 is considered 0 in Table 9.

Table 9 The second output of applying the first part of the third step of the SNTOCD framework to the test data set of Table 1: the weight of the relationships between each two users of topical subgroup "G2" of Table 4 with considering Table 5

	User3	User4	User5	User6	User7	User8
User3	-	2	2	-	0	-
User4	2	-	2	-	-	-
User5	2	2	-	1	-	-
User6	-	-	1	-	1	2
User7	0	-	-	1	-	1
User8	-	-	-	2	1	-

4.1.3.2. Part 2: Transforming the topical subgroups into semantic topical subnetworks

With having the connection weights from the previous step, each topical subgroup of the second step is transformed into a semantic topical subnetwork in this step. In order to describe the subnetworks, we use a graph model which is a extension of the formal graph model proposed in [8]. In this model, the extended graph is defined as EG = (U, O, E, W), Where: U is the set of users who are involved in social activities, O is the set of objects (or contents) that members communicate with each other through them, E is the set of edges that represent the relationships which exist between members, and W is the set of the mentioned relationships' weights that are calculated by the weighting strategy described in the previous step. It should be considered that O can consist of more than one social object for each relationship since it is common that each pair of users communicate with each other through several social objects.

Now, we check the output of applying the second part of the third step of the SNTOCD framework to the test data set. With considering the connections weights between each two users of topical subgroup "G1" and topical subgroup "G2" (from Table 8 and Table 9), we use the introduced graph model to create semantic topical subnetworks from the mentioned topical subgroups. The semantic topical subnetwork "G1" is shown in Figure 3. This subnetwork is created from the topical subgroup "G1". Also, the semantic topical subnetwork "G2", which is created from the topical subgroup "G2", is shown in Figure 4. According to Figure 3 and Figure 4, user 3 is the member of both semantic topical subnetworks of “G1” and “G2”. The reason is that, this user belongs to both topical subgroups of “G1” and “G2” (according to Table 3 and Table 4).

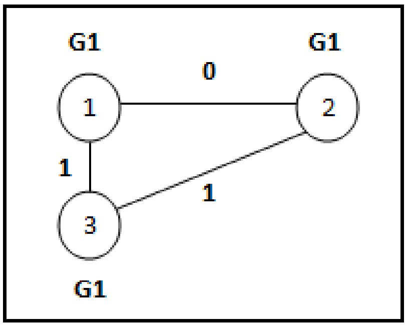


Fig. 3. The first output of applying the second part of the third step of the SNTOCD framework to the test data set of Table 1: the semantic topical subnetwork “G1” created from the topical subgroup “G1” of Table 3 using the connection weights represented in Table 8

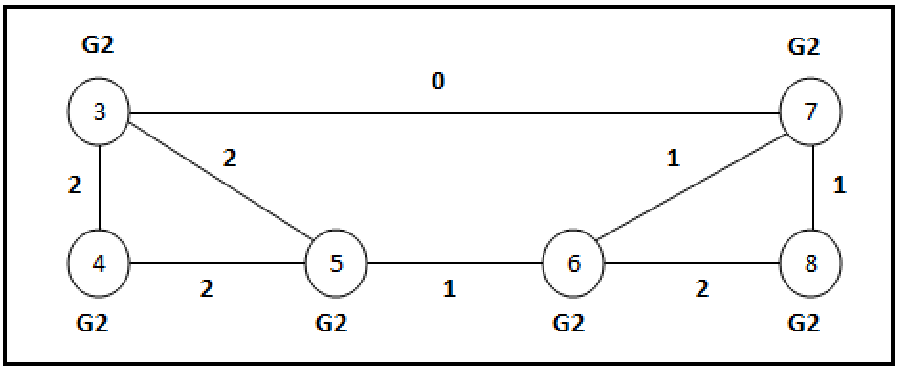


Fig. 4. The second output of applying the second part of the third step of the SNTOCD framework to the test data set of Table 1: the semantic topical subnetwork “G2” created from the topical subgroup “G2” of Table 4 using the connection weights represented in Table 9

4.1.4. Finding topical communities of the semantic topical subnetworks

This step aims to find communities in each of the semantic topical subnetworks which were created in the previous step. According to the previous step, members in each semantic topical subnetwork are connected to each other with different strengths. Based on the similarity of their viewpoints toward a social object, some members may have stronger connections, while some other may have weak or no connections. Thus, in this step, we apply a topological community detection algorithm to the previously created semantic topical subnetworks in order to identify their tightly connected groups of members (communities). Since the output of this step of the SNTOCD framework is the communities in which members are interested in the same topics of interests, we regard these communities as the topical communities.

In order to perform this step, many community detection algorithms can be employed. Newman and Girvan introduced Modularity and employed it in an important algorithm to partition network graphs of links and nodes into subgraphs [12]. Modularity evaluates communities from the perspective of the topological structure of a network since it is often used to evaluate whether the division is good in the sense that there are many edges within communities and only a few between them [8]. Since its introduction, many studies have been carried out to optimize Modularity as an objective function. For the weighted networks, Modularity has been defined as follows [16]:

$$Q=\frac{1}{2m}\sum_{i,j}\left[A_{ij}-\frac{k_i k_j}{2m}\right]\delta\left(c_i,c_j\right) \quad (2)$$

Where A_{ij} represents the weight of the edge between i and j, $k_i=\sum_j A_{ij}$ is the sum of the weights of the edges attached to vertex i, c_i is the community to which the vertex i is assigned, the δ function δ (u, v) will be 1 if u=v, and 0 otherwise, and also $m=\frac{1}{2}\sum_{ij}A_{ij}$.

The algorithm proposed by Newman and Girvan [12], attracted a large amount of attentions in the literature of community detection. Since the mentioned algorithm is very time-consuming, Blondel et al. suggest the modified version of the algorithm in order to make it faster, giving rise to what is known as the Louvain method [14]. This algorithm is a Modularity maximization algorithm which iteratively optimizes the Modularity in a local way, and aggregates the nodes of the same community [55]. Moving an isolated node i into a community C will cause a Modularity gain of ΔQ that can be calculated as follows [14]:

$$\Delta Q=\left[\frac{\sum_{in}+2k_{i,in}}{2m}-\left(\frac{\sum_{tot}+k_i}{2m}\right)^2\right]-\left[\frac{\sum_{in}}{2m}-\left(\frac{\sum_{tot}}{2m}\right)^2-\left(\frac{k_i}{2m}\right)^2\right] \quad (3)$$

Where $\sum_{in}$ is the sum of the weights of the links inside C, $\sum_{tot}$ is the sum of the weights of the links incident to node C, k_i is the sum of the weights of the links incident to node i, $k_{i,in}$ is the sum of the weights of the links from i to nodes in C, and m is the sum of the weights of all the links in the network.

Since the Louvain method is a promising Modularity-maximization algorithm, and outperforms many community detection algorithms in extensive number of experiments, this paper employs the Louvain method in order to find topical communities.

