Applied Network Science

RESEARCH Open Access

Improving topic modeling through homophily for legal documents

Kazuki Ashihara[1], Cheikh Brahim El Vaigh[4], Chenhui Chu[2*], Benjamin Renoust[2], Noriko Okubo[3], Noriko Takemura[2], Yuta Nakashima[2] and Hajime Nagahara[2]

*Correspondence:
chu@ids.osaka-u.ac.jp
[2] Inria, IRISA, Rennes, France
Full list of author information is available at the end of the article

Abstract

Topic modeling that can automatically assign topics to legal documents is very important in the domain of computational law. The relevance of the modeled topics strongly depends on the legal context they are used in. On the other hand, references to laws and prior cases are key elements for judges to rule on a case. Taken together, these references form a network, whose structure can be analysed with network analysis. However, the content of the referenced documents may not be always accessed. Even in that case, the reference structure itself shows that documents share latent similar characteristics. We propose to use this latent structure to improve topic modeling of law cases using document homophily. In this paper, we explore the use of homophily networks extracted from two types of references: prior cases and statute laws, to enhance topic modeling on legal case documents. We conduct in detail, an analysis on a dataset consisting of rich legal cases, i.e., the COLIEE dataset, to create these networks. The homophily networks consist of nodes for legal cases, and edges with weights for the two families of references between the case nodes. We further propose models to use the edge weights for topic modeling. In particular, we propose a cutting model and a weighting model to improve the relational topic model (RTM). The cutting model uses edges with weights higher than a threshold as document links in RTM; the weighting model uses the edge weights to weight the link probability function in RTM. The weights can be obtained either from the co-citations or from the cosine similarity based on an embedding of the homophily networks. Experiments show that the use of the homophily networks for topic modeling significantly outperforms previous studies, and the weighting model is more effective than the cutting model.

Keywords: Homophily network, Topic modeling, Legal documents

Introduction

Computational law applies quantitative computational methods to the study of laws, intersecting various research fields such as natural language processing, network science, machine learning, statistical methods, and sociophysics (Katz 2011). In the last decade, the interest in studying laws from the perspective of computational social science has significantly increased (Lettieri and Faro 2012). It has been shown that no matter the country, an efficient approach for searching patterns in citation analysis of legal documents will always become a complex network analysis study (Fowler et al. 2007; Kim 2013; Pelc 2014; Koniaris et al. 2017; Khanam and Wagh 2017; Lettieri et al. 2018; Lee et al. 2019).

In computational law, topic modeling is one key task for many higher-level goals such as law document retrieval, law document clustering for exploration, or the comparison of legal documents (Katz et al. 2011; Lu et al. 2011; O'Neill et al. 2016; Wang et al. 2017; Yoshioka et al. 2018). To date, LDA is still the favored model for topic modelling, but does not take into account the larger context of laws and cited cases. However, ambiguity may still arise in the topics because law is a specific domain: different real-world contexts can lead to the breach of a same law, while a same real-world context can breach different law cases—meaning that the relevance of the modeled topics strongly depends on the legal context they are used in Kanapala et al. (2019). In the case of judgement cases, the citations to preexisting cases and laws are as important as the content to render a decision—similarly to scientific work.

Academic evaluation has frequently used citation analysis (Hirsch 2005), despite the controversies that have been shown even recently (Renoust et al. 2017). Using scientific citation networks has also improved the quality of topic modeling (Chang and Blei 2009). In comparison to academic citation networks, the structure of legal cases often follows a directed acyclic network (or DAG). The structure of these DAG may however differ. While academic citation networks are often deep with many authors citing one another, and even self citations, the structure of legal citations may look more shallow. This results in a flat organization with little cross references between cases but a clear emphasis in precedent cases making jurisprudence (Pelc 2014). As a consequence, the very shape of the networks built from legal citations shows a different topology to those built from academic citations.

In addition, since the content of the cited documents is not always available, network analysis then relies on the investigation of the co-citation patterns (Kim 2013; Khanam and Wagh 2017). Sometimes, this is referred to as homophily (McPherson et al. 2001; Borgatti et al. 2009), also known as the property of entities to agglomerate when being similar and the implied similarity of two entities. Homophily always corresponds to a bipartite structure, which can be projected into a single type network. Topic modeling is one approach to using homophily in the projections (Renoust et al. 2014). In the context of computational law, retrieval and topic modeling give rise to open challenges and publicly datasets such as the COLIEE data (Yoshioka et al. 2018). The COLIEE dataset provides a testbed for legal information extraction and entailment. It provided over 6k cases from the Canadian Federal Court for about 40 years, with very rich annotations including among a lot of different entities, citations to past cases, rulings, and laws.

Our work contributes a methodology for building topic modeling for legal documents, when the content of cited documents is not available. We propose in the current work to automatically build networks from cases of the COLIEE dataset. We analyze the COLIEE dataset. We then construct a homophily network consisting of nodes for legal cases and edges with weights for the references. There are two major types of citations. The first one refers to prior cases, while the other one to statute laws. We further propose to use these two types of citations to explore the similarity of cases, by constructing homophily relationships between cases. Furthermore, we use case homophily to improve topic modeling for legal cases. In particular, we work on the relational topic model (RTM) (Chang and Blei 2009) that uses the links between documents during topic modeling. We compare different strategies of using the edge weights in the homophily network as link information for RTM.

This work is an invited extension of the original presentation (Ashihara et al. 2019). In this paper, we extend our previous work as follows:

- We improve the strategy of using homophily relationships for topic modeling. Previously, we only set a threshold based on edge weights in the homophily network to decide whether a link should be used for RTM or not, which we call the cutting model. In this paper, we also propose to use the weights to weight the link probability in RTM, which we call the weighting model. Experiments show that our weighting model significantly outperforms our previous cutting model.
- In addition to the product or the sum of prior case and status information, we propose a list of weighting methods with fuzzy logic aggregation (Detyniecki et al. 2000) for the weighting model, showing similar coherence score to the simple weighting model and better coherence scores than previous cutting model.
- We also investigate the use of a kernel such as Node2vec (Grover and Leskovec 2016) to embed homophily network in low-dimension space. Experiments show that our Node2vec model, also significantly outperforms our previous cutting model.
- We analyze the topic words in detail for the best performing topic model, and verify the effectiveness of the proposed models for legal case topic modeling.

The remainder of the paper is as follows: Sect. 2 presents the related work; Sect. 3 introduces the COLIEE dataset (Yoshioka et al. 2018) along with its characteristics; Sect. 4 presents homophily network modeling for this dataset. Section 5 describes topic modeling for legal cases and our proposed models for improving RTM using homophily networks; we report experiments and results in Sect. 6 and conclude the paper in Sect. 7.

Related work

In the current section, we present the related work to both legislation networks and topic modeling. In addition, we discuss the difference between our work and the related work.

