Filomat 32:5 (2018), 1843–1851
https://doi.org/10.2298/FIL1805843Z

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Topology and Semantic based Topic Dependency Structure Discovery

Anping Zhao[a], Suresh Manandhar[b], Lei Yu[a]

[a]*College of Computer and Information Science,Chongqing Normal University, China*
[b]*Department of Computer Science,University of York,UK*

Abstract. As an important enabler in achieving the maximum potential of text data analysis, topic relationship dependency structure discovery is employed to effectively support the advanced text data analysis intelligent application. The proposed framework combines an analysis approach of complex network and the Latent Dirichlet Allocation (LDA) model for topic relationship network discovery. The approach is to identify topics of the text data based on the LDA and to discover the graphical semantic structure of the intrinsic association dependency between topics. This not only exploits the association dependency between topics but also leverages a series of upper-level semantic topics covered by the text data. The results of evaluation and experimental analysis show that the proposed method is effective and feasible. The results of the proposed work imply that the topics and relationships between them can be detected by this approach. It also provides complete semantic interpretation.

1. Introduction

Topic modeling provides us a way to represent a large volume of unstructured texts because topics reveal the correlations between words. It cannot deal with other unstructured text information such as relationships between topics. Real applications often need to explain these large volumes of unstructured text information using both topics and their dependency relationships simultaneously. However, discovering such dependency relationships between topics, with the goal to automatically discover latent user requirements and to facilitate an enterprises decision, is a major bottleneck. This fact calls for the development of novel model to topic dependency structure discovery for maximizing the utility of observed textual information. Toward these challenges, topics need to be organized and semantically described in an effective and efficient manner. We develop a Topic Dependency Network Model (TDNM) that helps users to analyse how the correlated topics are linked by their semantically implied relationships. The TDNM explicitly ties the content of the topics to the connections between them for effectively implementing automatic topic search and discovery. It enables us to set only part of the intrinsic topic semantic association to depict the relationships distribution of topics and concentrates on navigation and collaboration for exploring related latent topics.

This paper focuses on how to model TDNM for discovering dependency relationships among topics by combining LDA and network coherence analysis of complex network. Our interest lies in a particular problem, one which the more powerful models are developed to solve, where the following various information is involved with texts:When

2010 *Mathematics Subject Classification*. 90B10.
Keywords. Topic; Dependency; LDA; Network coherence.
Received: 24 October 2017; Accepted: 30 January 2018
Communicated by Hari M. Srivastava
This research was supported by the Fundamental Research Funds for the Central University [No.2017CDJSK01XK22], the Chongqing Research Program of Basic Research and Frontier Technology [No.cstc2017jcyjAX0106], the Pre-research Project of Chongqing Normal University [No.16XYY20], the Scientific and Technological Research Program of Chongqing Municipal Education Commission [No.KJ1600306].
Email addresses: apzhao@cqnu.edu.cn (Anping Zhao), suresh@york.ac.uk (Suresh Manandhar), leiy@cqnu.edu.cn (Lei Yu)

topics and relationships are involved with texts, how do we develop a model of text data that accounts for both topics and their relationships and identifies the influence of links between words on topic relationship? Given the topics and their relationship links, how does our model provide a predictive relationships distribution of topics?

This paper defines this as a research problem and makes the following contributions:

1. We formulate a problem that aims to discover a topic dependency relationship network from text data and incorporate semantic communities and complex network coherence measures to deal with the problem.
2. We present TDNM as an analytic framework to tackle this problem, which includes both topics identifying and dynamic topic relationships discovery.
3. We present the TDNM-based predictive mechanism of topic relationships and conduct preliminary experiments and performance studies for evaluating TDNM.

The remaining of the paper is organized as follows: We review related work in section 2. The problem formulation of our work is presented in section 3. Section 4 describes the approach to model TDNM based on LDA and complex network analysis. We give the effectiveness of the proposed approach by the results of evaluation in section 5. Finally, we summarize our conclusions of the work and give the future research directions in section 6.