Now, we check the output of applying the fourth step of the SNTOCD framework to the test data set. For this reason, we apply the Louvain method, which is a topological community detection algorithm, to the semantic topical subnetworks that were created in the previous step (please refer to Figure 3 and Figure 4) in order to detect their topical communities. Figure 5 shows the topical community which is detected in the semantic topical subnetwork "G1". Also, the detected topical communities of semantic topical network "G2" are shown in Figure 6. According to Figure 5 and Figure 6, user 3 is a member of two topical communities (one in the semantic topical subnetwork of "G1", and the other in the semantic topical subnetwork of "G2"). Thus, it can be concluded that the SNTOCD framework is able to automatically detect topical overlapping communities.

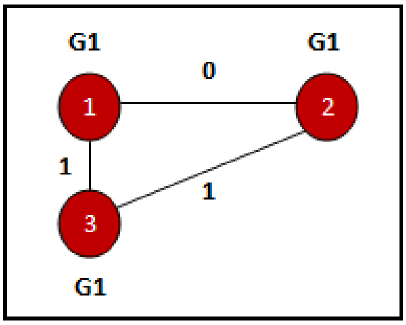


Fig. 5. The first output of applying the fourth step of the SNTOCD framework to the test data set of Table 1: the topical community which is detected in the semantic topical subnetwork "G1" of Figure 3 (the nodes with the same color are in the same community)

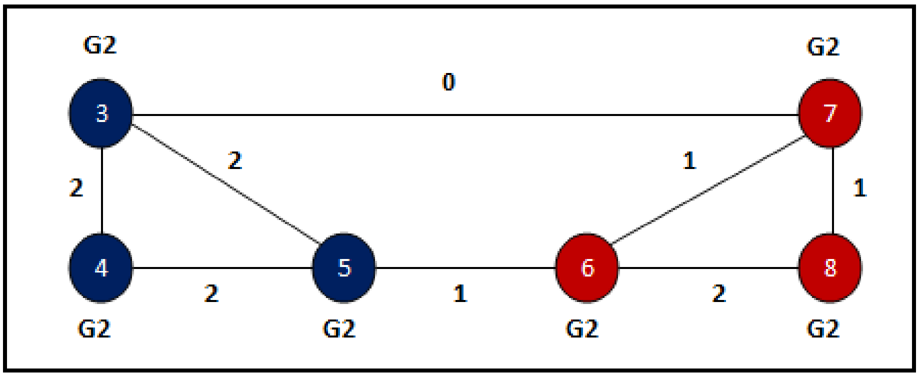


Fig. 6. The second output of applying the fourth step of the SNTOCD framework to the test data set of Table 1: the topical communities which are detected in the semantic topical subnetwork “G2” of Figure 4 (the nodes with the same color are in the same community)

4.2. The framework summarization

In the previous subsection, we explained how the SNTOCD framework operates. In order to implement the SNTOCD framework, the 4 steps are taken. First, the data sets are preprocessed, and each social object of it is annotated with its corresponding topic label. Then, members are partitioned into different subgroups with considering the topic labels of the social objects they are involved in. After that, the weight of the relationships between each pair of members of the created topical subgroups is computed according to the proposed weighting strategy, and the semantic topical subnetworks are constructed based on the assigned relationships’ weights. In the final step, the topical communities of the constructed semantic topical subnetworks are identified.

An alternative might be to create a semantic network first, and then partition the members of the semantic network into semantic topical subnetworks in order to find topical communities. This is considered as a candidate framework, and will be compared with the SNTOCD framework. The main strength of the SNTOCD framework, with respect to its process, is that it can detect more meaningful topical communities than the candidate framework.

We are going to explain the superiority of the SNTOCD framework in comparison with the candidate framework through an example. Figure 7 shows a basic network of 5 nodes and 6 edges. We assume that the topics of interest for each node (member) of the network are known and are labeled to it. We assume that all nodes in the network are interested in two topics. These topics are named as “T1” and “T2”. The number of communications between each pair of nodes is considered as the weights of the edges. For example, node 1 and node 2 have 6 communications. Thus, the weight of their related edge is 6. Moreover, we assume that one half of the communications between each two nodes is related to “T1”, and the other half is related to “T2”. Thus, as the weight of all edges in the network is 6, for each pair of nodes which have an edge between them, 3 communications are related to “T1”, and the other 3 communications are related to “T2”.

Figure 8 illustrates the processes of the candidate framework and the SNTOCD framework, respectively. It is considered that the first step of the both frameworks, which is preprocessing, has been performed in the creation of the basic network that is shown in Figure 7. According to Figure 8, we can see that both approaches find 4 topical communities from the network of Figure 7 in which 2 topical communities are related to “T1”, and the other 2 ones are related to “T2”. But there is a difference between these topical communities.

According to Figure 8, candidate approach builds a semantic network first. Thus, the weights of the edges are modified. Then, according to the topics of interest related to each member, two semantic topical subnetworks are created. Each of the two semantic topical subnetworks is related to a topic of interest. Since the weights of the edges in both semantic topical subnetworks are the same, the resulting topical communities are the same, too. But there is an acute problem here. As mentioned before, half of the communications between each pair of nodes is related to “T1”, and the other half is related to “T2”. However, the semantic network is build with no topic consideration. In other words, all 6 communications between each pair of nodes in the basic network of Figure 7 are analyzed in the process of building the semantic network. Thus, the semantic weights of the edges are directly related to all 6 communications between each pair of nodes. This is the reason why the semantic topical subnetworks, which are created considering the semantic network, are exactly the same.

We clarify the above explanations with an example. From 6 communications between node 1 and node 2, we consider that the semantic weight of the relationship for the 3 communications that are related to “T1” is 2, and the semantic weight of the relationship for the 3 communications that are related to “T2” is 1. As Figure 8 depicts, the weight of the relationship between node 1 and node 2 in both semantic topical subnetworks, which is created in the candidate approach, is 3. As mentioned before, it happens because modifying the weight edges (for creation of the semantic network) is performed before considering the topics of interest (in generation of the semantic topical subnetworks).

