An Extended Louvain Method for Community Detection in Attributed Social Networks

Yasser Sadri[1], Saeid Taghavi Afshord[1*], Shahriar Lotfi[2], Vahid Majidnezhad[1]
[1]Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran.
[2]Department of Computer Science, Faculty of Mathematics, Statistics and Computer Science, University of Tabriz, Tabriz, Iran.
Email :yasser.sadri@iaushab.ac.ir; sa.taghavi@iau.ac.ir; shahriar_lotfi@tabrizu.ac.ir,vahidmn@iaushab.ac.ir
Receive Date: 10 September 2022, Revise Date: 6 October 2022, Accept Date:20 November 2022

Abstract

Community detection is a significant way to analyze complex networks. Classical methods usually deal only with the network's structure and ignore content features. During the last decade, most solutions for community detection only consider network topology. Social networks, as complex systems, contain actors with certain social connections. Moreover, most real-world social networks provide additional data about actors, such as age, gender, preferences, etc. However, content-based methods lead to the loss of valuable topology information. This paper describes and clarifies the problems and proposes a fast and deterministic method for discovering communities in social networks to combine structure and semantics. The proposed method has been evaluated through simulation experiments, showing efficient performance in network topology and semantic criteria and achieving proportional performance for community detection.

Keywords: Social Networks Analysis, Complex Networks, Community Detection, Extended Louvain, Semantic.

1. Introduction

Over recent years, the development of online social networks, including millions of users with different characteristics, relationships, and interactions, has made a new epoch in dynamically complex systems. Analyzing these networks is one of this field's most challenging and significant topics. Social network analysis is the mapping and measuring relationships and collaborations between individuals, groups, organizations, websites, and any entity that can process information and knowledge. These networks have been analyzed and evaluated from different aspects. Some studies have considered the propagation of information procedures in the social network and have presented diverse models for this process. Some others have identified influential nodes in the network. Analysis of network changes over time is one of the other aspects of network investigation. Identifying hidden relations and predicting future communications are among the analyzing methods. With the increasing expansion of these networks, new approaches for evaluating the network are presented, and new challenges are raised in this field. Analysis of communication between people and social objects is one of the most important of these methods. One of the widely used, essential, and challenging subjects, which has attracted the attention of many researchers, is the identification of communities.

Discovering communities can be used in various fields, like business, security, and communication. It can also be helpful to identify significant research areas in various scientific fields, including sociology, biology, social sciences, computer science, and so on., and the dissemination of information among scientific communities, etc., and other virtual spaces. Furthermore, the role of community detection has been highlighted by the emergence of companies whose business model is based on social network portals. Commercial companies can provide goods and services to large communities. Therefore, the performance of advertising plans and the marketing industry can improve by determining and classifying suitable groups of users in a particular network. Moreover, it can be used to identify influential groups in society by focusing on the communication and beliefs of individuals. In addition to the above, it can be used in studying the spread of diseases and viruses to control infectious diseases.

A community is a dense mass of nodes having fewer relations with others. Examples include people with similar interests on social media, web pages with related content, and articles with similar topics. Despite the differences in size, behavior, and characteristics of networks, they share common principles and rules. This paper focuses on identifying communities in social networks with attributes. It is a challenging task to find meaningful communities that combine topology and content. To address this, an extended Louvain algorithm (EL), known for its efficiency in analyzing large networks, is proposed. Our approach solves the problems of network topology and node attributes, incorporating structure and semantics in detecting communities. This paper discusses related works in Section 2, outlines significant issues, and essential definitions and arguments in Section 3. Section 4 explains our proposed method, while Section 5 contains analyses of our approach and other methods. Finally, we give concluding remarks in Section 6.

2. Literature Review

There are several proposals for identifying communities in complex networks. Community detection methods are commonly classified into three kinds: classic, semantic, and combinatorial. The traditional algorithms are classic and based on network topology. The algorithms proposed in [1-5] are well-known methods in this area. Most other methods are somehow related to these algorithms. They do not consider semantic features and node attributes. Some of the algorithms, known as semantic or topical, regard only the node attributes without any structural considerations. Many perform on the supposition that entities in social media communicate with one another through-composed text. These approaches depend on various techniques to identify topics in the text, and they categorize similar texts that share common issues. In this context, two commonly used techniques are Latent Dirichlet Allocation (LDA) [6] and Latent Semantic Analysis (LSA) [7]. The algorithms proposed by [8-14] are some of the content-based techniques. Combinatorial methods intend to discover social network communities in a way that considers structure and content social network characteristics together. This section studies and reports the main features of the algorithms considered for comparative analysis in social network areas. Therefore, we focus on combinatorial algorithms. Besides, we illustrate some significant and newly released related techniques. Combinatorial algorithms aim to balance topology and content.