2. Related work

With the introduction of complex network theories, it is plausible to conduct large-scale empirical study into language networks[1][2]. The recent research works gave the semantic representation about human language by complex network model, which employed the dynamic complex network viewpoint to analyze natural human language[3][4]. The word co-occurrence analysis method was used as a real application example of language network in the scientific fields[5]. Researchers have been investigating the interplay between topic models and networks. The work related to the topical model for relationship and community detection in online social networks is presented in [6]. X. Wang et al. presented a directed graph model to cluster topical entities with relations among them[7]. McCallum et al. proposed the Author-Recipient-Topic (ART) to capture the dependencies between topics of conversation[8]. Dietz et al. created a model to analyze the citation networks, where the topical innovation and inheritance are used to generate documents[9]. Nallapati et al. similarly presented a model to find topical association relationships in documents by cititation[10]. Wang et al. presented a model to deal with situations where semantic links exist among the no similarity association documents[11]. For detecting entity-topic associations between the online news, Balog K et al. presented a system that helps to discover and analyze the relationship between entities and topics[12]. Z. Yin et al. incorporate community discovery into topic coherence analysis in undirected text-associated graphs[13].

There has been little work on topic topology and link analysis relating to topic dependency relationships. The Author-Topic Model was extended to infer social networks from meeting transcripts[14]. The topic maps also were proposed to structure the concept space[15]. Categorization unsupervised methods are used to build topic ontology for discovering similar topics by latent semantic indexing (LSI)[16]. Yan Liu et al. developed a hierarchical Bayesian approach Topic-Link LDA[17]. Yookyung Jo et al. designed the approach to capture the rich topology of topic evolution inherent in the corpus[18]. A topical community discovery approach was proposed in[19]. Recently, David I. Inouye et al. introduced a new topic model which can model dependencies between words as opposed to previous independent topic models[20]. Elijah Meeks combined networks with topic models to show how topics or documents associate to one another within the context of the digital humanities[21]. Qiaozhu Mei et al. defined the problem of topic modeling with network structure[22] to discover topical communities. The relational topic model (RTM) assumes that the distance between topic proportions of documents determined the links between them[23]. The Correlated Topic Model (CTM) is based on LDA with modeling of correlations between topics[24]. James Foulds et al. introduce latent topic networks in the social sciences[25]. S. Das presented an approach to topical relationship detection by unstructured equivocal sources[26].

Our work differs from them mainly on three points. First, it aims to discover not only topics, but also topic relationship network information. Second, we focus on tackling with the dynamic topic relationship changes, which are not addressed in those other studies. Finally, the ultimate goal of our work is to discover and predict the dynamic topic dependency relationship structure, rather than the relationship between documents or any other objects.

3. Problem statement and preliminaries

3.1. Problem statement

Topic relationship discovery is to find relational dependency links between topics from a group of discovered topics. Graph structure is an important characteristic of topic relationship dependency, in which the nodes represent topics and the edges represent associated relationships between topics. The nodes include a list of words which represent topics. The inherent semantic association dependencies between topics can be discovered by analysing the network coherence of a word co-occurrence network,to which a topic refers.

It's a great way to describe relationship by network models. In our problem, we focus on creating a topic graph model TDNM from a set of topics by extending LDA. Semantic and topological coherence are widely adopted to evaluate the relatedness between the communities[27]. With the help of semantic and topological coherence analysis on the word co-occurrence network of text data, we can use the semantic communities determined by the topics to pinpoint those of them that are most influential to each other. Based on word co-occurrence network of text data, the TDNM ties the content of the topics with the connections between them and embeds this data in a latent space that explains both semantic content of the topics and how they are connected.