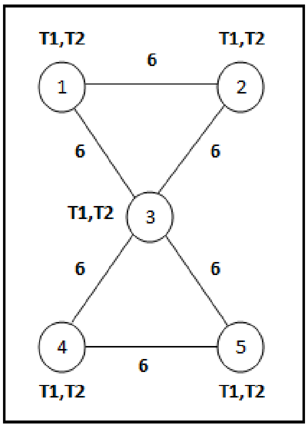
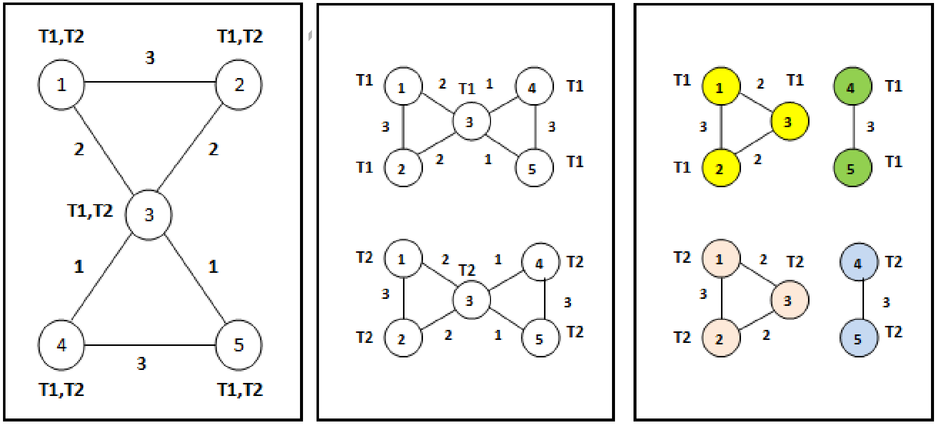
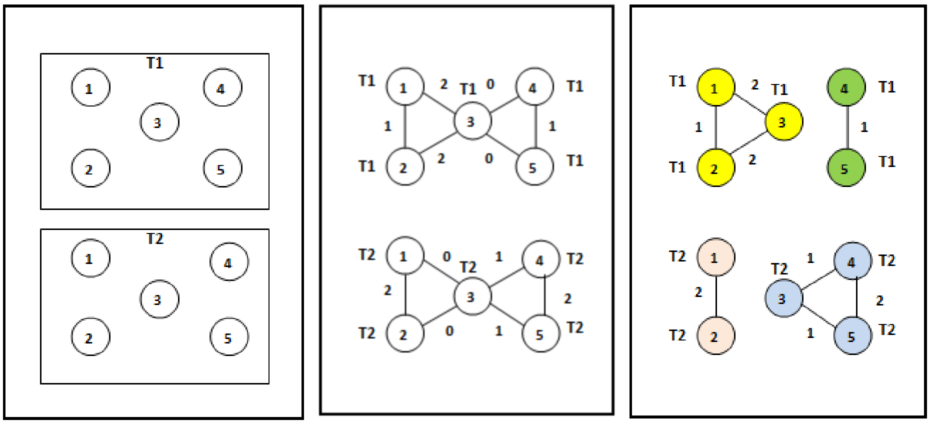


Fig. 7. An example of a basic network



(a) The process of community detection by the candidate framework. From left to right, Stage (1): Building a semantic network; Stage (2): Generating semantic topical subnetworks; Stage (3): Finding topical communities;



(b) The process of community detection by the SNTOCD framework. From left to right, Stage (1): Creating topical subgroups; Stage (2): Generating semantic topical subnetworks; Stage (3): Finding topical communities;

Fig. 8. The process of the candidate framework vs. the process of the SNTOCD framework

On the other hand, according to Figure 8, the topical subgroups are first created in the SNTOCD framework. As no communications' content analysis is done in the stage "Creating topical subgroups" of the SNTOCD framework, the communications can be split based on their related topics. Thus, in the next stage of the SNTOCD framework, the weights of the relationships in each semantic topical subnetwork are calculated directly based on the relevant topics.

According to Figure 8, the candidate approach finds 4 topical communities on 2 topics: T1={{1,2,3},{4,5}}, T2={{1,2,3},{4,5}}. The SNTOCD framework finds 4 topical communities on 2 topics, but with different members: T1={{1,2,3},{4,5}}, T2={{1,2},{3,4,5}}. As the SNTOCD framework assigns the weights to edges based on the related topics of interest in each of the semantic topical subnetworks, it can be concluded that the SNTOCD framework can detect more meaningful topical communities than the candidate framework.

5. Experiments

In this section, the results of our research are presented. First, 3 real-life data sets along with a performance metric are described. Then the results of applying the SNTOCD framework to the 3 mentioned data sets are analyzed. After that, the performance of the SNTOCD framework is compared with the performances of 3 other relevant approaches. All the experiments are conducted on a computer with Intel Core 2 Duo 2.20 GHz CPU and 2 GB RAM.

5.1. Real-life data sets

3 real-life data sets, which were used in our experiments, are described as follows:

MovieLens data set [58]: This rating data set was collected from the MovieLens web site (http://movielens.org). It consists of 100000 ratings from 943 users which were given to 1682 movies. Ratings are made on a 5-star scale with 1-star increments. Communications between users are made by rating the same movies.

Book-Crossing data set [59]: This data set was collected by Cai-Nicolas Ziegler from the Book-Crossing community (http://www.bookcrossing.com). It contains 278858 users providing 1149780 ratings about 271379 books. Ratings are expressed on a scale from 1 to 10. Communications between users are made by rating the same books.

CIAO data set [60]: The CIAO is a product review site (http://ciao.com), which provides a sensible platform to study trust in the online world. Users of this site, share their opinions about a product by means of rating or commenting. These products are divided into different categories, and the ratings are expressed on a scale from 1 to 5. There are 35773 ratings in this data set which are attached to 16850 products by 2248 users. Communications between users are made by rating the same products.

5.2. Performance metric

As described earlier in this paper, the SNTOCD framework considers the topics of interest, and uses the results of communications' content analysis for finding meaningful communities. The analysis of communications' content is used for estimating the depth of the semantic connections between members of a network. Thus, in order to evaluate the performance of the SNTOCD framework, two aspects should be considered: topic and linkage structure. It means that the expected results should keep each community's members with the same topic and strong semantic links.

In our experiments, we use the metric PurQ$_\beta$ which is introduced in [8]. This performance evaluation metric, which considers both topic and linkage structure, is defined as follows:

$$PurQ_\beta = \left(1+\beta^2\right)\left(Purity \cdot Q\right) / \left(\beta^2 \cdot Purity + Q\right) \quad (4)$$

The PurQ$_\beta$ has three parameters which are explained as follows:

- **Q** denotes the Modularity. As mentioned before, Modularity evaluates communities from the perspective of link structure [8]. In our experiments, for a partition of each semantic topical subnetwork, Modularity is calculated by equation 2. Since the SNTOCD framework may generate more than one semantic topical subnetwork for each data set, this paper proposes the following equation for calculating the total value of Modularity in the SNTOCD framework:

$$Q = \sum_{i=1}^{n} \frac{Weight_{STSN_i}}{Weight_T} \cdot Q_{STSN_i} \quad (5)$$

  Where n is the number of generated semantic topical subnetworks. Q_{STSN_i} is the value of Modularity for the semantic topical subnetwork STSN$_i$. $Weight_{STSN_i}$ is the sum of the weights of the edges in the semantic topical subnetwork STSN$_i$. Weight$_T$ is the sum of the weights of the edges in the semantic network. This network is the same as the basic network except for its edge weights which are calculated with the weighting strategy of the SNTOCD framework. The larger the Modularity, the better the communities are divided from the perspective of topological structure.

- **Purity** represents the purity of topics in the detected communities. It evaluates the communities from the perspective of topic consideration and is calculated as follows [8]:

$$Purity = 1/N_{cm} \cdot \sum_{i=1}^{N_{cm}} \max_{1 \le j \le k} \{n_{ij} / n_i\} \quad (6)$$

  Where N$_{cm}$ represents the number of detected communities, n$_{ij}$ refers to the number of nodes belonging to topic j and community i, n$_i$ refers to the number of nodes in community i, and k represents the number of topics in the network. The higher the Purity, the better the communities are partitioned from the perspective of topics.

- **β** is a parameter to adjust the weights of Purity and Modularity, and $\beta \in [0, \infty)$. If we consider the purity of topics and the topology of the network to be equally important, the value of β shall be set to 1. If we want to pay more attention to Purity in comparison with

Modularity, the value of β shall be set to a number in the range $[0,1)$. On the other hand, if we want to pay more attention to Modularity in comparison with Purity, the value of β shall be set to a number in the range $(1,\infty)$.