Some of these methods have been presented in [15-18]. Zhao, Feng [15], known as Topic-oriented, first apply k-means clustering to classify social objects into pre-existing subjects. Then, the algorithm of [3] is exerted to appear the communities clustered based on maximizing modularity. Most of the current methods are an extension of this method. A probabilistic generative model considers (CESNA) that communities generate network structure and attributes [19]. Chai, Yu [20] proposed a popularity–productivity stochastic block model with a discriminative-content model PPSB-DC for general structure detection. Reihanian and Minaei-Bidgoli [21] investigate a rating-based way similar to [15], except that Louvain's algorithm is used in the second step.

Table 1 Comparison of different methods.

Method	Partitioning	Deterministic	Overlapping	Topic dependency	Size of Datasets	Datasets Used	Quality measures
Topic-oriented	No	Yes	No	No	Medium	Enron Pol-Blog Cora	Modularity Purity PurQ
PPSB-DC	Yes	No	No	No	Medium	Cora CiteSeer WebKB	NMI Pairwise F-measure Accuracy
CESNA	No	Yes	No	No	Medium	ego-Facebook ego-Twitter Flickr	Evaluation
Rating based	No	Yes	No	No	Small Medium	Movielens Book-Crossing CIAO MovieTweetings	Modularity Purity F-score
MOEA-SA	No	No	No	No	Small Medium	Synthetic Cora Citeseer Political books Political Blogs ego-Facebook	NMI Cumulative NMI Density Entropy
WCMFA	Yes	Yes	No	No	Small	Consult London Gang Montreal Gang	RI ARI NMI
Pourabbasi, Majidnezhad [17]	No	No	No	No	Small Medium	WebKB Cora CiteSeer Politics-UK	NMI Accuracy Jaccard ARI
Wang, Jin [18]	No	No	No	No	Medium	WebKB Cora CiteSeer	NMI Accuracy Jaccard F-score
Akachar et al., (2021)	No	Yes	No	No	Medium Large	Enron Cora DBLP	Modularity NMI F-score
C. He et al. (2022)	No	Yes	No	No	Medium	Synthetic Facebook Computer science Engineering	ONMI F-score
Reihanian, Feizi-Derakhshi [22]	No	No	Yes	No	Small Medium	Pol- Books Football Adj-noun Epinion Politics-UK UK-Faculty	Modularity Purity F-score
2PCD	No	Yes	No	Yes	Small Medium Large	Pol- Books football Adj-noun Lazega Pol-Blogs Cora Epinion Blog Catalog CIAO Flickr DBLP	Modularity Semantic coherency F-score

The method of [16] first divides the network into different clusters based on topics; then, communities are re-identified using the Louvain method. Wang and Jin [18] offer an optimization function to merge the topology and node attributes. For this purpose, two types of semantic and topological communities are defined separately. Thus, two matrices containing crisp values represent node membership in each semantic or structural community used. First, the algorithm considers topology as fixed and optimizes the content feature, then assumes that the semantic is constant and optimizes the structure criterion.

A multiobjective evolutionary algorithm (MOEA-SA) in [23] is proposed to maximize modularity and attribute similarity. Co-association matrix-based multi-layer fusion for community detection (WCMA) in attributed networks is proposed in [24]. Pourabbasi and Majidnezhad [17] propose a single-chromosome evolutionary algorithm to identify node connectivity and similar traits to identify communities. In [25], a neural network-based semi-supervised method is designed to discover overlapping communities. Reihanian, Feizi-Derakhshi [22], and Reihanian, Feizi-Derakhshi [26] propose evolutionary approaches to find, respectively, disjoint and overlapping communities containing nodes with similar attributes and structural connections. In [27], we suggest a two-phase community discovery (2PCD) method to account for the impact of node attribute relationships on decision-making. The basic idea is that linked functions increase the similarity of their related nodes. Therefore, fuzzy structure and content combiner were introduced. It incorporates networking features in an advanced web format. Finally, a community detection process is performed on this enriched network. A detailed comparison of different methods is illustrated in Table 1.

3. Problems Statement and Definitions

This section discusses some background information about social network analysis, especially community detection. We present the formal definitions and basic concepts. Then briefly summarize the problems and challenges of discovering communities.

3.1. Basic concepts

A network is a particular graph that outlines relationships in real-world systems.

Definition 1 Network is a dyadic $G = (V, E)$. The set of network nodes V denotes users, actors, and active members. Each node can be associated with several other nodes. E represents the edges, links, connections, and interactions between the network nodes. Connections are made through different networking activities such as blogging, tagging, movie uploading, and so on [28].

Definition 2 Weighted Network is represented as $G = (V, E, W)$. W denotes the corresponding weights of E in terms of the intensity or capacity of the connections.

Definition 3 Attributed Network is a triangle $G = (V, E, A)$. A is the set of attributes, features, and topics of nodes.