In the TDNM, the process begins with learning topics by LDA. Then the topic relationship is learned from network coherence analysis process. The critical word correlations between semantic communities of a word co-occurrence network are extracted to encode the topic dependency. The semantic links among topics are then modeled as binary variables, one for each pair of topics. These binary variables are allocated according to a distribution that is based on the dependent semantic links among words, which represent every topic in a word association complex network. Because of this dependence, the content of the topics is statistically connected to the link structure between them. Thus, topic relationship of TDNM depends both on the content of the topic as well as the pattern of its links. In this paper, We focus on jointly modeling the topics from text data and their dependency relationship in the unified model.

3.2. Preliminaries

Word co-occurrence networks are one of the most common linguistic networks. An undirected graph $G = (V, E)$ can be used to represent **a Word co-occurrence network**, where V is the set of nodes representing the word forms in the language data, E is the set of edges that represent adjacency relations of the word forms. The two nodes $u, v \in V$ are linked by an edge $e \in E$ if the two corresponding word forms are adjacent within at least one document.

A semantic community c ($c \in C$) is a group of words in the word co-occurrence network with more associations and common topics within the group than between groups. Where C is the community set. A network is said to have community structure in which the nodes of the network are joined together in tightly knit groups, between which there are only looser connections.

Network coherence is introduced to analyze the interconnected networks where the interconnections couple the nodes based on a discrete time evolution process.

4. Topic dependency network modeling

4.1. Definition of the model

The Topic Dependency Network Model (TDNM) is a hierarchical model of links and topic attributes, in which the nodes are topics, and the edges represent the existence of relationship dependency between the topics. The topics are $\beta_{1:K}$,where each β_k is a distribution over the vocabulary. A $K-$dimensional Dirichlet parameter α, the topic proportions for the dth document are θ_d , where $\theta_{d,k}$ is the topic proportion for topic k in document d , each document is associated with a distribution θ over topics, chosen from a Dirichlet(α). The topic assignments for the dth document are z_d , where $z_{d,n}$ is the topic assignment for the nth word in document d. The observed words for document d are w_d , where $w_{d,n}$ is the nth word in document d , which is an element of the fixed vocabulary. We denote the topic dependency relationship network as G by the conditional distribution of a link given topic relatedness, where $L_{i,j}$ is the conditional distribution of a link between the ith and jth topic. Based on the basic definitions, we describe the data generation process for TDNM as follows:

$$\rho_{i,j} \sim Binomial(\rho) \tag{1}$$
$$\theta_d|\alpha \sim Dirichlet(\alpha) \tag{2}$$
$$z_{d,n}|\theta_d \sim Multi(\theta_d) \tag{3}$$
$$w_{d,n}|z_{d,n}, \beta_{1:K} \sim Multi(\beta_{z_{d,n}}) \tag{4}$$
$$L_{i,j}|\beta_i, \beta_j, \theta_i, \theta_j \sim Bernoulli(\sigma(\rho_{i,j})) \tag{5}$$

where σ is sigmoid function which have been explored in [23] to capture the main trend of the observations, $\sigma(x) = 1/(1+\exp(-x))$ and $\rho_{i,j}$ is a function of the topic relatedness between topic β_i and β_j,which follows a binomial distribution with parameter ρ. The link formation of G associates the existence of a link with a binomial distribution, whether a link exists between two topics follows a binomial distribution ρ parameterized by the relatedness between topics. And the function σ is to provide binary probabilities for links between topics.

With this notation, the joint distribution of the hidden and observed variables for TDNM as follows:

$$p(G, \beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = p(G|\beta_{1:K}) \times \prod_i^K p(\beta_i) \prod_d^D p(\theta_d) \left(\prod_n^N p(z_{d,n}|\theta_d) p(w_{d,n}|\beta_{1:K}, z_{d,n}) \right) \tag{6}$$

where the conditional distribution of a topic dependency relationship network G is

$$p(G|\beta_k) = \prod_i^K \prod_{j \neq i}^K (\sigma(\rho_{i,j}))^{L_{i,j}} (1 - \sigma(\rho_{i,j}))^{1-L_{i,j}} \tag{7}$$

K is the number of topics, D is the number of documents and N is the number of words.