Legislation networks

The interest in network analysis for legal documents has been significantly increasing recently. Many previous studies have shown that the analysis of legal networks is closely related to complex networks (Fowler et al. 2007; Kim 2013; Pelc 2014; Koniaris et al. 2017; Khanam and Wagh 2017; Lettieri et al. 2018; Lee et al. 2019). Fowler et al. (2007) developed a centrality measure based on *Authorities* and *Hubs* (Kleinberg 1999), which is dedicated to citations of cases in the US Supreme Court consisting of 26k+ cases in a citation network. In order to find complex network properties and homophily behavior in a treaty network, Kim (2013) explored a structure consisting of 1k citations for 747 treaties. Pelc (2014) investigated the fundamental precedent concept, i.e., previous deliberations being cited in cases, in the international commercial cases. They also did a centrality study of *Authorities* and *Hubs*, which confirms that the network structure is relevant to predicting case output. From the Official Journal of the European Union, Koniaris et al. (2017) built a law reference network. They showed that it has the temporal evolution and multi-scale structure property of multilayer complex networks. With betweenness centrality, Khanam and Wagh (2017) proposed an analysis on citation for judgements in Indian courts. The relevance of the EUCaseNet project (Lettieri et al. 2018) should also be underlined. It combines network analysis and centrality-based visualization to explore the entire EU case law corpus. Lee et al. (2019) explored the court decision versus constitution article patterns in Korea, which conducts topic analysis on the main clusters. Because in this paper we investigate the the Federal Court of Canada case law network, our target is close to these studies. We investigate the homophily in our analysis, which has been illustrated by all of the studies above. Although applying topic modeling for legislation networks is not new, we take the further step of using network analysis to improve topic modeling. Different from previous studies, we improve topic modeling with the case co-citation structure, and feed back to homophily of documents in ways of topic proximity.

Topic modeling

Latent Dirichlet allocation (LDA) is the first topic model introduced by Blei et al. (2003). As a graphical model, LDA can learn from observed documents to infer hidden word and document-topic distributions. In Sect. 5, we give a description of LDA in detail. The correlated topic model, proposed by Blei and Lafferty (2007), models topic occurrences using the logistic normal for LDA. The dynamic topic model proposed by Blei et al. (2006) models temporal information in sequence data. Most topic models are unsupervised, but supervised topic modeling has been studied too. The supervised LDA proposed by Blei and McAuliffe (2007) can model topics of responses and documents. Supervised LDA is suitable for data such as product reviews, which has both evaluation scores and corresponding descriptions of products. Ideal point topic models, proposed by Nguyen et al. (2015), assume that the responses are also hidden. RTM models the topic of a document pair, which shares links, e.g. references, between a document pair (Chang and Blei 2009). We describe RTM in detail in Sect. 5. Collaborative topic models proposed by Wang and Blei (2011), can make recommendation for user preferences using user data.

Recent studies have also tried to bridge topic models to text representation methods based on word embeddings. E.g., Das et al. (2015) modeled topics with distributions on word embeddings instead of word types and showed that the proposed model is more robust for handling out-of-vocabulary words; similarly, Dieng et al. (2019) developed an embedded topic model that models words with categorical distributions on word embeddings and topic embeddings. We think that this can be one interesting direction for our future work.

In the context of joint network and topic modeling, Liu et al. (2009) proposed a framework to perform LDA-based topic modeling and author community discovery simultaneously; Zhu et al. (2013) proposed a mixed-topic link model for joint topic modeling and link prediction; Brochier et al. (2020) proposed a topic-word attention mechanism to generate document network embeddings via the interaction between topic and word

Table 1 An example of reference identification within legal cases

Prior case	New Brunswick (Board of Management) v. Dunsmuir, [2008] 1 S.C.R. 190; 372 N.R. 1; 329 N.B.R.(2d) 1; 844 A.P.R. 1, refd to. [para. 11]
⟶	New Brunswick (Board of Management) v. Dunsmuir, (2008)
Statute	Immigration and Refugee Protection Act, S.C. 2001, c. 27, sect. 113(b) [para. 14.]
⟶	Immigration and Refugee Protection Act, S.C. (2001)

embeddings. Different from previous studies, in this paper, we apply RTM for legal case analysis and improve it via co-citation homophily networks. To the best of our knowledge, this is the first work that utilizes co-citation homophily networks for topic modeling.

Data

We collect our data from the Competition on Legal Information Extraction/Entailment (COLIEE 2018) (Yoshioka et al. 2018).[1] For our task we study the *Case Law Competition Data Corpus*, which has also been used in *Task 1* and *2* in COLIEE 2018. The data consists of 6154 cases from the Federal Court of Canada over the period of approximately 40 years, ranging between 1974 and 2016. Note that most cases in the corpus are the ones with a date after 1986. This data corpus is very rich. Each case is a textual document containing multiple parts, including a summary of the court, case content, references to relevant past cases and statutes, rulings, counsels, legal topics of interest, solicitors, miscellaneous information, and important facts.

In this paper, we only focus on the prior cases and statutes noticed to form our networks. From the text input, they are divided by paragraph titles as follows:

- **Cases Noticed**, they correspond to the past trials which are relevant to this trial.
- **Statutes Noticed**, they correspond to laws referred to give the verdict of the trial.

Each consecutive line has one reference. Recall that as they are Canadian cases, they may be written in both English and French. We found that there are only 5576 cases that refer to prior cases and statutes noticed among all the cases.

The reference destination is always very detailed, and references can be separated into paragraphs or chapters. If we directly use these as basic units for analyzing network modeling, only a small number of references are redundant across the cases, making the network very sparse. Therefore, we consider the references to the full case or statute articles. The identification of each case or statute can be made based on a year, a title, and references. The parsing is conducted based on looking for the year structure, and make titles at a high granularity as nodes (Table 1). The cases without information about the year are discarded. In total, these correspond to 39 cases being cited. We also save the year information along with the nodes.

[1] https://sites.ualberta.ca/~miyoung2/COLIEE2018/.

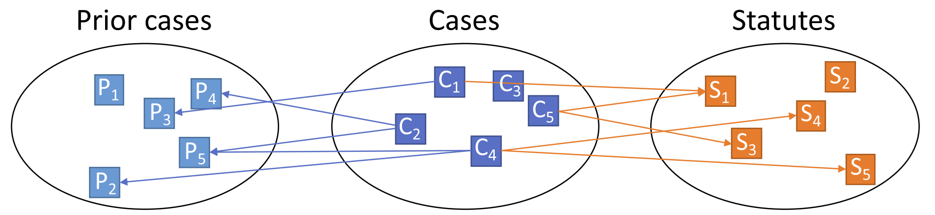

Fig. 1 An overview of the structure of the legal case data in the COLIEE dataset

Legal network

In this section, we first describe the network model in Sect. 4.1. Second, we provide the details about the construction of the underlying homophily network in Sect. 4.2. Finally, in Sect. 4.3 we lay out on the use of Node2vec to embed the homophily network in a low dimensional space.

Network structure

In our data, each case refers to a set of prior cases and statutes noticed. In our network, $G = (V,E)$, every case is represented as a node $v_1 \in V$, and a prior case or statute noticed is represented as another node $v_2 \in V$. We treat each citation as a link $(v_1, v_2) = e \in E$. Figure 1 shows an overview of our network modeling.

Our initial set contains $|C| = 5539$ cases. Each case can refer to multiple prior cases and statutes noticed. Each case $c \in C$ may refer to several prior cases $p \in P$, where $|P| = 25{,}112$ in total. They also can have reference to a statute $s \in S$, where $|S| = 1288$. The citations to cases can be from the eighteenth century. Note that it cannot be guaranteed that the year information is reliable for cases before the eighteenth century, which constitute a small number of 78 cases.

With the above network modeling, our network includes 31,976 nodes, and 53,554 links. Note that reliable year information is only available for 29,319 nodes. We can separate the network into two sub-networks. The first, G_P, is constructed using only the cases and their cited prior cases, consisting of 29,952 nodes with 53,554 edges. The second, G_S, only considers the cases and their cited statutes, consisting of 4441 nodes with 6453 edges.