Definition 4 Subnetwork g ⊆ G is a part of the network that maintains the properties of the original network.

Community, in Traditional Sense, is a significant substructure in which nodes share dense connections compared to external ones.

Community, in Semantic Definitions, reflects semantic concepts, interest topics, and attributes shared by individuals.

Community, in Real-World, refers to a group of objects with high interaction, similar semantic characteristics, and interest in everyday issues.

Definition 5 Community, in Social Networks, $C = \{c_1, c_2, \dots, c_q \,|\, c_i \subseteq \mathrm{G}\}$ is a subnetwork where nodes share dense relations and reflect similar attributes.

3.2. Problem's statement and challenges

Problem I Structure or Semantic: Finding meaningful communities in social networks has become an important and challenging issue. The first issue in this field is whether the community detection strategy should be based on network topology or semantics, and/or both. The next challenge is discovering communities in a way that fairly integrates structure and content. This integration is a complex and complicated problem for community detection in general [29, 30]. On the other hand, in real-life, semantic features and topics are related. These dependencies have been ignored in most of the existing approaches.

Problem II Disjoint or Overlapping: The next problem that complicates the scrutiny is whether combinatorial communities should be overlapping or non-overlapping. Unlike disjoint methods, a vertex can belong to more than a community using overlapping algorithms [22].

Problem III Complexity and Scalability: Nowadays, online social networks like Facebook, Twitter, LinkedIn, and so on contain a wide range of information, including millions of users with various attributes and relationships between them. In this context, two issues of time and space complexity are raised. Presented techniques must have lower intricacy and greater flexibility. Distributed computing and parallelism can be another solution for computational complexity.

Problem IV Number of Communities: Knowing the number of communities in advance is a problem; despite some solutions, it has not been fully resolved. Some solutions require a predefined number of communities. Despite the lower computational complexity, these solutions are practically inefficient for identifying communities in social networks with dynamic changes in topology and attributes.

Problem V Weighted or Unweighted Network: Some real-world networks have weighted edges, nodes, and even signed relationships. The problem is that these characteristics must be treated differently and appropriately.

Problem VI Deterministic or non-Deterministic: Unlike non-deterministic methods, such as evolutionary methods, the output is the same every time the algorithm is executed in deterministic methods.

4. Proposed Method

In this section, we outline the proposed mechanism for identifying communities in social networks with various attributes and topological requirements. We extend the Louvain algorithm proposed in [2] due to its high modularity value in a reasonable time. The network we generate to identify communities also considers semantic relationships and structural relationships; i.e., the proposed algorithm will increase modularity by putting nodes with high topological and topical relations in a community. Figure 1 illustrates the flowchart of the proposed extended Louvain method.

Step 1 receives two matrices as input: Structural Adjective (SA) of nodes (V) and Attribute Relation (AR), which are represented by equations 1 and 2, respectively. AR displays the relationships between V and attributes A.

$$SA = \begin{array}{c c} & \begin{array}{ccc} V_1 & \cdots & V_n \end{array} \\ \begin{array}{c} V_1 \\ \vdots \\ V_n \end{array} & \begin{bmatrix} sa_{1,1} & \cdots & sa_{1,n} \\ \vdots & \ddots & \vdots \\ sa_{n,1} & \cdots & sa_{n,n} \end{bmatrix} \end{array} \quad (1)$$

$$AR = \begin{array}{c c} & \begin{array}{ccc} A_1 & \cdots & A_k \end{array} \\ \begin{array}{c} V_1 \\ \vdots \\ V_n \end{array} & \begin{bmatrix} ra_{1,1} & \cdots & a_{1,k} \\ \vdots & \ddots & \vdots \\ ra_{n,1} & \cdots & ra_{n,k} \end{bmatrix} \end{array} \quad (2)$$

Step 2 calculates Topical Weights (TW$(i\ ,j)$) between node pairs (vi, vj). TW$(i\ ,j)$ is based on equation 3 and determined by the ratio of shared attributes between vi and vj to the maximum number of attributes that one of vi or vj possesses.

$$\mathrm{TW}(i\ ,j) = \frac{\sum_{a=1}^{l}(AR(i,a).AR(j,a))}{\sum_{a=1}^{l}(AR(i,a) + AR(j,a) - (AR(i,a).AR(j,a)))} \quad (3)$$

Step 3 combines SA and TW in the Topological and Topical matrix (TT) according to equation 4 to support structure and semantics simultaneously.

$$TT(i,j) = MAX\big(\mathrm{TW}(i\ ,j), \mathrm{SA}(i\ ,j)\big) \quad (4)$$

Step 4 places each isolated node vi in separate communities C_{vi}.