4.2. Relatedness measurement for topic relationship

By creating this model, we aim to quantify the topic relatedness to form a link between the topics. Every semantic community is naturally determined by the topic over a word co-occurrence network. The question here is how related a set of communities could be if the link between them is based on shared common words and links between words. Within the scope of our work, the topic relatedness problem is transformed to a graph-based network community relatedness problem. We choose to use the network coherence for graph-based network community relatedness measurement.

Topological relatedness. In a word co-occurrence network, the link between words signifies both a semantic relationship and network topology. we employ the link-based Jaccard coefficient to calculate topic relatedness between semantic communities by the links found within their corresponding communities. It is based on the observation that the closer the two communities, the denser the links between them are. The link-based Jaccard coefficient between the community c_i and c_j is defined as:

$$Coef_{link}(c_i, c_j) = \frac{2 \times N_{c_i,c_j}}{N_{c_i} + N_{c_j}} \tag{8}$$

where N_{c_i}, N_{c_j} represent the total number of links of the nodes in community c_i and c_j which is determined by the topic β_i and β_j respectively. N_{c_i,c_j} is the number of intercommunity links between c_i and c_j .

Semantic relatedness. The assumption is that the closer the two communities are, the more common terms two communities share. Based on this, the term-based Jaccard coefficient between the community c_i and c_j is defined as:

$$Coef_{term}(c_i, c_j) = \frac{N_{t_i} \cap N_{t_j}}{N_{t_i} \cup N_{t_j}} \tag{9}$$

where N_{t_i}, N_{t_j} represents the total number of terms in community c_i and c_j.

Based on the above two measures, the function defines a distribution over the link between two semantic communities c_i and c_j of topic. This function is dependent on the word communities over a network that is generated by topic β.

$$\rho(\beta_i, \beta_j) = f(c_i, c_j) = Coef(c_i, c_j) \tag{10}$$

The above formulated topic relatedness function models each per-pair binary variable as a logistic regression with a Jaccard coefficient of communities.

4.3. Posterior inference

The conditional distribution of the topic relationship structure given the observed word co-occurrence network communities is called the posterior. Variational methods are used to find the member of a parameterized family of distributions over the hidden structure that is closest to the posterior. We use the mean-field variational method which has been explored in [17] to solve similar problem to our work, which efficiently obtain an approximation of the objective distribution. In this approach, the posterior is a fully factorized approximation of the form

$$q(\theta, z) = q(\theta|\alpha) \prod_{1}^{N} q(z_n|\phi_n) \tag{11}$$

where $\theta \sim Dirichlet(\alpha)$, $z_n \sim Multi(\phi)$. Our goal is to solve this optimization problem by minimizing the Kullback-Leibler(KL) divergence between the true distribution p and the variational distribution q.

$$\min_{q_1,\ldots,q_D} \mathbb{KL}(q\|p) \tag{12}$$

where we optimize over the parameters of each marginal distribution q_i. And we derive a coordinate descent method, where at each step we make the following update:

$$\log q_j(x_j) = E_{-q_j}[\log \widetilde{p}(x)] + const \tag{13}$$

In our problem, the expectation of the complete log likelihood has a similar formulation in variational inference for the LDA model. For the expectation of the logistic function which is the link based topic relatedness function that provides binary probabilities, we use a first-order approximation that the update is identical to that in variational inference for RTM.

5. Evaluation and experimental results

5.1. Datasets

For better illustration of the experiments, we used two publicly available datasets. The first data set is drawn from the Dublin Core metadata for papers published in the Journal of Statistical Software (JSS). The final JSS data set contains 348 documents and 4279 terms which will be used here to illustrate topic relationship prediction analysis. The second data set is from New York Times corpus (NYT)[1)]. The final NYT dataset comprises of 10,000 total documents with 7123 unique terms and an average document length of 1865. The datasets were pre-processed by separating sentences and removing non-alphabetic characters and single-character words.