These two networks present one main connected component G consisting of 30,456 nodes with 52,453 links, which covers most nodes and edges. The node/link number for G_P is 27,353/44,871, and 4125/6150 for G_S. We further investigate the possibility of looking for case communities from these networks via Louvain clustering (Blondel et al. 2008) and modularity (Newman 2006). The main components of G, G_P, and G_S show a modularity $Q_G = 0.739$ with 34 communities, a modularity $Q_{G_P} = 0.762$ for 45 communities, and a modularity $Q_{G_S} = 0.747$ for 27 communities, respectively. This is illustrated in Fig. 2.

If the communities were extremely imbalanced—consider an extremely large community surrounded by very small others—there would be more chances to be unsuccessful looking for homophily because most documents would have higher chances to share the same few common characteristics. Finding community structures within the

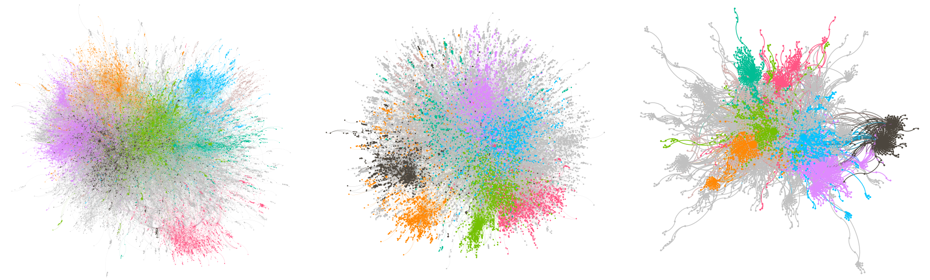

Fig. 2 Community structure through citations (left: G, middle: G_P, right: G_S)

citation network confirms that we can leverage on these structures by using homophily. In other words, there are groups of documents that share latent characteristics, and that may be differentiated enough to other groups.

Homophily network

Prior case and statute citations indicate double bipartite structures in our network model, i.e., from G_P are *case–prior case*, and from G_S are *case–statute* relationships. Bipartite projections into one-mode networks indicate a complex network structure (Guillaume and Latapy 2006). Thus, we further derive three one-mode networks: $G'_P = (C, E'_P)$, $G'_S = (C, E'_S)$, and $G' = (C, E')$, in order to analyze homophily; where C represents nodes, E represents edges, and G' is a combination of G'_P and G'_S,

Homophily is the property for entities that are linked together to share some existing characteristics, that may also be shared among a group or a cluster. Although this may be captured through an *entity–characteristics* association, naturally forming bipartite networks, other models, including multilayer networks and hypergraphs could also be investigated (Renoust 2013). In the context of homophily, multilayer networks and bipartite graphs have equivalence (Renoust 2014), and there exists the incidence/Levi graph for hypergraphs (Levi 1942). All may be projected into 1-mode networks connecting the entities which may be linked by homophily. This single mode projection is the artifact which enables us to alleviate the limitation of not having access to the cited content but still embedding the network structure within our topic modeling.

To this end, we project the other two bipartite relationships onto *case–case* relationships. That is, let $u \in C$, $v \in C$ be two initial cases that we are investigating. We can assign a set of references $R_u = \{r_1, r_2, r_3, \ldots\}$ to each of these cases, where r_x represents either a prior case or a statute noticed. In a projected network G', the original cases $\{u, v\} \subseteq C$ become the nodes, and there exists a link $(u, v) = e \in E'$ if and only if the intersection of their respective reference sets is non empty: $R_u \cap R_v \neq \varnothing$. Each reference $r_x \in P$ is a prior case in the network induced by prior cases G'_P. Each reference $r_x \in S$ is a statute law in the network induced by statutes G'_S. A reference $r_x \in S \cup P$ can be of either a case or a statute noticed in the general projected network G'.

We can obtain the weight of each link with the following two methods. In the first one, w_n is the number of shared citations between two cases. In the second one, w_j is the Jaccard index of these cases (Jaccard 1901).

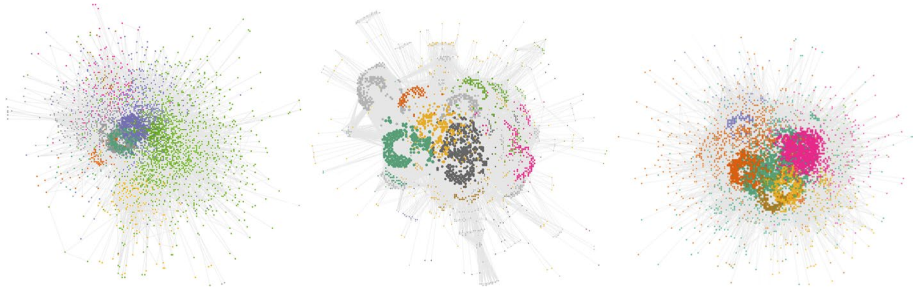

Fig. 3 Network projection and communities (left: G'_P, middle: G'_S, right: G')

$$w_n = |R_u \cap R_v| \quad and \quad w_j = \frac{w_n}{|R_u \cup R_v|} \tag{1}$$

We find that the resulting networks are very dense, where the numbers of nodes and edges are 4803/286,435 for G'_P, 3138/379,447 for G'_S, and 5576/643,729 for G', respectively. We can see that there is only a little overlap between links induced by prior cases and statutes. After investigating the main components of these networks, we get a size of 4244/286,403, 3033/379,426, and 4870/643,725 nodes and edges for G'_P, G'_S and G', respectively. G'_P has modularity $Q_{G'_P} = 0.428$ for 14 communities, G'_S has modularity $Q_{G'_S} = 0.542$ for 13 communities, and G' has modularity $Q_{G'} = 0.502$ for 7 communities. Figure 3 visualizes these networks with their communities.

Embedding of homophily network with Node2vec

Homophily networks can be embedded in a low dimensional space using kernels such as Node2vec (Grover and Leskovec 2016). Node2vec aims at embedding networks in low dimensional representations while preserving their properties such as node neighborhood, roles, or communities based on homophily. Node2vec is based on the model skip-grams (Mikolov et al. 2013) used for embedding words leveraging their contexts. To compute node embedding, node2vec operates the model skip-grams over random walks in the graph, allowing it to represent the neighbourhood and the overall position of each node in the graph. Our aim is to embed the homophily networks and exploit the semantics of their embeddings, as provided by Node2vec. Thus we can obtain the weight of each link by computing the cosine similarity between the two nodes of the link under consideration. For a link $l = (n_i, n_j)$, one can compute the weight

$$w_s = sim(v_{n_i}, v_{n_j}), \tag{2}$$

where v_x is the embedding of the node x provided by the Node2vec embedding,[2] and $sim(\cdot, \cdot)$ is the cosine similarity. Using cosine similarity to weight links in the network is more general and robust, as this weight incorporates the similarity of two nodes and also the similarity of their neighborhood. The final Node2vec embedding of the case and status homophily networks is shown in Fig. 4 using UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) (McInnes et al. 2018), a dimension

[2] We used the best hyper-parameters for Homophily following (Grover and Leskovec 2016) to get v_x.