Step 5 calculates the modularity gain ΔQ using equation 5 for each node vi. It is done to move vi into the neighboring community C_{vj} that contain node vj, which has at least one edge between them. Σ_{in} and Σ_{tot} are the sums of the weights of the links inside and incident to C_{vi} respectively, while $k_{vj,in}$is the sum of the weights of the links from vj to nodes in C_{vi}.

$$\Delta\mathrm{Q} = \left[\frac{\Sigma_{in} + 2k_{vj,in}}{2m} - \left(\frac{\Sigma_{tot}+k_{vj}}{2m}\right)^2\right] - \left[\frac{\Sigma_{in}}{2m} - \left(\frac{\Sigma_{tot}}{2m}\right)^2 - \left(\frac{k_{vj}}{2m}\right)^2\right] \quad (5)$$

Step 6 finds the merge with the maximum modularity Q from Eq. 6. The total number of output network edges is denoted by $m = \frac{1}{2}\sum_{vivj} TT_{vivj}$, k_{vi} represents the degree of node vi. If nodes vi and vj belong to the same community, $\delta(C_{vi}, C_{vj})$ equals one; otherwise, it is zero [15].

$$Q = \frac{1}{2m}\sum_{vivj}\left(TT_{vivj} - \frac{k_{vi}k_{vj}}{2m}\right)\delta(C_{vi}, C_{vj}) \quad (6)$$

Step 7 merges node vi with the target community C_{vj}. If no increase is possible, vi remains in its original community.

Step 8 merges all nodes in the same communities as a node and repeats step 4.

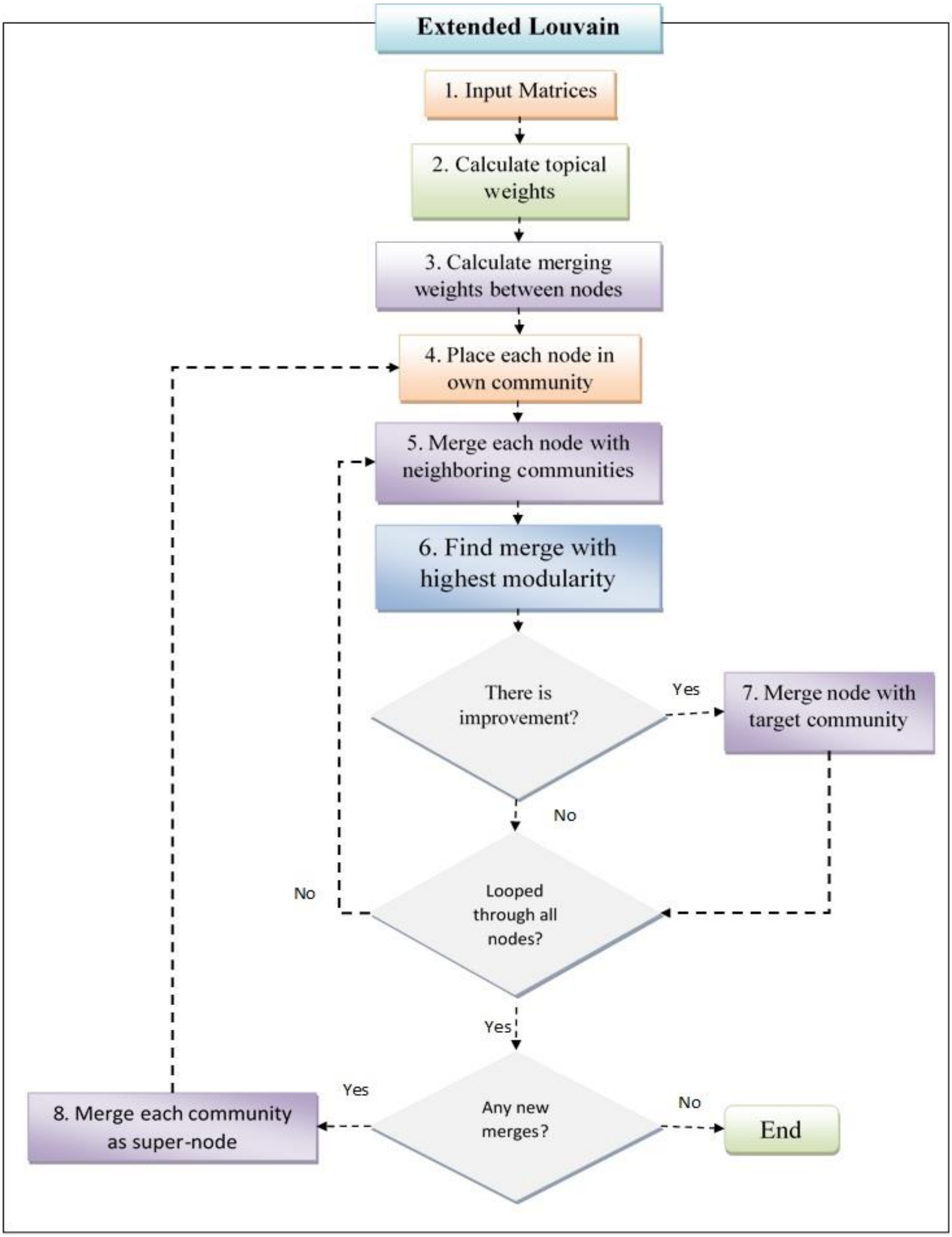


Fig. 1 The flowchart of the proposed extended Louvain to topology and content

5. Experiments and Evaluations

This section comprises an assessment of the outcomes of the introduced method. Our primary objective is to compare the algorithm's efficiency with other approaches while also analyzing the complexity of the algorithms.