As a brief, $PurQ_\beta$ can make a balance between Purity and Modularity. β adjusts the emphasis of the two aspects which are topics and link structure [8].

5.3. Results of the SNTOCD framework

In order to apply the SNTOCD framework to the 3 data sets, we had to go through the 4 steps of the framework.

The first step was to preprocess the data sets. For MovieLens data set, movies were considered as social objects. Since the genre of a movie represents the general topic, in which a movie is made about, they were extracted and considered as topics for MovieLens data set. These extracted genres are the same as the genres attached to each movie by IMDB (http://www.imdb.com). Then, all the movies which were in the genres of “Documentary” or “Western” were retrieved. In this step, we achieved 77 movies. There were 50 movies in the genre of “Documentary” and 27 movies in the genre of “Western”. For Book-Crossing data set, books were considered as social objects. Thus, we extracted the categories of 93 books from Amazon (http://www.amazon.com). Since the category of a book represents its topic, they were extracted and considered as topics for Book-Crossing data set. In this data set, the books were partitioned into two categories of “Fiction” and “Non-Fiction”. The “Fiction” category contained 80 books, while the “Non-Fiction” category contained 13 books. For CIAO data set, products were considered as social objects. Each product’s category was attached to it in the data set, and was considered as its topic. The products in CIAO data set were partitioned into six categories of “DVDs”, “Books”, “Beauty”, “Music”, “Travel”, and “Food and Drink”. The “DVDs” category contained 2057 products, The “Books” category contained 2803 products, the “Beauty” category contained 2333 products, the “Music” category contained 1801 products, the “Travel” category contained 3922 products, and the “Food and Drink” category contained 3937 products.

The second step was to create topical subgroups. Thus, for each data set, the users who rate the social objects of the same genre (for the MovieLens data set) or the same category (for the Book-Crossing and CIAO data sets) were partitioned into the corresponding topical subgroup. For example, all users who rate the movies in the genre of “Documentary” were partitioned into the topical subgroup of “Documentary”. Thus, according to the number of topics, we achieved two topical subgroups for the MovieLens and Book-Crossing data sets and 6 topical subgroups for the CIAO data set.

The first part of the third step was to compute the weights of the semantic relationships between members of each topical subgroup. Thus, for each topical subgroup, we first retrieved the ratings which the members of the topical subgroup had given to the social objects related to the topic of the subgroup. Then we analyzed these ratings. For the MovieLens and CIAO data sets, if a member had rated a social object with a number greater than 3, we considered his/her sentiment toward that social object as positive. Also, if a member had rated a social object with a number less than 3, we considered his/her sentiment toward that social object as negative. Finally, if a member had rated a social object with 3, we considered his/her sentiment toward that social object as neutral. On the other hand, for the Book-Crossing data set, if a member had rated a social object with a number greater than 6, we considered his/her sentiment toward that social object as positive. Also, if a member had rated a social object with a number less than 5, we considered his/her sentiment toward that social object as negative. Finally, if a member had rated a social object with 5 or 6, we considered his/her sentiment toward that social object as neutral. After that, for each pair of members in a topical subgroup who rated the same social objects, which were related to that subgroup, we calculated their relationship weights with the weighting strategy introduced in Subsection 4.1.3.1.

The second part of the third step was to transform the topical subgroups, which were created in the second step, into semantic topical subnetworks. Thus, for each topical subgroup, we draw an edge between each two members who rated the same social objects related to that subgroup. The weight of these edges was the weights of the semantic relationships which were computed in the previous part.

The last step was to detect topical communities of the semantic topical subnetworks. Thus, we applied the Louvain method to each semantic topical subnetwork, which was created in the previous step, in order to find its topical communities. In our experiments, the Louvain method was applied 10 times to each semantic topical subnetwork, and 10 values for Modularity were obtained for each of these subnetworks. The mean of these achieved values of Modularity for each semantic topical subnetwork is reported in the "Modularity of Subnetwork (Q_{STSN})" column of Table 10.

Table 10 gives the results achieved by applying the SNTOCD framework to the MovieLens, Book-Crossing and CIAO data sets. In this table, the columns "Semantic Topical Subnetwork", "No. of Edges", and "No. of Nodes" represent the created semantic topical subnetworks in the process of applying the SNTOCD framework to the 3 data sets, the number of edges, and the number of nodes existing in each of the created subnetworks, respectively. Moreover, the columns "Total Modularity (Q)" and "Purity" denote the overall Modularity and Purity values for each data set which were calculated according to equation 5 and equation 6, respectively.

According to Table 10, Purity has its maximum value for each of the 3 data sets. The reason is that, the SNTOCD framework partitions the social network of each data set into semantic topical subnetworks, in which the members are interested in the same topic; As a result, the detected topical communities in each of these semantic topical subnetworks also incorporate members who are interested in the same topic. Thus, according to equation 6, the purity of topics for each data set of Table 10 reaches to 1 which is the maximum value of Purity.

Table 10 The results achieved by applying the SNTOCD framework to MovieLens, Book-Crossing, and CIAO data sets

Data set	Semantic Topical Subnetwork	No. of Edges	No. of Nodes	Modularity of Subnetwork (Q_{STSN})	Total Modularity (Q)	Purity
MovieLens	**Documentary**	15833	352	0.3129	0.1946	1
	Western	69369	491	0.1700		
Book-Crossing	**Fiction**	8531	1021	0.8836	0.8708	1
	Non-Fiction	1587	191	0.8110		
CIAO	**DVDs**	53916	1356	0.2617	0.3706	1
	Books	8999	904	0.4373		
	Beauty	5267	811	0.6936		
	Music	2076	569	0.7817		
	Travel	12905	867	0.4200		
	Food & Drink	29763	1193	0.4007		

It should be considered that it is possible for a certain user to be in several semantic topical subnetworks since the interest of people in several different topics is common. Thus, some of the members of the semantic topical subnetworks of each data set may be the same. For example, consider a case that a user of the MovieLens data set, named as “User X”, rated several different movies. Some of these movies are in the genre of “Documentary”, and the others are in the genre of “Western”. Thus, “User X” belongs to both semantic topical subnetworks of the MovieLens data set. Since each user in a semantic topical subnetwork is a member of a topical community, “User X” is a member of two topical communities (one in the semantic topical subnetwork of “Documentary”, and the other in the semantic topical subnetwork of “Western”). With the above explanations, it can be concluded that the SNTOCD framework is able to automatically detect topical overlapping communities.