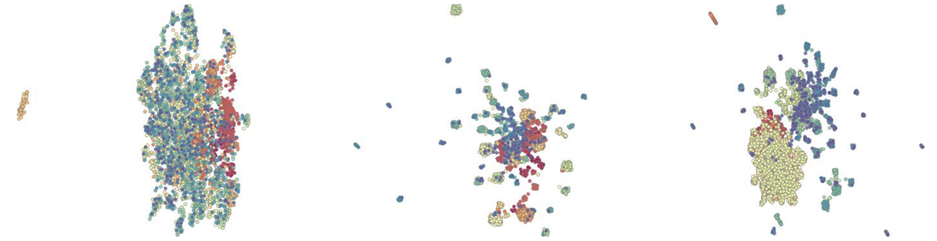

Fig. 4 Node2vec embedding of the networks using UMAP (left: G'_P, middle: G'_S, right: G')

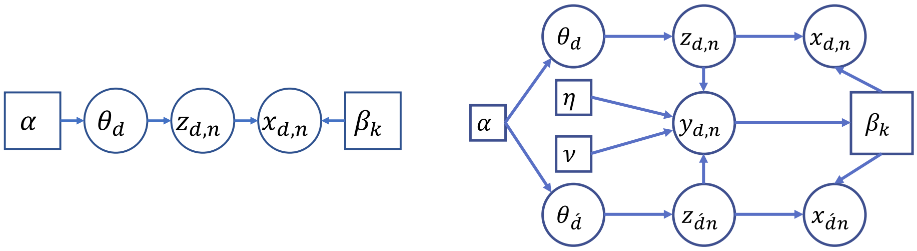

Fig. 5 Statistical model of LDA (left) and RTM (right)

reduction technique that can be used for visualisation similar to t-SNE. Node2vec embeddings preserve the property of homophily. As shown in Fig. 4, one can see the different communities based, on prior case for G'_P, status for G'_S, and both prior case and status for G'. Node2vec can force words in similar documents to fall in the same topics because the nodes are cases/status and Node2vec leverage only random walks to clusters nodes.

Relational topic model with complex network

In this section, the classical topic model of LDA (Blei et al. 2003) is first introduced, and second the RTM (Chang and Blei 2009) is described (as illustrated in the left and right parts of Fig. 5, respectively). We then integrate the weights obtained by the homophily relationships to RTM for legal document analysis.

LDA (Blei et al. 2003) is a generative model. In LDA, documents are represented as a mixture of latent topics, and each topic is characterized by a distribution over the vocabulary in the documents. α and β are parameters in corpus-level, which are sampled when generating the corpus. θ_d are document-level variables, which are sampled for each document. We assume that words in documents are generated by the topic probability. From α and θ_d, the topic appearance probability is generated in document d. Then, the word topic probability $z_{d,n}$ is generated in document d. At last, the word $x_{d,n}$ for the nth word in the dth document is generated from the occurrence probabilities of the vocabulary in the topic k of β_k and $z_{d,n}$. The LDA model only focuses on documents themselves, without considering the relationships among documents.

The link relationship between documents is considered in the RTM (Chang and Blei 2009). Same as LDA, firstly documents are generated from topic distributions in RTM. Next, document links are modeled as binary variables, with one link for one pair of documents. Given a pair of documents d, and $\acute{d}$, an indicator of binary link is drawn as:

$$y|\mathbf{z}_d, \mathbf{z}_{\acute{d}} \sim \psi(\cdot|\mathbf{z}_d, \mathbf{z}_{\acute{d}}),$$

where ψ is the probability function of distributed link between the document pair. Sigmoid or exponential have been used for ψ. We adopt the exponential, because it performs better in the experiments reported in Chang and Blei (2009). ψ depends on the topic assignments z_d and $z_{\acute{d}}$, which generate their words. It is calculated as:

$$\psi = \exp\left(\eta^{\mathrm{T}}(\overline{\mathbf{z}}_d \circ \overline{\mathbf{z}}_{\acute{d}}) + \nu\right), \tag{3}$$

where $\overline{\mathbf{z}}_d = \frac{1}{N_d}\sum_n z_{d,n}$, $\circ$ denotes element-wise product, and coefficients η and intercept ν are parameters, which can be estimated as:

$$\begin{aligned} \nu &= \log(1 - \mathbf{1}^{\mathrm{T}}\overline{\Pi}) - \log\left(\rho\frac{K-1}{K} + 1 - \mathbf{1}^{\mathrm{T}}\overline{\Pi}\right) \\ \eta &= \log(\overline{\Pi}) - \log\left(\overline{\Pi} + \frac{\rho}{K^2}\mathbf{1}\right) - \mathbf{1}\nu \end{aligned}$$

In the above two equations, $\mathbf{1}$ is the vector whose elements are all 1, ρ is a scalar to control the frequency of the negative observations of no links, and $\overline{\Pi}$ is given by

$$\overline{\Pi} = \sum_{(d,\acute{d})} \mathbb{E}[\overline{z}_d] \circ \mathbb{E}[\overline{z}_{\acute{d}}]$$

where the summation is computed over all possible documents pairs.

We use the homophily network structure presented in Sect. 4.2 to improve on RTM. We propose two models, called the cutting model and weighting model, in order to strengthen the effect of co-citation patterns in the networks. Prior cases and statutes can have different influences to the judgement of current cases, but the inferences are case by case and thus data-driven; therefore, we design different weighting schemes to model the different influences. The motivation behind the cutting/weighting models is to take the best of both prior cases and statutes. Besides using the networks of prior cases only or the networks of statutes only, we also propose to use both of them by aggregating their weights.

Cutting model

The cutting model sets a threshold to filter noisy co-citations with low weights, while keeping the most influential statute laws and prior cases. In this model, we use either w_n or w_j in Eq. (1) to cut inefficient links, only keeping the links with edge weights higher than a threshold. The links above will be used in RTM as document links. We tried different thresholds of edge weights in the three homophily networks, G'_P, G'_S, and G', respectively, in our experiments.

Weighting model

In this model, instead of cutting the links with edge weights, we keep all the links but use edge weights to weight the link probability function, i.e., Eq. (3). Note that we only use w_j for weighting, because of the probability characteristic of Eq. (3). Formally, the weighting is performed as follows:

$$\psi' = w \times \psi,$$

we compare different types of weighting methods:

- $w = w_j^P$: only use the edge weights w_j in G'_P.
- $w = w_j^S$: only use the edge weights w_j in G'_S.
- $w = w_s^P$: only use the edge weights w_s in G'_P.
- $w = w_s^S$: only use the edge weights w_s in G'_S.
- $w = w_j^P + w_j^S$: use the sum of the edge weights w_j in G'_P and G'_S.
- $w = w_j^P \times w_j^S$: use the product of the edge weights w_j in G'_P and G'_S.
- $w = w_s^P + w_s^S$: use the sum of the edge weights w_s in G'_P and G'_S.
- $w = w_s^P \times w_s^S$: use the product of the edge weights w_s in G'_P and G'_S.

Fuzzy logic aggregation, extensively studied in Detyniecki's thesis (Detyniecki et al. 2000), can help balancing between the different types of weights. Fuzzy logic, especially triangular norm fuzzy logic (*t-norm*) which guarantees triangular inequality in probabilistic spaces, generalizes intersection in a lattice and conjunction in logic, offering many aggregation operators to define conjunction for values within [0,1] (Schweizer and Sklar 2011). Each t-norm operator is associated with an s-norm (t-conorm) with respect to De Morgan's law (Hurley 2005): $S(x,y) = 1 - T(1-x, 1-y)$. The t-norm is the standard semantics for conjunction in fuzzy logic and thus the couple t-norm/s-norm acts as *AND*/*OR* operators on real values in [0,1]. We experiment with several fuzzy operators: beside the classical sum/product, we also consider the family of Hamacher t-norms (Hamacher product (Hamacher 1976)) defined for $\lambda \geq 0$ as

$$T_{\mathrm{H},\lambda}(x,y) = \frac{xy}{\lambda + (1-\lambda)(x+y-xy)}, \tag{4}$$

the family of Yager t-norms (Yager 1980) defined for $\lambda > 0$ as

$$T_{\mathrm{Y},\lambda}(x,y) = \max \begin{cases} 0 \\ 1 - \sqrt[\lambda]{(1-x)^{\lambda} + (1-y)^{\lambda}} \end{cases}$$

and the Einstein summation (Einstein 1916)

$$T_{\mathrm{E}}(x,y) = \frac{xy}{1 + (1-x)(1-y)}.$$

recall that $T_{H,0}$ is the classical product/sum. Fuzzy logic aggregators can also be used as weighting methods. We thus consider in addition to the methods presented above, the following weighting strategies:

- $w = T_{H,\lambda}(w_j^P, w_j^S)$: the aggregation of w_j in G'_P and G'_S, using Hamacher t-norm.
- $w = T_{\mathrm{Y},\lambda}(w_j^P, w_j^S)$: the aggregation of w_j in G'_P and G'_S, using Yager t-norm.
- $w = T_{\mathrm{E}}(w_j^P, w_j^S)$: the aggregation of w_j in G'_P and G'_S, using Einstein sum.