5.1. Experiments

Moving on to the experiments and simulation results, we implemented the algorithms in MATLAB (2016a) and applied them to various datasets. Subsequently, we compared the proposed EL with other relevant algorithms. To conduct these experiments, we utilized a computer equipped with an Intel Core i3 2.20 GHz CPU and 8 GB RAM DDR4.

5.1.1. Measurement metrics

The quality of algorithms analyzes using measurement criteria. Modularity is the most popular quality criterion for structural measurements. Purity is one of the most used criteria to evaluate attribute homogeneity as a semantic measurement. Measures based on F-measure can be used for the quality of combinatorial algorithms.

- **Modularity (Q)** measures the ratio of the number of edges within a community to the number of edges between them. It was first introduced by Girvan and Newman [5] to calculate the power of dividing a network into different communities. However, its application is limited to an undirected and unweighted graph in a social network. It also fails in the calculation of the modularity of overlapping communities. In [31] a measurement is defined for directed graphs. Blondel, Guillaume [2] enhance modularity for weighted networks. Some modularity measures are proposed for overlapping and fuzzy communities in [32-35]. More details can be found in [36].

$$Q = \frac{1}{2m}\sum_{c\in C}\sum_{uv}\left(SA_{uv} - \frac{k_u k_v}{2m}\right) a_{u,c} a_{v,c} \quad (7)$$

- **Purity** measures the degree to which communities contain users with the same interests and class [37]. The greater purity, the nodes of communities are similar in attributes.

- $$Purity = \frac{1}{N_{cm}}\sum_{i=1}^{N_{cm}} \max_{1\le j\le K}\left\{\frac{n_{ij}}{n_i}\right\} \quad (8)$$

- **F-score (F-measure)** is a measure of accuracy. The F-score is utilized for measuring classification efficiency [38]. Earlier works concentrated mainly on the F_1 score, but with the extension of large-scale engines, performance goals altered to place more emphasis [39], so F_β is used in wide applications. PurQ is a F_β presented in [15] for evaluating the performance of algorithms that must consider both aspects of the topology and content.

5.1.2. Evaluation networks

This subsection briefly explains some datasets that are popular in the field of human social interaction, social blogging, co-purchasing, co-rating, lexical, and photo-sharing is used to evaluate algorithms.

Political Books [40] is a co-purchasing network comprising 105 nodes showing US policy books sold by Amazon Online Bookstore. Edges (441 edges) are the number of simultaneous purchases of similar books by the same buyers. The nodes have three attributes, "l", "n" or "c" indicating they are "liberal", "neutral" or "conservative" which were assigned by Mark Newman based on comments, descriptions, and reviews posted to the Amazon site.

American college football [5] is a human social interaction network taken from American football games between division IA colleges fall of 2000. There are 115 nodes representing teams and 613 edges denoting games between pair teams. Nodes attributes convey conferences they attended and labeled as follows: "Atlantic Coast" =0, "Great East" =1, "Big Ten" =2, "Twelve Great" =3, "American Conference" =4, "Independent" =5, "Middle America" =6, "Western Mountain" =7, "Ten Calm" =8, "Southeast" =9, "Sun Belt" =10, "Western Stadium" =11.

Lazega [41] comes from a network investigation on 71 lawyers in a US New England corporate Law firm containing their coworker network, advice network, friendship network, and indirect control networks. Some node attributes, such as status (partner or associate), gender, practice (litigation or corporate), and so on, are available [1]. For example, advice-practice contains 717 links between 71 nodes, a friend-status network with 399 friendship relations, and Work-practice involves 378 co-working connections. In this paper, we use a friend-status network.

Table 2. Dataset's information

Network	Node	Edge	Attributes
Pol- Books	105	441	3
football	115	613	12
Lazega	71	399	3
Flickr	100267	3781947	195
CIAO	7375	264225	28

Flickr (www.flickr.com) is a social photo-sharing network built by featuring tag images from Flickr. The Flickr site provides users to share their content, upload tags and subscribe to different interest groups [42]. This dataset contains 100267 nodes and 195 interest groups as the Node attributes [43].