5.4. Comparison

In order to prove the superiority of the SNTOCD framework, in this subsection, we compare its results with the results of 3 other related frameworks. The first of these frameworks, which we call it the Classical Community Detection framework, applies a topological community detection algorithm to a basic network for the sake of community detection. As previously mentioned, the weight of an edge in a basic network represents the number of communications between its relevant nodes. This is a framework for classical community detection methods, in which finding the communities is done by only considering the topological structure of a network. In the process of this framework, no content analysis is performed. The second framework, which is called the Topic-oriented Community Detection framework, partitions the members involved in social objects into different topical clusters. Since the members in each topical cluster are connected with different strengths, a topological community detection algorithm is applied to each topical cluster in order to detect its communities. This framework has been proposed in [8]. The third framework, which is called the Semantic Network-based Community Detection framework, considers the communications’ content analysis in order to build a semantic network. In this condition, no topic consideration is performed. After creating a semantic network, a topological community detection algorithm is applied to this network in order to find meaningful communities. This framework has been proposed in [48]. The 3 mentioned frameworks were implemented and were applied to the 3 data sets which were explained in Subsection 5.1. After that, the performance metric defined in Subsection 5.2 was used for evaluating the experimental results.

We first applied the Louvain method [14] to the basic networks of the MovieLens, the Book-Crossing, and the CIAO data sets in order to find their communities (implementing the Classical Community Detection framework). Also, we partitioned the members of the 3 data sets into topical clusters. Each topical cluster included members which had the same topic of interest. Afterwards, the Louvain method was applied to these topical clusters for finding their communities (implementing the Topic-oriented Community Detection framework). On the other hand, the results of the communications’ content analysis (ratings analysis) were applied to the basic networks of the 3 data sets, and their semantic networks were built. Then, the Louvain method was applied to these semantic networks in order to identify their communities (implementing the Semantic Network-based Community Detection framework). After achieving the performance results of the 3 frameworks, we used $PurQ_\beta$ to compare them with the performance results of the SNTOCD framework. The corresponding results of $PurQ_\beta$ are given in Table 11.

Table 11 The values of Modularity, Purity, and $PurQ_\beta$ which are achieved by applying the SNTOCD framework along with the other 3 related frameworks to each of the 3 data sets used in the experiments

Data set	Framework	Modularity (Q)	Purity	$PurQ_\beta$				
				β=0.5	β=0.75	β=1	β=1.5	β=2
MovieLens	**Classical**	0.1086	0.9777	0.3760	0.2519	0.1955	0.1495	0.1321

	Topic-oriented	0.1244	1	0.4154	0.2830	0.2213	0.1703	0.1509
	Semantic Network-based	0.1627	0.8948	0.4710	0.3415	0.2753	0.2174	0.1945
	SNTOCD	**0.1946**	**1**	**0.5472**	**0.4017**	**0.3259**	**0.2588**	**0.2320**
Book-Crossing	**Classical**	0.8375	0.9050	0.8906	0.8795	0.8699	0.8572	0.8502
	Topic-oriented	0.8469	1	0.9651	0.9389	0.9171	0.8888	0.8737
	Semantic Network-based	0.8627	0.9833	0.9566	0.9362	0.9191	0.8965	0.8844
	SNTOCD	**0.8708**	**1**	**0.9712**	**0.9493**	**0.9309**	**0.9069**	**0.8939**
CIAO	**Classical**	0.2899	0.8279	0.6038	0.4963	0.4294	0.3624	0.3332
	Topic-oriented	0.3086	1	0.6906	0.5535	0.4716	0.3920	0.3581
	Semantic Network-based	0.3167	0.9723	0.6876	0.5571	0.4778	0.3996	0.3661
	SNTOCD	**0.3706**	**1**	**0.7464**	**0.6205**	**0.5407**	**0.4596**	**0.4239**

As mentioned before, in order to adjust the weights of Purity and Modularity in $PurQ_\beta$, β has been introduced. This parameter can be set to a value more than or equal to 0 ($\beta \in [0, \infty)$). When the topic is more important for us, and we want the purity of them to have more effect on the values of $PurQ_\beta$, we set β to the values less than 1 (0< β <1). According to equation 7, when β approaches zero, $PurQ_\beta$ approaches Purity:

$$\lim_{\beta \to 0} PurQ_\beta = \lim_{\beta \to 0} \frac{(1+\beta^2)(Purity \cdot Q)}{(\beta^2 \cdot Purity + Q)} = \frac{Purity \cdot Q}{Q} = Purity \quad (7)$$

Thus, it can be concluded that when the value of β tends toward 0, the effect of Purity on the values of $PurQ_\beta$ will increase. On the other hand, when the link is more important for us, and we want Modularity to have more effect on the values of $PurQ_\beta$, we set β to the values more than 1 (β >1). According to equation 8, when β approaches ∞, $PurQ_\beta$ approaches Modularity (Q):

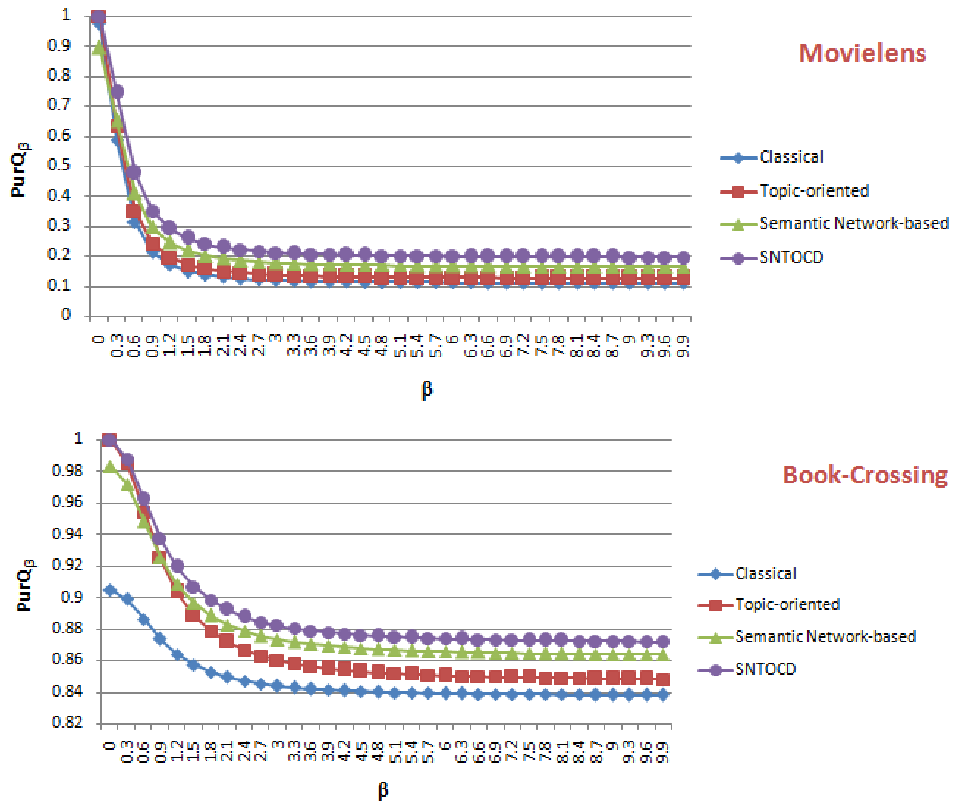
$$\lim_{\beta \to \infty} PurQ_\beta = \lim_{\beta \to \infty} \frac{(1+\beta^2)(Purity \cdot Q)}{(\beta^2 \cdot Purity + Q)} = \lim_{\beta \to \infty} \frac{\beta^2 \cdot Purity \cdot Q + Purity \cdot Q}{(\beta^2 \cdot Purity + Q)} = \frac{Purity \cdot Q}{Purity} = Q \quad (8)$$