In summary, the sum and product methods model the use of prior cases and statutes in an OR and AND relation, respectively. The fuzzy logic aggregation can further take a trade-off between the OR and AND relations. The same weighting strategies are also used to aggregate the weights obtained with Node2vec embedding, where w_s is used instead of w_j.

Experiments

We used the Canadian law corpus (Section 3), to conduct experiments for topic modeling.

Settings

For preprocessing, we performed the following. Words were lemmatized and lower-cased using NLTK (Loper and Bird 2006). We also discarded word tokens containing non-alphabetic characters. In addition, we excluded stop words by using the English stop word list in NLTK.[3]

We compared the homophily networks G'_P, G'_S, and G'. Using the edge weights obtained from Eqs. (1) and (2), we either judged which edge to use according to a threshold in the cutting model (Sect. 5.1) or weight the edges using the methods in the weighting model (Sect. 5.2). We used a publicly available RTM implementation for all our experiments.[4] The parameters of RTM were trained using the obtained network information and documents.[5] The number of topics and max iterator were set to 200, and 10, respectively.

The Node2vec kernel is used with 50 dimensions, an inout hyper-parameter set to 0.5 to preserve homophily property, and the remaining parameters are kept by default.

As a baseline, we compared with LDA (Blei et al. 2003),[6] which does not use link information. Furthermore, we compared to another baseline RTM ($w \rightarrow \infty$), which is an RTM that does not use any link information. For all the experiments, we used the default values for the corpus-level parameters α and β (see Fig. 5), which were 0.1 and 0.01, respectively.

Evaluation metrics

To evaluate the output topics of the different models, we essentially used the coherence score (Newman et al. 2010). Coherence measures the similarity among the output topic words, which is commonly used for evaluating topic modeling performance. Coherence can be computed as:

$$coherence = \sum_{k=2}^{N} \sum_{l=1}^{k-1} sim(word_k, word_l),$$

where $word_k$ and $word_l$ are the kth and lth topic words output by a topic model , and N is the number of output topic words. $sim(\cdot, \cdot)$ is the cosine similarity of two words. The cosine similarity is calculated by representing the two word with word embeddings by

[3] https://gist.github.com/sebleier/554280.

[4] https://github.com/dongwookim-ml/python-topic-model/blob/master/notebook/RelationalTopicModel_example.ipynb.

[5] https://github.com/dongwookim-ml/python-topic-model.

[6] https://radimrehurek.com/gensim/models/ldamodel.html.

Table 2 Coherence scores of topics from G'_P and G'_S against w_n, w_j, cutting and weighting models

	G'_P, w_n		G'_P, w_j		G'_S, w_n		G'_S, w_j	
LDA		0.131		0.131		0.131		0.131
RTM	$w_n \to \infty$	0.159	$w_j \to \infty$	0.159	$w_n \to \infty$	0.159	$w_j \to \infty$	0.159
Cutting	$w_n \geq 100$	0.166	$w_j \geq 0.75$	0.160	$w_n \geq 10$	*0.167*	$w_j \geq 0.75$	0.164
	$w_n \geq 50$	0.165	$w_j \geq 0.50$	0.166	$w_n \geq 5$	0.159	$w_j \geq 0.50$	0.164
	$w_n \geq 5$	*0.167*	$w_j \geq 0.25$	0.161	$w_n \geq 0$	0.164	$w_j \geq 0.25$	0.165
	$w_n \geq 0$	0.166	$w_j \geq 0$	0.166			$w_j \geq 0$	0.164
Weighting			w_j^P	*0.180*			w_j^S	*0.176*
			w_s^P	0.170			w_s^S	0.171

GloVe840B (Pennington et al. 2014).[7] Note that we used the top ten words output by topic models to calculate coherence.

In addition, we also reported the C_V (Röder et al. 2015), the C_{UMass} (Mimno et al. 2011), and the C_{UCI} (Newman et al. 2010) scores to confirm the performance consistency among different evaluation metrics. The C_V metric is reported as the measures with strongest correlations with human ratings. C_V is based on a sliding window, a one-set segmentation of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosine similarity.

C_{UMass} is an intrinsic, asymmetrical confirmation measure between top word pairs that accounts for the ordering among the top words of a topic. C_{UMass} is computed as:

$$C_{UMass} = \frac{2}{N \times (N-1)} \sum_{k=2}^{N} \sum_{l=1}^{k-1} \log \frac{P(w_k, w_l) + \epsilon}{P(w_l)},$$

where $P(w_x, w_y)$ donates the probability that the co-occurrence of the words w_x and w_y while $P(w_y)$ donate the frequency of the word w_y in the corpus. Word probabilities are estimated based on document frequencies of the original documents used for learning the topics.

Unlike C_{UMass}, C_{UCI} is an extrinsic measure based on pointwise mutual information (PMI) using an external resource such as Wikipedia to estimate the pointwise mutual information. C_{UCI} coherence is calculated by:

$$C_{UCI} = \frac{2}{N \times (N-1)} \sum_{k=2}^{N} \sum_{l=1}^{k-1} \log \frac{P(w_k, w_l) + \epsilon}{P(w_k) \times P(w_l)},$$

C_{UCI} is similar to coherence where the $sim(.,\ .)$ is replaced with the pointwise mutual information.

[7] https://nlp.stanford.edu/projects/glove/.

Table 3 Coherence, Cv, UMass, and C_{UCI} scores of topics on G' against cutting and weighting models

Model	G', w_n, w_j	Coherence	Cv	UMass	C_{UCI}
LDA		0.131	0.509	−0.457	−4.476
RTM	$w \rightarrow \infty$	0.159	0.611	−0.412	−1.078
Cutting	$w_n^P \geq 0, w_n^S \geq 0$	0.163	0.616	−0.394	−1.061
	$w_n^P \geq 5, w_n^S \geq 10$	0.163	0.619	−0.395	−1.029
	$w_j^P \geq 0.50, w_j^S \geq 0.25$	0.167	0.613	−0.402	−1.057
Weighting	$w_j^P + w_j^S$	0.165	0.620	−0.387	−1.031
	$w_j^P \times w_j^S$	*0.174*	0.610	−0.419	−1.052
	$T_{H,0}(w_j^P, w_j^S)$	0.173	0.607	−0.416	−1.083
	$T_{Y,1}(w_j^P, w_j^S)$	*0.174*	*0.622*	*−0.371*	−1.030
	$T_E(w_j^P, w_j^S)$	0.171	0.614	−0.395	−1.030
	$w_s^P + w_s^S$	0.172	0.619	−0.387	−1.020
	$w_s^P \times w_s^S$	0.170	0.614	−0.392	−1.050
	$T_{H,0}(w_s^P, w_s^S)$	0.171	0.613	−0.402	−1.082
	$T_{Y,1}(w_s^P, w_s^S)$	0.171	0.617	−0.399	−1.055
	$T_E(w_s^P, w_s^S)$	0.171	0.620	−0.387	*−1.012*

Results

The coherence scores are shown in Table 2, comparing case G'_P with statute G'_S networks, for the baselines, and both the cutting and weighting models. In the table, w_x with $x \in \{n, j, s\}$ denotes the node weight as in Eqs. 1 and 2. In the cutting model, an edge is considered only when its weight $w_x \geq w_e$; an edge between two nodes is created in RTM and trained with RTM. $w_x \geq 0$ means that all edges are considered without considering their weights, $w_x \rightarrow \infty$ means none of the links is used for the model, which is the RTM baseline described in Sect. 6.1. In the weighting model, all edges between two nodes are created and trained using RTM, but they are weighted by the different weighting methods described in Sect. 5.2.