CIAO [40] is a co-rating network where users share their views on a product by evaluating and recording their opinions on the CIAO site. These products are categorized into different classes, such as fashion, office equipment, etc. This dataset retains 284,086 reviews supplied by 7,375 users to 105,114 products (out of 28 classifications).

5.1.3. Simulation results

We evaluate and compare the proposed EL with the Louvain [2] as a structural algorithm, semantic clustering [9], and topic-oriented method [15] as a combinatorial method on mentioned datasets.

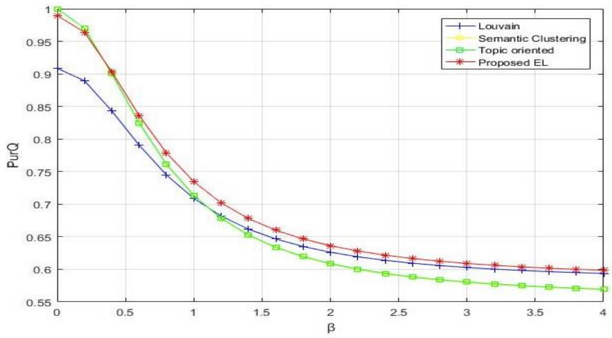
In Table 3, we can see the modularity of identified communities in a real-world data set. Based on this table, it is evident that the structural and our methods perform better in terms of modularity compared to the semantic and even the combined algorithms. The greedy structural algorithms prioritize modularity over the attributes of nodes when detecting communities. Our proposed algorithm performs better than other algorithms in most cases because it considers the network topology, leading to high modularity. The results are sometimes weak with the semantic method, possibly due to nodes with common features being more connected.

Table 3. Modularity results from the comparison of the set of networks

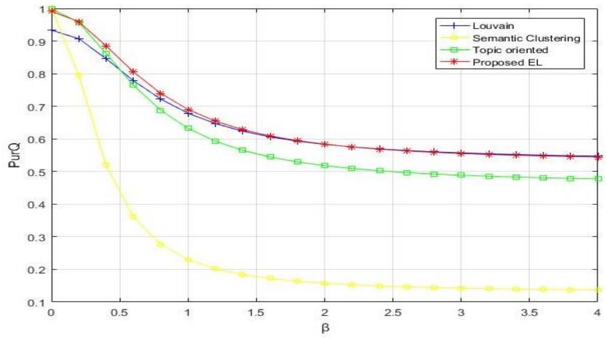
Network	Louvain	Semantic Clustering	Topic Oriented	Proposed EL
football	0.5811	0.5540	0.5546	0.5841
CIAO	0.5333	0.1299	0.4623	0.5294
Pol- Books	0.4727	0.4149	0.2969	0.5123
Flickr	0.4003	0.2918	0.2863	0.4251
Lazega	0.3397	0.2553	0.2378	0.3394

Table 4. Purity results of the comparison on the set of networks

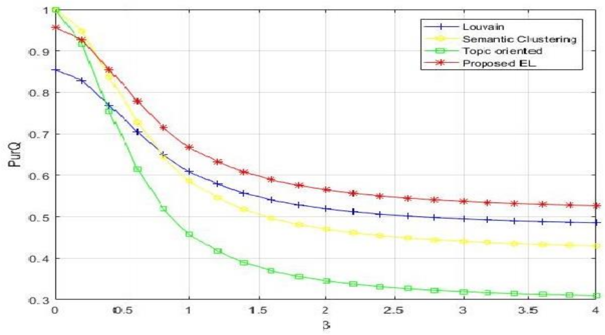
Network	Louvain	Semantic Clustering	Topic Oriented	Proposed EL
football	0.9087	1.0000	1.0000	0.9897
CIAO	0.9334	1.0000	1.0000	0.9915
Pol-Books	0.8547	1.0000	1.0000	0.9571
Flickr	0.8300	1.0000	1.0000	0.9982
Lazega	0.5070	1.0000	1.0000	0.9750



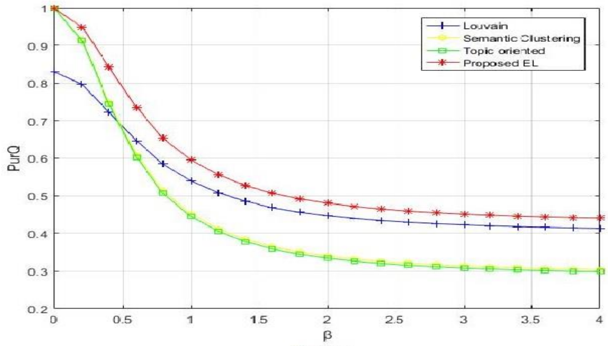
a) Football



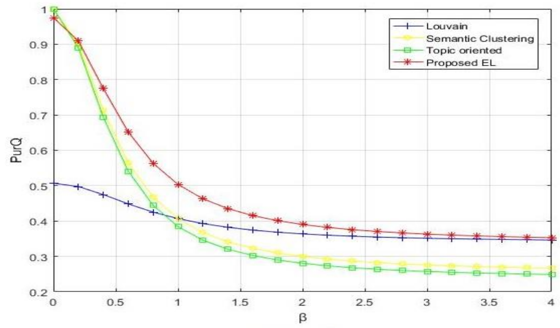
b) CIAO



c) Pol-Books



d) Flickr



e) Lazega

Fig. 2 Performance analysis: The $PurQ_{\beta}$ of returned communities from different algorithms for different βs