[1)] http://archive.ics.uci.edu/ml/

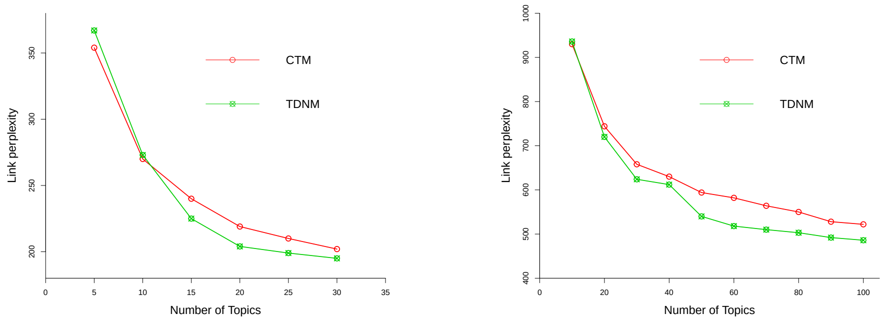

Figure 1: The link perplexity comparison under TDNM and CTM on JSS (Left) and NYT (Right).

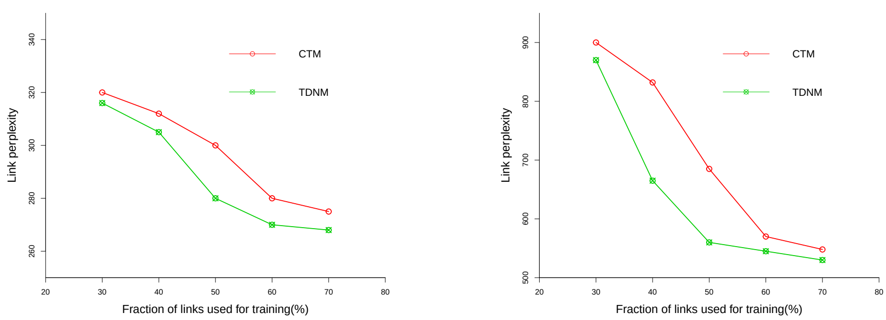

Figure 2: The link perplexity of the predictive distribution comparison on JSS (Left) and NYT (Right).

5.2. *Evaluation of approach*

We choose to compare different topic relationship schemes to experimentally analyze our approach. Correlated Topic Model (CTM)[24] is a typical topic model to model correlations between topics. Our evaluation compares the results of the CTM approach on the two data sets.

Perplexity is a commonly used metric for topic models, which is a measure of the ability of a model to generalize to unseen data[28]. We use the metrics for link perplexity to evaluate the topic dependency relationship structure and to compare the results from TDNM with CTM. Suppose we observe topic relationship links $l_{1:P}$ from a topic dependency network and are interested in which model provides a better relationship predictive distribution $p(l|l_{1:P})$ of the remaining dependency relationships. To compare these predictive distributions, we use link perplexity of set of links L of topic relationship network, which can be defined as follows:

$$linkPerp = exp - \frac{\sum_{i=P+1}^{L} \log p(\tilde{l}_i|l_{1:P})}{|L|} \tag{14}$$

It is a score to qualify link prediction in the relationship dependency network and signifies the capability of the model to capture links among topics in the topic relationship network. It can be thought of as the effective number of equally likely dependency relationship links according to the model. Models that give lower link perplexity to the unseen relationship links better capture the dependency structure of topics. The result of the predictive distribution comparison under TDNM and CTM is shown as in Figure 1 and Figure 2.