In all cases, we can see that our RTM model given link information shows higher performance than the baselines, i.e., LDA and RTM with $w_x \rightarrow \infty$. In addition, the cutting model that uses a weight threshold to cut links improves compared to the all link inclusion $w_x \geq 0$ without weighting. This means that there is noise information in some links which give negative effect when we treat them equally. As for the settings of creating nodes either from prior cases G'_P or statute laws G'_S, we do not observe significant difference. This indicates that both citations are good sources for improving topic modeling. There is also no big difference comparing between w_n and w_j, where w_n is the common citation number between two cases, and w_j treats homophily as similarity, respectively. In the cutting model, we only use prior cases citations, $w_n \geq 5$ performs the best in G'_P; when we only use statutes, $w_n \geq 10$ performs the best in G'_S. Also, $w_j \geq 0.50$ and $w_j \geq 0.25$ performs the best for G'_P and G'_S, respectively, in the cutting model. Comparing the performance of the cutting model to that of the weighting model, we can see that the weighting model outperforms using citations from either prior cases G'_P or statute laws G'_S. This indicates the importance of giving weights

Table 4 Coherence scores over different number of topics for G'_P, G'_S and G' in the cutting model

	G'_P, $w_n \geq 5$	G'_S, $w_n \geq 10$	G', ($w_j^P \geq 0.50$, $w_j^S \geq 0.25$)
LDA $\|T\| = 200$	0.102	0.102	0.102
LDA $\|T\| = 100$	0.113	0.113	0.113
LDA $\|T\| = 50$	0.125	0.125	0.125
LDA $\|T\| = 10$	0.151	0.151	0.151
RTM($w \rightarrow \infty$) $\|T\| = 200$	0.159	0.159	0.159
RTM($w \rightarrow \infty$) $\|T\| = 100$	0.165	0.165	0.165
RTM($w \rightarrow \infty$) $\|T\| = 50$	0.169	0.169	0.169
RTM($w \rightarrow \infty$) $\|T\| = 10$	0.174	0.174	0.174
RTM $\|T\| = 200$	0.167	0.167	0.167
RTM $\|T\| = 100$	0.166	0.174	0.173
RTM $\|T\| = 50$	*0.171*	0.179	0.167
RTM $\|T\| = 10$	0.159	*0.182*	*0.180*

to distinguish good and noise links for RTM. Comparing w_j and w_s in the weighting model, we can see that w_j performs slightly better than w_s.

We also evaluated the best balance between G'_P and G'_S in the cutting model for the combined links of G'. For the cutting model, we use the best thresholds in Table 2 in order to balance the weights w_x^P and w_x^S. Table 3 shows the results. We also report the C_V, C_{UMass}, and C_{UCI} scores to confirm the consistency among different evaluation metrics. As shown in Table 3, we can see in general they are very consistent, and thus we discuss the results based on the coherence scores only. Although the performance is better than the no link model $w \rightarrow \infty$, it is similar to when using them separately. It indicates that in the cutting model, the coherence score is not necessarily improved with both types of links, though the best parameters are used as each citation information. For the weighting model, the results shown in Table 3 indicate a trend similar to the cutting model that the coherence score is not necessarily improved by using both types of links no matter w_j or w_s are used or which weighting methods are used. However, the weighting model still outperforms the cutting model. The best weighting methods are the simple product and Hamacher product (smooth product) (see Eq. (4)), which means simple weighting strategies work well if there are both prior cases and statutes (an AND operator). Using cosine similarity based on Node2vec is very consistent, as the different weighting models have close coherence scores. The results obtained with Node2vec similarity w_s can be interpreted by the fact that the kernel generalizes well the homophily networks. Meanwhile, Node2vec does not improve over simple and fuzzy weighting strategies. The best weighting strategies for Node2vec is the sum, and the interpretation can be that prior cases or statutes are enough to generate coherence topic models (an OR operator).

In addition, we investigated the impact of topic numbers for the topic models. Table 4 lists the coherence scores for the baseline LDA, RTM ($w \rightarrow \infty$), and the cutting model using different number of topics $|T|$ of 200, 100, 50, and 10 with the best weight setting for G'_P, G'_S and G', respectively. It performs the best at $|T| = 50$ when using G'_P, and

Table 5 Coherence scores of different weighting methods for the weighting model

	w_j^P	w_j^S	$w_j^P + w_j^S$	$w_j^P \times w_j^S$	w_s^P	w_s^S	$w_s^S + w_s^P$	$w_s^S \times w_s^P$
$\lvert T\rvert = 200$	0.180	0.176	0.165	0.174	0.170	0.171	0.172	0.170
$\lvert T\rvert = 100$	0.166	0.175	0.176	0.169	0.170	0.169	0.171	0.168
$\lvert T\rvert = 50$	0.172	0.171	0.164	0.167	0.175	0.170	0.170	0.174
$\lvert T\rvert = 10$	*0.190*	*0.181*	*0.192*	*0.187*	*0.177*	*0.180*	*0.175*	*0.179*

Table 6 Standard deviation for the baseline models and our best model, with respect to the coherence evaluation metric with $|T| = 10$

	LDA	***RTM*** ($w = \infty$)	***RTM*** $w = (w_j^P + w_j^S$)
Coherence	0.151	0.174	0.192
SD	0.0140	0.0143	0.0146

Table 7 Examples of topic words and their predicated topic from the best performing model (i.e., the weighting model with $w = w_j^P + w_j^S$ **and** $|T| = 10$**)**

Topic ID	Topic words	Predicated topic
0	Trademark, patent, claim, statement	Patent
1	Citizenship, resident, immigration, refuge	Immigration and refugee
2	Document, sale, product, person, subject, ship	Contract
4	Property, land, right, aboriginal	Property
5	Act, human, right, protect, nation	Human right
6	Commission, employ, applicant	Labour and employment
10	Procedure, rule, action, report, file	Procedural

$|T| = 10$ when using G'_S and G'. Though there is some variability, we can see that as the topic number decreases coherence tends to increase. Table 5 shows the coherence for the weighting model using different topic numbers $|T|$ with different weighting methods.[8] We observe the same trend as the cutting model, where the small topic number at $|T| = 10$ shows the best coherence score. In addition, the weighting model still outperforms the cutting model when changing $|T|$, and $w = w_j^P + w_j^S$ at $|T| = 10$ shows the best performance among all the models. The weighting models show the small topic number at $|T| = 10$ to be better with both w_j and w_s obtained respectively from the co-citations homophily networks and cosine similarity based on the Node2vec embedding of the homophily networks. We did not add the results for the fuzzy aggregation for different topic number as the results are similar to the simple weighting schemes. As a conclusion, *if available we recommend to use both prior cases and statutes to generate topic models; otherwise, the prior cases if available and finally the statutes. We recommend also to use small topic numbers e.g., 10.*

[8] As the three t-norm weighting methods show similar performance compared to the four simple weighting methods as shown in Table 3, we did not compare them here.