Table 4 shows the purity of different algorithms. The combinatorial algorithms and the method presented in this article, which consider the attributes of nodes in identifying communities, perform better than methods with a structure that has a purity close to one. The semantic algorithm has the highest purity with one value, but it does not consider network structure and modularity. Combinatorial algorithms are slightly better than the proposed EL algorithm in terms of purity because they only cluster based on the attributes of nodes in the initial step.

Based on the results of modularity and purity, we calculate PurQβ to determine which algorithm performed better. The value of β determines the weight of the similarity value of the attributes of the nodes and the value of structural links. When β equals 1, the purity of node attributes and the modularity of structural connections are equally significant. When β is between 0 and 1, it indicates that node attribute values have a higher weight. And when β is greater than 1, it signifies that topological relations have a higher weight. Figure 2 displays the average PurQβ for each algorithm across all datasets at different β. The proposed algorithm performs better than other algorithms because it considers the network topology requirements and the nodes' attributes in the community detection procedure. The topic-oriented algorithm is unsuitable due to inefficient modularity, resulting in a lower PurQβ.

Conclusions

There are three main types of methods for network analysis: structural, semantic, and merging solutions. However, most studies focus solely on the network structure and ignore the attributes of individual nodes. Semantic methods can also be problematic as they may lose valuable information and lead to poor modularity. Many merging solutions also suffer from poor efficiency and high complexity.

Our approach extends the Louvain algorithm, known for its speed and efficiency in analyzing large networks, to combine structural and semantic elements. We collect the topology and content data in a structural format before extracting the communities. Our method has been shown to capture both the topology and content of the network, as demonstrated by the excellent PurQ results. Our algorithm is practical for analyzing social networks and content features, and its certainty is an additional advantage as a merging method.

References

[1] Zhang, H., et al., Limited Random Walk Algorithm for Big Graph Data Clustering. Journal of Big Data, 2016. **3**.

[2] Blondel, V., et al., Fast Unfolding of Communities in Large Networks. Journal of Statistical Mechanics Theory and Experiment, 2008. **2008**.

[3] Clauset, A., M. Newman, and C. Moore, Finding community structure in very large networks. Physical review. E, Statistical, nonlinear, and soft matter physics, 2005. **70**: p. 066111.

[4] Newman, M. and M. Girvan, Finding and Evaluating Community Structure in Networks. Physical review. E, Statistical, nonlinear, and soft matter physics, 2004. **69**: p. 026113.

[5] Girvan, M. and M. Newman, Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. Proc. Natl Acad. Sci. USA 99, 7821-7826. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**: p. 7821-6.

[6] Blei, D., A. Ng, and M. Jordan, Latent Dirichlet Allocation. Vol. 3. 2001. 601-608.

[7] Hastings, P., Latent Semantic Analysis. 2004.

[8] Abdelbary, H.A. and A. El-Korany, Semantic Topics Modeling Approach for Community Detection. Social networks, 2013. **81**.