Thus, it can be concluded that when the value of β tends toward ∞, the effect of Modularity on the values of $PurQ_\beta$ will increase. In the case that we want the purity of topics and the topology of the network to have equal effects on the values of $PurQ_\beta$, β should be set to 1. In this case $PurQ_\beta$ is the harmonic mean of Purity and Modularity (Q). According to equation 9, when β is set to 1, $PurQ_\beta$ is calculated as follows:

$$PurQ_1 = (1+1^2)(Purity \cdot Q)/(1^2 \cdot Purity + Q) = \frac{2Purity \cdot Q}{Purity + Q} \quad (9)$$

Consequently, as it is shown in Table 11, β was set to 0.5, 0.75, 1, 1.5, and 2, respectively, which represents the consideration of different strengths for topics and links of the networks in question. The values of Purity, Modularity, and $PurQ_\beta$ have been calculated for all of the frameworks after applying them to the 3 data sets used in the experiments. These calculated values are given in Table 11. According to this table, considering the first 3 frameworks, Purity has reached to higher values in the second framework since the basic networks of the data sets are partitioned into topical clusters, in which each identified community includes members who have the same topic of interest. Modularity has reached to higher values in the third framework in comparison with its values in the first and the second frameworks. Because in this framework, the communications' content analysis of the 3 data sets has been used for building the semantic networks in which the weights of the edges are directly based on the communications' content analysis, and represent the semantic relationship between the corresponding nodes. The higher values of Modularity in the third framework -in which the semantic network is created- show that the community structures detected by this framework are more significant than those detected by the first and the second frameworks. As mentioned before, the SNTOCD framework combines both topic consideration and communications' content analysis for finding communities. According to Table 11, the SNTOCD framework reaches to higher values of Purity -because of topic consideration- and Modularity -because of communications' content analysis- , and thus has higher values of $PurQ_{\beta}$ for all five values of β in comparison with the other 3 frameworks.

Figure 9 graphically illustrates the results of $PurQ_{\beta}$ and its mean values for all of the frameworks with considering the 3 data sets used in the experiments. In order to better clarify the performance of each framework in different conditions, β was set on a scale from 0 to 9.9 with 0.3 unit increments.



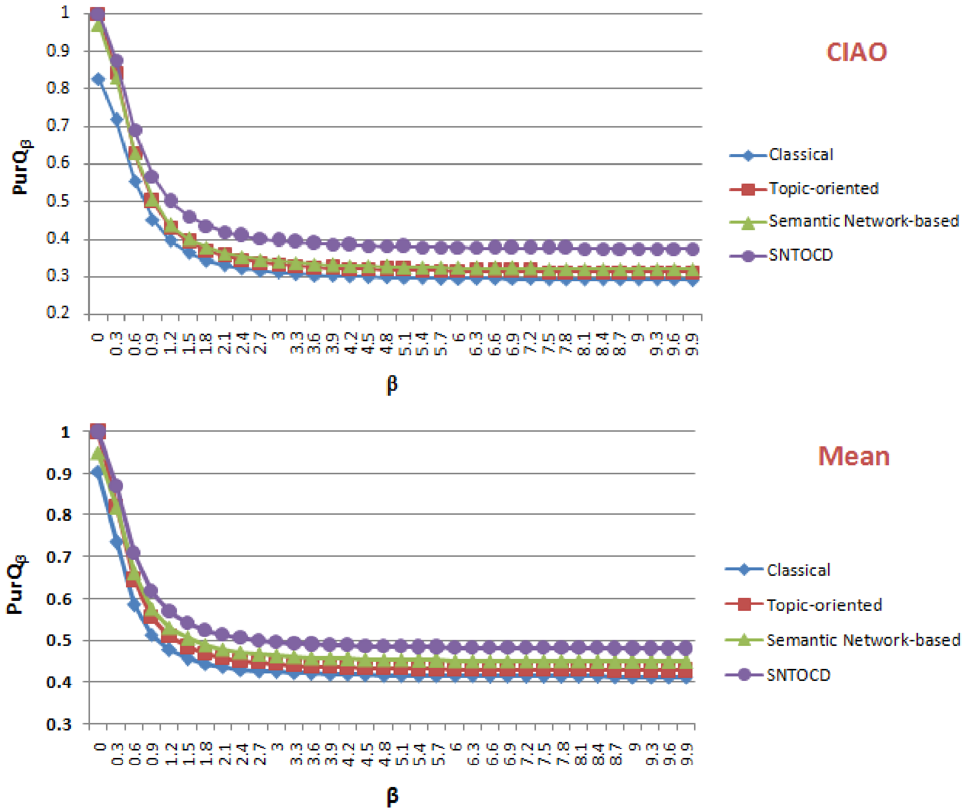


Fig. 9. The performance comparison of the 4 frameworks (Classical, Topic-oriented, Semantic Network-based, and SNTOCD) by considering their achieved values of PurQ_β for each of the 3 data sets (MovieLens, Book-Crossing, and CIAO) along with their achieved mean values of PurQ_β for all of the 3 data sets (Mean) under different values of β from 0 to 9.9 with 0.3 unit increments

According to Table 11 and Figure 9, the performance of the Topic-oriented framework is better than those of the Semantic Network-based and the Classical frameworks, when the value of β tends toward 0. As a matter of fact, when β is set to zero, the performance of the SNTOCD framework and the Topic-oriented framework will be the same. The reason is that, the SNTOCD framework and the Topic-oriented framework identify communities, in which all of the members have the same topic of interest, and thus, when β is set to zero, the values of Purity (Total Purity) for these two frameworks will be equal to 1 (as shown in Table 11). But according to Table 11, the values of Purity achieved by the Classical and the Semantic Network-based frameworks for all of the data sets are less than 1 since they identify communities in which the members may be interested in different topics.

According to Table 11 and Figure 9, when the value of β increases, the corresponding values of PurQ_β for all of the frameworks tend to move toward the relevant values of Modularity which is given in Table 11. According to the Mean section of Figure 9, as the value of β increases, the Semantic Network-based framework shows a better performance in comparison with the Classical and the Topic-oriented frameworks. The reason is that, with considering Table 11, the mean value of achieved Modularity values in the Semantic Network-based framework is higher than those in the Classical and the Topic-oriented frameworks since the semantic networks are created in the Semantic Network-based framework.

According to Table 11 and Figure 9, the Classical framework shows the worst performance among all frameworks in the experiments. The reason is that, in the process of Classical framework, no content analysis is performed. Thus, it can be concluded that the performance of the approach which performs content analysis in its process is superior to that of the approach, in which no content analysis is performed.

According to Table 11 and Figure 9, the SNTOCD framework shows a better performance in comparison with the other frameworks. The structure of the SNTOCD framework, which is explained in Section 4, allows it to effectively use the strengths of the Topic-oriented and the Semantic Network-based frameworks, altogether. Actually, the SNTOCD framework partitions a social network into semantic topical subnetworks to use the advantage of topic consideration in the Topic-oriented framework (for obtaining the high values of Purity) and the advantage of creating the semantic network in the Semantic Network-based framework (for achieving the high values of Modularity), and then finds communities in each of the semantic topical subnetworks. This is the reason why the SNTOCD framework gains the higher values of $PurQ_\beta$, and thus shows a better performance in comparison with the other 3 related frameworks.