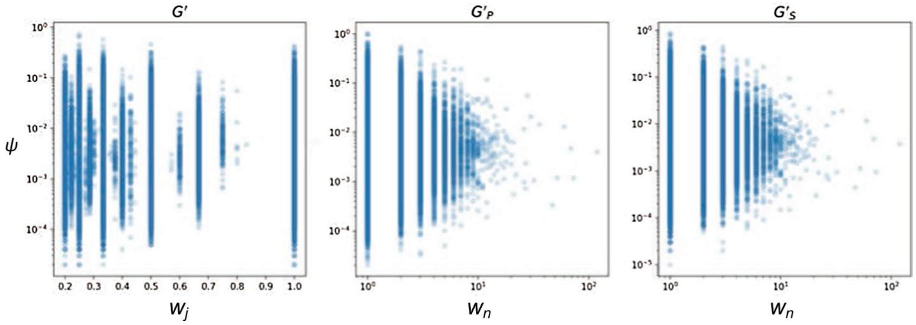

Fig. 6 Comparison of the best w_n and w_j against the topic similarity ψ

To understand the uncertainty of topic models on our data, we calculated the standard deviation.[9] Table 6 shows standard deviation results for the baseline models and our best model, with respect to the coherence evaluation metric with $|T| = 10$. Our standard deviation results show a consistent small score, meaning that the results are stable.

To understand the performance of the topic models qualitatively, we also analyzed the output topic words. Table 7 shows topic word examples from the best performing model (i.e., the weighting model with $w = w_j^P + w_j^S$ and $|T| = 10$) along with their topic IDs. We can see that these topic words are very informative and related to the areas of law in Canada.[10] Given these topic words, we can easily predict the topics of patent, immigration and refugee, contract, property, human right, labour and employment, and procedural. Through these predicted topics, we can easily understand the topics of important legal cases covered in the COLIEE dataset.

We further investigated the difference between topic similarity (ψ) and homophily (w_x) as outputs of our models. The difference for the resulting weights is shown in Fig. 6. We can see that the shapes of the topic similarity are very similar in the perspective of their best weight used w_n. However, if we investigate the most similar cases, different results may be obtained. With the cases-based topic similarity, with $\psi = 0.65$, $w_n \geq 1$, and $w_j \geq 0.05$ for G'_P the closest cases are *Bargig v. Can. (M.C.I.) (2015) case #4127* and *Barrak v. Can. (M.C.I.) (2008) #4984*. These cases are about *immigration*. To be more specific, both cases are investigating exception requests under the *Immigration and Refugee Protection Act*, and both were rejected under *insufficient humanitarian and compassionate grounds*. With the cases-based topic similarity, with $\psi = 0.48$, $w_n \geq 1$, and $w_j \geq 0.5$ for G'_S the closest cases are *Can-Am Realty Ltd. v. MNR (1993) case #4580* and *Deconinck v. MNR (1988) case #475*. These cases are about *tax*. In both cases, a taxpayer is contesting a *tax assessment*, but one case accepted the plaintiff's appeal (#475) while the other was rejected (#4580). With the statutes and case-based topic similarity, with $\psi = 0.019$ for G' the closest cases are *Diabate v. Can. (M.C.I.) (2013) case #3451* and *De Araujo v. Can. (M.C.I.) (2007) case #1276*. These cases are also about *immigration*. In

[9] We did not conduct bootstrapping, because in our context, bootstrapping may not be straightforwardly applicable. We would need a very specific sampling method that would preserve the citation graph structure and its modularity for homophily to be relevant.

[10] https://en.wikipedia.org/wiki/Law_of_Canada#Procedural_law.

both examples, the applicants asked a judicial review for a *humanitarian and compassionate* relief. One application was accepted (#3451) and the other one rejected (#1276).

Conclusion

We presented a novel analysis method for the COLIEE corpus of the law dataset. Thanks to homophily, we improve topic modeling using the citation structure even without having access to the cited content. We built networks composed of thousands of cases and references. The references belonged to two types of citations, i.e., prior cases, and statutes laws. We explored these two types of citations to investigate citation homophily among cases. We further proposed a cutting model and a weighting model for using these references to improve the RTM. Experiments indicated that the weighting model outperforms the cutting model, which significantly improves topic modeling performance. In addition, the predicted topics are very informative for legal case analysis. We publish both our data and codes online[11] for further research.

In our future work, we first intend to use a multilayer network model with Detangler (Renoust et al. 2015) to visualize the overlapping of topics based on topics content and similarity in order to evaluate the capacity of our topic models to relate similar documents. Second, we plan to combine all the different items contained in the data, such as counselors. Lastly, we plan to further investigate the extraction of new links in the dataset with our topic modeling, allowing to explore the homophily between the cited cases and laws.

We have so far applied this method in the context of legal documents and computational law only. However, there are many similar data that our framework could work for. For instance, in scientific paper co-citations, there can be different types of co-citations in different parts of the paper such as “introduction” or “related work” and “core” parts that have different functions in the paper. There also can be co-citations types such as co-citations from authors themselves, co-authors, and other researchers. These different types of co-citations can be used for homophily-based topic modeling in our framework. In news data, there can be co-references to news agencies, events, and name entities. These different types of co-references play different roles and can be used for homophily-based topic modeling in our framework as well. Therefore, this work invites further investigation of a more general method of homophily-based topic modeling that would fit a larger set of application contexts, including citation network, but could also extend other relations implying homophily.

Abbreviations

RTM: Relational topic model; DAG: Directed acyclic network; LDA: Latent Dirichlet allocation; COLIEE: Competition on legal information extraction/entailment; UMAP: Uniform manifold approximation and projection for dimension reduction; NPMI: Normalized pointwise mutual information; PMI: Pointwise mutual information.

Acknowledgements

We thank the reviewers of Complex Networks 2019 for their constructive comments.

Open Access

[11] https://figshare.com/articles/Improving_Topic_Modeling_through_Homophily_for_Legal_Documents/12408104.

Author Contributions

KA conducted most of the topic modeling experiments. CBEV performed the node2vec and fuzzy logic aggregation experiments. CC conceptualized this work and was a major contributor in writing the manuscript. BR mainly contributed to the legal network part and writing the manuscript. NO, NT, YN and HN frequently discussed this work and provided valuable feedback. All authors read and approved the final manuscript.

Funding

This work was supported by an interdisciplinary research funding from the Institute for Datability Science, Osaka University.

Availability of Data and Materials

The datasets generated during and/or analysed during the current study are available at https://figshare.com/articles/Improving_Topic_Modeling_through_Homophily_for_Legal_Documents/12408104.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Graduate of Information Science and Technology, Osaka University, Osaka, Japan. [2] Inria, IRISA, Rennes, France. [3] Institute for Datability Science, Osaka University, Osaka, Japan. [4] Graduate School of Law and Politics, Osaka University, Osaka, Japan.