We can see that the TDNM model provides more predictive power with lower numbers by observing link perplexity. There is higher numbers of link perplexity for both of TDNM and CTM when a small number of topics have been observed. However, the numbers of the TDNM on link perplexity gradually stable as the number of topics is increased to approach the the optimal number and those of the CTM decrease with the increase of the numbers of topics. Figure 2 shows the link perplexity with varying amounts of the relationship links in the topic network used for training on

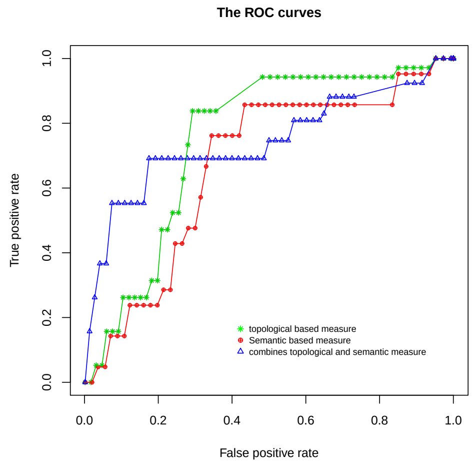

Figure 3: Roc curves of different topic relationship measures.

the datasets. It can be seen that TDNM obtains lower link perplexities than the CTM as the number of relationship links used for training is increased. This is because the CTM uses topic correlation link information to influence the predictive distribution of words and infers that the words in a related topic may be probable. This makes it less effective at predicting links from relationship links observations as the number of relationship links between topics is decreased. In contrast, the TDNM use both of words and its links information in the word occurrence network for links prediction. Thus, it gives predictions less dependent of the number of links between topics.

One aspect of the model is its ability to predict new relationships in the presence of a dependency network. Link prediction is a natural generalization task in topic relationship networks, and another way to measure the quality of our model. Given the words of a new topic, how probable is it links to other topics? The accuracy can be used to measure the performance of the TDNM on link prediction with different link function. We follow other work in this area by defining the accuracy of our model as the receiver operating characteristic curve (ROC). Through the comparisons with the different topic relatedness measures described above, we aims to justify that the effectiveness of integrating external semantic information and extended structured information into the context of the network communities can better leverage the semantic to benefit topic dependency relationship network discovery. The resulting ROC curve displays the overall impact for topic link prediction, as shown in Figure 3. Analysis of the results shown in the figure reveals that the performance of topological measure based and semantic measure based topic relationship prediction is very similar at low false positive rates. However, with the increase of the false positive rate, the true positive rate of topology based measure is larger than that of semantic based measure. We can thus conclude that topology based measure shows better results than semantic based measure. The reason for these results is that topology based measure actually implies semantic information that semantic communities are from topics. Though it is not very obvious in the overall performance, the measure which combines topology and semantic when considering the semantic attribute information and the structure topology information of word co-occurrence network is distributed significantly closer to the upper left corner than the first two measures we analyzed in the beginning. Even though the total area under the ROC curve is roughly the same for the combined topological and semantic measure and for the other topic relatedness measures, the shape is different. In general, we can draw a conclusion that by incorporating words and its links information of a word occurrence network, the model is able to generalize to new topics for which no relationship dependency information was previously known.

5.3. Experimental results

we constructed a word co-occurrence network which shows the existence of topological communities that represent highly interlinked local regions in the network. Network coherence is introduced to analyze the interconnected networks for assessing the impact of topic relationship dependency on TDNM. The part of the resulting word co-occurrence network of two topics is shown in Figure 4(Left). The results clearly show that there are some word

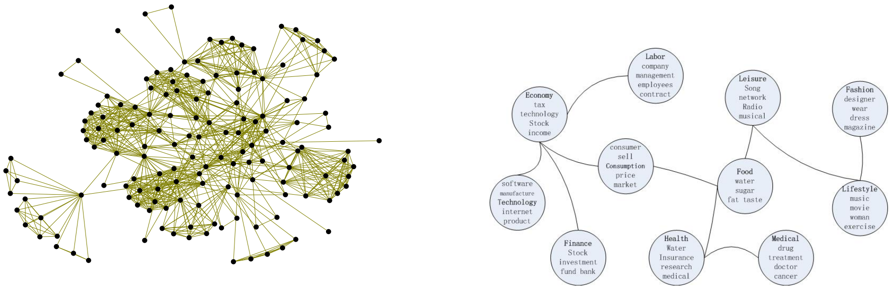

Figure 4: Word co-occurrence network of two topics (Left) and the part of the topics dependency relationships graph on NYT dataset(Right). the class labels of resulting topics are manually assigned. Each topic is labeled with the five most probable terms of its distribution.

community structures which are used to discover the dependency relationship between the two topics by analysis of network coherence.