Now, it can be concluded that the SNTOCD framework achieves a better performance in all of these three conditions:

- When the topic is as important for us as the link (β =1).
- When the topic is more important for us than the link (0< β <1).
- When the link is more important for us than the topic (β >1).

6. Conclusion

In this paper, a generic framework has been proposed for overlapping community detection in social networks with special focus on rating-based social networks. This framework has been called the Semantic Network-based Topical Overlapping Community Detection framework (the SNTOCD framework). For implementing the SNTOCD framework, first, a data set is preprocessed. In this process, social objects of the data set are recognized and their topics are retrieved and are labeled to them. Then, all individuals involved in those social objects are partitioned into topical subgroups. Each of these topical subgroups contains members interested in social objects with the same topic. In order to measure the semantic relationships between members of each topical subgroup, a new weighting strategy, which is based on rating, has been proposed. For each topical subgroup, a semantic topical subnetwork is built with the edge between each two individuals who rate the same social objects. The weights of the edges in these semantic topical subnetworks are calculated with the proposed weighting strategy. The final part of the presented SNTOCD framework is the detection of topical overlapping communities that is performed by applying the Louvain method, which is a topological community detection algorithm, to each semantic topical subnetwork. The structure of the SNTOCD framework enables it to automatically detect overlapping communities.

To evaluate the performance of the SNTOCD framework, experiments were conducted on 3 real-life data sets. Furthermore, the approaches of 3 relevant frameworks were implemented in order to compare them with the SNTOCD framework. With considering the qualities of the detected communities in the experiments, the SNTOCD framework performs better than the other 3 relevant frameworks. The reason is that two different aspects of content analysis, which are topic consideration and communications' content analysis, are considered in the process of SNTOCD framework for the detection of overlapping communities. Quantitative evaluations also indicated that the SNTOCD framework has a better performance in comparison with the other 3 relevant frameworks.

References

[1] K. Choi, K.-A. Toh, H. Byun, Incremental face recognition for large-scale social network services, Pattern Recognition, 45 (2012) 2868-2883.

[2] A. Reihanian, B. Minaei-Bidgoli, H. Alizadeh, Topic-oriented community detection of rating-based social networks, Journal of King Saud University-Computer and Information Sciences, 28 (2016) 303-310.

[3] A. Lancichinetti, S. Fortunato, Consensus clustering in complex networks, Scientific reports, 2 (2012).

[4] A. Reihanian, M.-R. Feizi-Derakhshi, H.S. Aghdasi, Community detection in social networks with node attributes based on multi-objective biogeography based optimization, Engineering Applications of Artificial Intelligence, 62 (2017) 51-67.

[5] J. Leskovec, K.J. Lang, M. Mahoney, Empirical comparison of algorithms for network community detection, Proceedings of the 19th international conference on World wide web, ACM2010, pp. 631-640.

[6] M. Newman, Communities, modules and large-scale structure in networks, Nature Physics, 8 (2012) 25-31.

[7] C. Jia, M.B. Carson, X. Wang, J. Yu, Concept Decompositions for Short Text Clustering by Identifying Word Communities, Pattern Recognition, (2017).

[8] Z. Zhao, S. Feng, Q. Wang, J.Z. Huang, G.J. Williams, J. Fan, Topic oriented community detection through social objects and link analysis in social networks, Knowledge-Based Systems, 26 (2012) 164-173.

[9] A. Reihanian, B. Minaei-Bidgoli, M. Yousefnezhad, Evaluating the effect of topic consideration in identifying communities of rating-based social networks, Information and Knowledge Technology (IKT), 2015 7th Conference on, IEEE2015, pp. 1-6.

[10] Y. Li, Y. Wang, J. Chen, L. Jiao, R. Shang, Overlapping community detection through an improved multi-objective quantum-behaved particle swarm optimization, Journal of Heuristics, 21 (2015) 549-575.

[11] X. Zhou, Y. Liu, J. Zhang, T. Liu, D. Zhang, An ant colony based algorithm for overlapping community detection in complex networks, Physica A: Statistical Mechanics and its Applications, 427 (2015) 289-301.

[12] M.E. Newman, M. Girvan, Finding and evaluating community structure in networks, Physical review E, 69 (2004) 026113.

[13] A. Arenas, J. Duch, A. Fernández, S. Gómez, Size reduction of complex networks preserving modularity, New Journal of Physics, 9 (2007) 176.

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

[15] E.A. Leicht, M.E. Newman, Community structure in directed networks, Physical review letters, 100 (2008) 118703.

[16] M.E. Newman, Analysis of weighted networks, Physical Review E, 70 (2004) 056131.

[17] E. Le Martelot, C. Hankin, Multi-scale community detection using stability optimisation within greedy algorithms, arXiv preprint arXiv:1201.3307, (2012).

[18] D. Chen, F. Zou, R. Lu, L. Yu, Z. Li, J. Wang, Multi-objective optimization of community detection using discrete teaching–learning-based optimization with decomposition, Information Sciences, (2016).

[19] E.A. Hassan, A.I. Hafez, A.E. Hassanien, A.A. Fahmy, Community detection algorithm based on artificial fish swarm optimization, Intelligent Systems' 2014, Springer2015, pp. 509-521.

[20] A.-M. Karimi-Majd, M. Fathian, B. Amiri, A hybrid artificial immune network for detecting communities in complex networks, Computing, 97 (2015) 483-507.

[21] Z. Li, J. Liu, A multi-agent genetic algorithm for community detection in complex networks, Physica A: Statistical Mechanics and its Applications, 449 (2016) 336-347.

[22] L. Ma, M. Gong, J. Liu, Q. Cai, L. Jiao, Multi-level learning based memetic algorithm for community detection, Applied Soft Computing, 19 (2014) 121-133.

[23] M. Tasgin, A. Herdagdelen, H. Bingol, Community detection in complex networks using genetic algorithms, arXiv preprint arXiv:0711.0491, (2007).

[24] B. Amiri, L. Hossain, J.W. Crawford, R.T. Wigand, Community detection in complex networks: Multi–objective enhanced firefly algorithm, Knowledge-Based Systems, 46 (2013) 1-11.

[25] Q. Cai, M. Gong, L. Ma, S. Ruan, F. Yuan, L. Jiao, Greedy discrete particle swarm optimization for large-scale social network clustering, Information Sciences, 316 (2015) 503-516.

[26] M. Gong, Q. Cai, X. Chen, L. Ma, Complex network clustering by multiobjective discrete particle swarm optimization based on decomposition, IEEE Transactions on Evolutionary Computation, 18 (2014) 82-97.

[27] M. Gong, L. Ma, Q. Zhang, L. Jiao, Community detection in networks by using multiobjective evolutionary algorithm with decomposition, Physica A: Statistical Mechanics and its Applications, 391 (2012) 4050-4060.

[28] C. Liu, J. Liu, Z. Jiang, A multiobjective evolutionary algorithm based on similarity for community detection from signed social networks, IEEE transactions on cybernetics, 44 (2014) 2274-2287.