[9] Xia, Z. and Z. Bu, Community detection based on a semantic network. Knowledge-Based Systems, 2012. **26**(Complete): p. 30-39.
[10] Ereteo, G., F. Gandon, and M. Buffa, SemTagP: Semantic Community Detection in Folksonomies. 2011. 324-331.
[11] Dai, T. et al., Explore semantic topics and author communities for citation recommendation in bipartite bibliographic network. Journal of Ambient Intelligence and Humanized Computing, 2018. **9**(4): p. 957-975.
[12] Yu, X., J. Yang, and Z.-Q. Xie, A semantic overlapping community detection algorithm based on field sampling. Expert Syst. Appl., 2015. **42**(1): p. 366–375.
[13] Xin, Y., et al., An overlapping semantic community detection algorithm base on the ARTs multiple sampling models. Expert Systems with Applications, 2015. **42**(7): p. 3420-3432.
[14] Han, X., D. Chen, and H. Yang, A Semantic Community Detection Algorithm Based on Quantizing Progress. Complexity, 2019. **2019**: p. 1-13.
[15] Zhao, Z., et al., Topic-oriented community detection through social objects and link analysis in social networks. Knowledge-Based Systems, 2012. **26**: p. 164-173.
[16] Akachar, E., B. Ouhbi, and B. Frikh, ACSIMCD: A 2-phase framework for detecting meaningful communities in dynamic social networks. Future Generation Computer Systems, 2021. **125**: p. 399-420.
[17] Pourabbasi, E., et al., A new single-chromosome evolutionary algorithm for community detection in complex networks by combining content and structural information. Expert Systems with Applications, 2021. **186**: p. 115854.
[18] Wang, X., et al., Semantic Community Identification in Large Attribute Networks. Proceedings of the AAAI Conference on Artificial Intelligence, 2016. **30**.
[19] Yang, J., J. McAuley, and J. Leskovec. Community detection in networks with node attributes. In 2013 IEEE 13th international conference on data mining. 2013. IEEE.
[20] Chai, B.-f., et al., Combining a popularity-productivity stochastic block model with a discriminative-content model for general structure detection. Physical Review E, 2013. **88**(1): p. 012807.
[21] Reihanian, A., B. Minaei-Bidgoli, and H. Alizadeh, Topic-oriented community detection of rating-based social networks. Journal of King Saud University - Computer and Information Sciences, 2016. **28**(3): p. 303-310.
[22] Reihanian, A., M.-R. Feizi-Derakhshi, and H.S. Aghdasi, An enhanced multiobjective biogeography-based optimization for overlapping community detection in social networks with node attributes. Information Sciences, 2023. **622**: p. 903-929.
[23] Li, Z., J. Liu, and K. Wu, A multiobjective evolutionary algorithm based on structural and attribute similarities for community detection in attributed networks. IEEE transactions on cybernetics, 2017. **48**(7): p. 1963-1976.
[24] Luo, S. et al., Co-association matrix-based multi-layer fusion for community detection in attributed networks. Entropy, 2019. **21**(1): p. 95.
[25] He, C., et al., Semi-supervised overlapping community detection in attributed graph with graph convolutional autoencoder. Information Sciences, 2022. **608**: p. 1464-1479.
[26] Reihanian, A., M.-R. Feizi-Derakhshi, and H.S. Aghdasi, Community detection in social networks with node attributes based on multiobjective biogeography based optimization. Engineering Applications of Artificial Intelligence, 2017. **62**: p. 51-67.
[27] Sadri, Y. et al., Handling topic dependencies alongside topology interactions using fuzzy inferences for discovering communities in social networks. Expert Systems with Applications, 2022. **208**: p. 118188.
[28] Kanavos, A., et al., Emotional community detection in social networks. Computers & Electrical Engineering, 2018. **65**: p. 449-460.
[29] Kalanat, N., E. Khanjari, and A. Khanshan, Extracting Actionable Knowledge From Social Networks Using Structural Features. IEEE Access, 2020. **8**: p. 59637-59647.
[30] Khan, A., et al., Compact group discovery in attributed graphs and social networks. Information Processing & Management, 2019. **57**: p. 102054.
[31] Leicht, E.A. and M.E. Newman, Community structure in directed networks. Physical review letters, 2008. **100**(11): p. 118703.
[32] Nepusz, T., et al., Fuzzy communities and the concept of bridgeness in complex networks.

Physical review. E, Statistical, nonlinear, and soft matter physics, 2008. **77**: p. 016107.
[33] Chen, D. et al., Detecting overlapping communities of weighted networks via a local algorithm. Physica A: Statistical Mechanics and its Applications, 2010. **389**(19): p. 4177-4187.
[34] Shen, H.-W., X.-Q. Cheng, and J.-F. Guo, Quantifying and identifying the overlapping community structure in networks. Journal of Statistical Mechanics: Theory and Experiment, 2009. **2009**(07): p. P07042.
[35] Shen, H. et al., Detect overlapping and hierarchical community structure in networks. Physica A: Statistical Mechanics and its Applications, 2009. **388**(8): p. 1706-1712.
[36] Devi, J.C. and E. Poovammal, An Analysis of Overlapping Community Detection Algorithms in Social Networks. Procedia Computer Science, 2016. **89**: p. 349-358.
[37] Christopher, D.M., R. Prabhakar, and S. Hinrich, Introduction to information retrieval. An Introduction To Information Retrieval, 2008. **151**(177): p. 5.
[38] Beitzel, S., On understanding and classifying web queries (pp. 3216-3216). Chicago: Illinois Institute of Technology, 2006.
[39] Li, X. Ye-YiWang, and Alex Acero. 2008. Learning query intent from regularized click graphs. In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM.
[40] Krebs, V., Political books network. 2004.
[41] Lazega, E., The collegial phenomenon: The social mechanisms of cooperation among peers in a corporate law partnership. 2001: Oxford University Press on Demand.
[42] Wang, X., et al., Learning with multi-resolution overlapping communities. Knowledge and information systems, 2013. **36**(2): p. 517-535.
[43] Tang, L. and H. Liu. Relational learning via latent social dimensions. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009.