Furthermore, We conducted an experiment on 10,000 documents randomly selected from the NYT dataset as an example for a more specific topic relationship structure analysis. Our goal is to demonstrate whether it might be the case that the method works well over a specific type of documents, such as those published in the New York Times, which might have specific structures that might not be observed in the other model. Part of this network is illustrated in Figure 4(Right). For convenience to further analysis, each topic is labeled with the five most probable terms from its distribution. We can see that there are dependency relationships between eleven topics. We move down the edge of the network, topics become closely related. The topic of Economy depends on the topics about Technology, Finance, Consumption and Labor, while the Health topic is dependent on the Medical and Food topics. The dependency relationship between topics can be interpreted based on both topological relatedness and Semantic relatedness. With the dependency network representation, some parts of the dependency structure are implicit, for example, whether the topic of Lifestyle should be dependent on the topic of Health or vice versa. Such kind of implicit dependency structure in the topic dependency relationship graph can be further identified by our model in the context of relationship prediction. This shows that our model is flexible and better reflects the nature of the topic dependency structure.

6. Conclusion and future works

A semantically structured methodology is explored for topic relationship dependency structure discovery in this paper. The methodology revolves around a set of high-level topics dependency relationship problems as a probabilistic edge generation problem. The proposed framework combines a complex network analysis approach with the LDA model for topic dependency structure discovery. The work shows the potential of the method that dynamically discovers topic dependency based on probabilistic and complex network technology. Our framework makes assumptions that some topic dependency relationships can be measured in the semantic community of the word co-occurrence. Therefore, the optimality of the semantic communities will have an impact on the efficacy of the obtained results. This shows how the semantics of the content properties and the topology of the graphical structure of the intrinsic semantic dependency relationship between words can be incorporated to identify a semantic community.

While the graphical structure of dependency network may dynamically vary with the change in semantic context because of there are the complex semantic context interrelationships among topics. The TDNM should be adapted to the dynamic topic dependency relationships in the semantic context evolutionary process. This is a worth investigating issue for our future work.

References

[1] H. Liu, Linguistic complex networks: A new approach to language exploration, Grkg/Humankybernetik 52 (4) (2011) 151-170.
[2] J. Cong, H. Liu, Approaching human language with complex networks,Physics of Life Reviews 11 (4) (2014) 598-618.
[3] G. A. Wachs-Lopes, P. S. Rodrigues, Analyzing natural human language from the point of view of dynamic of a complex network, Expert Syst. Appl.45 (C) (2016) 8-22.