[29] C. Pizzuti, A multiobjective genetic algorithm to find communities in complex networks, IEEE Transactions on Evolutionary Computation, 16 (2012) 418-430.

[30] C. Shi, Z. Yan, Y. Cai, B. Wu, Multi-objective community detection in complex networks, Applied Soft Computing, 12 (2012) 850-859.

[31] Z. Yuxin, L. Shenghong, J. Feng, Overlapping community detection in complex networks using multi-objective evolutionary algorithm, Computational and Applied Mathematics, (2015) 1-20.

[32] X. Zhou, Y. Liu, B. Li, G. Sun, Multiobjective biogeography based optimization algorithm with decomposition for community detection in dynamic networks, Physica A: Statistical Mechanics and its Applications, 436 (2015) 430-442.

[33] L. Li, L. Jiao, J. Zhao, R. Shang, M. Gong, Quantum-behaved discrete multi-objective particle swarm optimization for complex network clustering, Pattern Recognition, 63 (2017) 1-14.

[34] R. Shang, W. Zhang, L. Jiao, R. Stolkin, Y. Xue, A community integration strategy based on an improved modularity density increment for large-scale networks, Physica A: Statistical Mechanics and its Applications, 469 (2017) 471-485.

[35] R. Shang, H. Liu, L. Jiao, A.M.G. Esfahani, Community mining using three closely joint techniques based on community mutual membership and refinement strategy, Applied Soft Computing, 61 (2017) 1060-1073.

[36] T. Evans, R. Lambiotte, Line graphs, link partitions, and overlapping communities, Physical Review E, 80 (2009) 016105.

[37] T.S. Evans, R. Lambiotte, Line graphs of weighted networks for overlapping communities, The European Physical Journal B-Condensed Matter and Complex Systems, 77 (2010) 265-272.

[38] Y.-Y. Ahn, J.P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks, arXiv preprint arXiv:0903.3178, (2009).

[39] J. Liu, Detecting the fuzzy clusters of complex networks, Pattern Recognition, 43 (2010) 1334-1345.

[40] R. Shang, S. Luo, Y. Li, L. Jiao, R. Stolkin, Large-scale community detection based on node membership grade and sub-communities integration, Physica A: Statistical Mechanics and its Applications, 428 (2015) 279-294.

[41] J. Wu, L. Jiao, C. Jin, F. Liu, M. Gong, R. Shang, W. Chen, Overlapping community detection via network dynamics, Physical Review E, 85 (2012) 016115.

[42] J. Wu, Y. Jiao, Clustering dynamics of complex discrete-time networks and its application in community detection, Chaos: An Interdisciplinary Journal of Nonlinear Science, 24 (2014) 033104.

[43] S.C. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, R.A. Harshman, Indexing by latent semantic analysis, JASIS, 41 (1990) 391-407.

[44] T. Hofmann, Probabilistic latent semantic indexing, Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM1999, pp. 50-57.

[45] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, the Journal of machine Learning research, 3 (2003) 993-1022.

[46] Y. Ding, Community detection: Topological vs. topical, Journal of Informetrics, 5 (2011) 498-514.

[47] A. Zhao, Y. Ma, A Semantically Structured Approach to Service Community Discovery, Semantics, Knowledge and Grids (SKG), 2012 Eighth International Conference on, IEEE2012, pp. 136-142.

[48] Z. Xia, Z. Bu, Community detection based on a semantic network, Knowledge-Based Systems, 26 (2012) 30-39.

[49] Z. Bu, Z. Xia, J. Wang, A sock puppet detection algorithm on virtual spaces, Knowledge-Based Systems, 37 (2013) 366-377.

[50] Y. Zhu, X. Yan, L. Getoor, C. Moore, Scalable text and link analysis with mixed-topic link models, Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM2013, pp. 473-481.

[51] Z. Bu, C. Zhang, Z. Xia, J. Wang, A fast parallel modularity optimization algorithm (FPMQA) for community detection in online social network, Knowledge-Based Systems, 50 (2013) 246-259.

[52] P.A. Grabowicz, L.M. Aiello, V.M. Eguíluz, A. Jaimes, Distinguishing topical and social groups based on common identity and bond theory, Proceedings of the sixth ACM international conference on Web search and data mining, ACM2013, pp. 627-636.

[53] J. Yang, J. McAuley, J. Leskovec, Community detection in networks with node attributes, 2013 IEEE 13th International Conference on Data Mining, IEEE2013, pp. 1151-1156.

[54] D. Tchuente, M.-F. Canut, N. Jessel, A. Peninou, F. Sèdes, A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP, Social Network Analysis and Mining, 3 (2013) 667-683.

[55] D. Wang, K. Kwon, J. Sohn, B.-G. Joo, I.-J. Chung, Community Topical "Fingerprint" Analysis Based on Social Semantic Networks, Advanced Technologies, Embedded and Multimedia for Human-centric Computing, Springer2014, pp. 83-91.

[56] L.M. Smith, L. Zhu, K. Lerman, A.G. Percus, Partitioning Networks with Node Attributes by Compressing Information Flow, arXiv preprint arXiv:1405.4332, (2014).

[57] M. Atzmueller, S. Doerfel, F. Mitzlaff, Description-oriented community detection using exhaustive subgroup discovery, Information Sciences, 329 (2016) 965-984.

[58] F.M. Harper, J.A. Konstan, The movielens datasets: History and context, ACM Transactions on Interactive Intelligent Systems (TiiS), 5 (2016) 19.

[59] C.-N. Ziegler, S.M. McNee, J.A. Konstan, G. Lausen, Improving recommendation lists through topic diversification, Proceedings of the 14th international conference on World Wide Web, ACM2005, pp. 22-32.

[60] J. Tang, H. Gao, H. Liu, A. Das Sarma, eTrust: Understanding trust evolution in an online world, Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM2012, pp. 253-261.

Ali Reihanian received his B.Sc. degree in Information Technology (IT) from Mazandaran University of Science and Technology, Iran, in 2011. He also received his M.Sc. degree in Information Technology (IT), with a sheer focus on Artificial Intelligence, from Mazandaran University of Science and Technology, Iran, in 2014. He is currently a Ph.D. candidate in artificial intelligence and robotics at University of Tabriz, Iran. His research interests include machine learning and pattern recognition, social network analysis, and data mining.

Mohammad-Reza Feizi-Derakhshi received his B.Sc. degree in Software Engineering from University of Isfahan, Iran. He also received his M.Ss. and Ph.D. degrees in Artificial Intelligence from Iran University of Science and Technology. He is currently an associate professor of Computer Engineering department at University of Tabriz, Tabriz, Iran. His research interests include natural language processing, optimization algorithms, intelligent methods for fault detection, and intelligent databases.

Hadi S. Aghdasi received his B.Sc. degree in computer engineering in 2006 from Sadjad University of Technology, Mashhad, Iran and received his M.Sc. and Ph.D. degrees in computer engineering from Shahid Beheshti University, Tehran, Iran, in 2008 and 2013, respectively. He is currently an assistant professor of Computer Engineering department in University of Tabriz, Tabriz, Iran, since 2013. His research interests include wireless visual sensor networks, humanoid robots in rescue systems and cognitive technology, and evolutionary methods.