[4] S. Martini-Ipi, D. Margan, A. Metrovi, Multilayer network of language: A unified framework for structural analysis of linguistic subsystems, Physica A: Statistical Mechanics and its Applications 457 (2016) 117-128.
[5] M. Sedighi, Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of informetrics), Library Review 65 (1/2) (2016) 52-64.
[6] J. Bi, J. Huang, Z. Qin, A relationship strength-aware topic model for communities discovery in online social networks, in: Advances in Computer Science and its Applications, Vol. 279 of Lecture Notes in Electrical Engineering, Springer Berlin Heidelberg, 2014, pp. 709-715.
[7] X. Wang, N. Mohanty, A. Mccallum, Group and topic discoveryfrom relations and their attributes, Advances in Neural Information Processing Systems 18 (2006) 1449-1456.
[8] A. McCallum, X. Wang, C. Emmanuel, Topic and role discovery in social networks with experiments on enron and academic email, Journal of Artificial Intelligence Research 30 (1) (2007) 249-272.
[9] L. Dietz, S. Bickel, T. Scheffer, Unsupervised prediction of citation inffluences,in: Proceedings of the 24th International Conference on Machine Learning, ICML '07, 2007, pp. 233-240.
[10] R. M. Nallapati, A. Ahmed, E. P. Xing, W. W. Cohen, Joint latent topic models for text and citations, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'08, 2008, pp. 542-550.
[11] E. Wang, J. Silva, R. Willett, L. Carin, Dynamic relational topic model for social network analysis with noisy links, in: IEEE the Statistical Signal Processing Workshop (SSP), 2011, pp. 497-500.
[12] K. Balog, M. Rijke, R. Franz, H. Peetz, B. Brinkman, I. Johgi, M. Hirschel,Sahara: Discovering entity-topic associations in online news, in: In 8th International Semantic Web Conference (ISWC 09), 2009.
[13] Z. Yin, L. Cao, Q. Gu, J. Han, Latent community topic analysis: Integration of community discovery with topic modeling, ACM Trans. Intell. Syst.Technol. 3 (4) (2012) 63:1-63:21.
[14] D. Broniatowski, C. Magee, Towards a computational analysis of status and leadership styles on fda panels, in: Social Computing, Behavioral Cultural Modeling and Prediction, Vol. 6589 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2011, pp. 212-218.
[15] D. Dicheva, C. Dichev, Helping courseware authors to build ontologies:The case of tm4l, in: Proceedings of the 2007 Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work, IOS Press, Amsterdam, The Netherlands, The Netherlands,2007, pp. 77-84.
[16] B. Fortuna, D. Mladeni, M. Grobelnik, Semi-automatic construction of topic ontologies, in: Semantics, Web and Mining, Vol. 4289 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, pp. 121-131.
[17] Y. Liu, A. Niculescu-Mizil, W. Gryc, Topic-link lda: Joint models of topic and author community, in: Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09, ACM, New York, NY, USA,2009, pp. 665-672.
[18] Y. Jo, J. E. Hopcroft, C. Lagoze, The web of topics: Discovering the topology of topic evolution in a corpus, in: Proceedings of the 20th International Conference on World Wide Web, WWW '11, ACM, New York, NY, USA, 2011, pp. 257-266.
[19] 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, Know.-Based Syst. 26 (2012) 164-173.
[20] D. I. Inouye, P. Ravikumar, I. S. Dhillon, Admixture of poisson mrfs: A topic model with word dependencies, in: International Conference on Machine Learning (ICML), 2014, p. 683691.
[21] E. MEEKS, S. B. WEINGART, The digital humanities contribution to topic modeling, Journal of Digital Humanities 2 (1).
[22] Q. Mei, D. Cai, D. Zhang, C. Zhai, Topic modeling with network regularization,in: Proceedings of the 17th International Conference on World Wide Web, WWW '08, ACM, New York, NY, USA, 2008, pp. 101-110.
[23] J. Chang, D. M. Blei, Relational topic models for document networks., in:AISTATS, Vol. 5 of JMLR Proceedings, JMLR.org, 2009, pp. 81-88.
[24] D. M. Blei, J. D. Lafferty, Correlated topic models, in: In Proceedings of the 23rd International Conference on Machine Learning, MIT Press, 2006,pp. 113-120.
[25] J. Foulds, S. Kumar, L. Getoor, Latent topic networks: A versatile probabilistic programming framework for topic models, in: International Conference on Machine Learning (ICML), 2015, pp. 777-786.
[26] S. Das, Topics and relationship discovery from unstructured equivocal sources for situation assessment, in: 19th International Conference on Information Fusion, FUSION 2016, Heidelberg, Germany, July 5-8, 2016,pp. 122-129.
[27] N. Shibata, Y. Kajikawa, I. Sakata, Measuring relatedness between communities in a citation network, Journal of the American Society for Information Science and Technology 62 (7) (2011) 1360-1369.
[28] G. Heinrich, Parameter estimation for text analysis, Tech. rep. (2004).