Received: 3 March 2020 Accepted: 7 October 2020
Published online: 17 October 2020

References

Ashihara K, Chu C, Renoust B, Okubo N, Takemura N, Nakashima Y, Nagahara H (2019) Legal information as a complex network: Improving topic modeling through homophily. In: Proceedings of the 8th international conference on complex networks and their applications, pp 28–39

Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323(5916):892–895

Blei DM, McAuliffe JD (2007) Supervised topic models. In: Proceedings of the 20th international conference on neural information processing systems. NIPS'07, pp 121–128. Curran Associates Inc., USA. http://dl.acm.org/citation.cfm?id=2981562.2981578

Blei DM, Lafferty JD (2006) Dynamic topic models. In: Proceedings of the 23rd international conference on machine learning. ICML '06, pp 113–120. ACM, New York, NY, USA. https://doi.org/10.1145/1143844.1143859

Blei D, Lafferty J (2007) A correlated topic model of science. Ann Appl Stat 1:17–35

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. J Mach Learn Res 3:993–1022

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 10:P10008

Brochier R, Guille A, Velcin J (2020) Inductive document network embedding with topic-word attention. arXiv:2001.03369

Chang J, Blei DM (2009) Relational topic models for document networks. In: International conference on artificial intelligence and statistics, pp 81–88

Detyniecki M, Bouchon-meunier DB, Yager DR, Prade RH et al (2000) Mathematical aggregation operators and their application to video querying

Das R, Zaheer M, Dyer C (2015) Gaussian LDA for topic models with word embeddings. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (vol 1: Long Papers). Association for Computational Linguistics, Beijing, pp 795–804. https://doi.org/10.3115/v1/P15-1077. https://www.aclweb.org/anthology/P15-1077

Dieng AB, Ruiz FJR, Blei DM (2019) Topic modeling in embedding spaces. CoRR arXiv:1907.04907

Einstein A et al (1916) The foundation of the general theory of relativity. Ann Phys 49(7):769–822

Fowler JH, Johnson TR, Spriggs JF, Jeon S, Wahlbeck PJ (2007) Network analysis and the law: measuring the legal importance of precedents at the US Supreme Court. Polit Anal 15(3):324–346

Grover A, Leskovec J (2016) node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 855–864

Guillaume J-L, Latapy M (2006) Bipartite graphs as models of complex networks. Physica A 371(2):795–813

Hamacher H (1976) On logical connectives of fuzzy statements. In: Proceedings of the 3rd European meeting cybernetics and systems

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci 102(46):16569–16572

Hurley P (2005) A concise introduction to logic. Cengage Learning

Jaccard P (1901) Étude comparative de la distribution florale dans une portion des alpes et des jura. Bull Soc Vaudoise Sci Nat 37:547–579

Katz DM (2011) What is computation legal studies? University of Houston, workshop on law and computation

Katz DM, Bommarito MJ, Seaman J, Candeub A, Agichtein E (2011) Legal n-grams? A simple approach to track the 'evolution'of legal language. In: Proceedings of JURIX

Kanapala A, Pal S, Pamula R (2019) Text summarization from legal documents: a survey. Artif Intell Rev 51(3):371–402

Khanam N, Wagh RS (2017) Application of network analysis for finding relatedness among legal documents by using case citation data. i-ManagerΓÇÖs Journal on Information Technology 6(4):23
Kim RE (2013) The emergent network structure of the multilateral environmental agreement system. Glob Environ Change 23(5):980–991
Kleinberg JM (1999) Authoritative sources in a hyperlinked environment. J ACM (JACM) 46(5):604–632
Koniaris M, Anagnostopoulos I, Vassiliou Y (2017) Network analysis in the legal domain: A complex model for european union legal sources. Journal of Complex Networks 6(2):243ΓÇô268
Lee B, Lee K-M, Yang J-S (2019) Network structure reveals patterns of legal complexity in human society: the case of the constitutional legal network. PLoS ONE 14(1):0209844
Lettieri N, Faro S (2012) Computational social science and its potential impact upon law. Eur J Law Technol 3(3)
Lettieri N, Faro S, Malandrino D, Faggiano A, Vestoso M (2018) Network, visualization, analytics. A tool allowing legal scholars to experimentally investigate EU case law, 543–555
Liu Y, Niculescu-Mizil A, Gryc W (2009) Topic-link lda: joint models of topic and author community. In: Proceedings of the 26th annual international conference on machine learning. ICML '09. Association for Computing Machinery, New York, NY, USA, pp 665–672. https://doi.org/10.1145/1553374.1553460
Levi FW (1942) Finite geometrical systems: six public lectures Delivered in February, 1940, at the University of Calcutta. The University of Calcutta
Loper E, Bird S (2006) NLTK: the natural language toolkit. In: Proceedings of the annual meeting of the association for computational linguistics, pp 69–72. arXiv:0205028v1
Lu Q, Conrad JG, Al-Kofahi K, Keenan W (2011) Legal document clustering with built-in topic segmentation. In: Proceedings of the 20th ACM international conference on information and knowledge management, pp 383–392
McInnes L, Healy J, Melville J (2018) Umap: uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426
McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol 27(1):415–444
Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: Proceedings of advances in neural information processing systems, pp 3111–3119
Mimno D, Wallach HM, Talley E, Leenders M, McCallum A (2011) Optimizing semantic coherence in topic models. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, pp 262–272
Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582
Newman D, Lau JH, Grieser K, Baldwin T (2010) Automatic evaluation of topic coherence. In: Proceedings of the annual conference of the North American Chapter of the Association for Computational Linguistics, pp 100–108
Nguyen V-A, Boyd-Graber J, Resnik P, Miler K (2015) Tea party in the house: a hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (vol 1: Long Papers). Association for Computational Linguistics, Beijing, pp 1438–1448. https://doi.org/10.3115/v1/P15-1139
O'Neill J, Robin C, O'Brien L, Buitelaar P (2016) An analysis of topic modelling for legislative texts. In: CEUR Workshop Proceedings
Pelc KJ (2014) The politics of precedent in international law: a social network application. Am Polit Sci Rev 108(3):547–564
Pennington J, Socher R, Manning C (2014) GloVe: global vectors for word representation. In: Proceedings of the conference on empirical methods in natural language processing, pp 1532–1543. arXiv:1504.06654
Renoust B (2013) Analysis and visualisation of edge entanglement in multiplex networks. University of Bordeaux
Renoust B (2014) Voisinage et intrication dans les réseaux multiplexes. In: Modèles et Analyses Réseau: Approches Mathématiques et Informatiques (MARAMI) 2014, Proceedings RMPD
Renoust B, Melançon G, Viaud M-L (2014) Entanglement in multiplex networks: understanding group cohesion in homophily networks, pp 89–117
Renoust B, Melançon G, Munzner T (2015) Detangler: visual analytics for multiplex networks. In: Computer graphics forum, vol 34. Wiley, pp 321–330
Renoust B, Claver V, Baffier J-F (2017) Multiplex flows in citation networks. Appl Netw Sci 2(1):23
Röder M, Both A, Hinneburg A (2015) Exploring the space of topic coherence measures. In: Proceedings of the Eighth ACM international conference on web search and data mining, pp 399–408
Schweizer B, Sklar A (2011) Probabilistic metric spaces. Dover Publications
Wang C, Blei DM (2011) Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '11, pp 448–456. ACM, New York, NY, USA. https://doi.org/10.1145/2020408.2020480
Wang Y, Ge J, Zhou Y, Feng Y, Li C, Li Z, Zhou X, Luo B (2017) Topic model based text similarity measure for chinese judgment document. In: International conference of pioneering computer scientists, engineers and educators. Springer, pp 42–54
Yager RR (1980) On a general class of fuzzy connectives. Fuzzy Sets Syst 4(3):235–242
Yoshioka M, Kano Y, Kiyota N, Satoh K (2018) 'Overview of Japanese statute law retrieval and entailment task at coliee-2018'. In: Twelfth international workshop on Juris-informatics (JURISIN 2018)
Zhu Y, Yan X, Getoor L, Moore C (2013) Scalable text and link analysis with mixed-topic link models. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '13. Association for Computing Machinery, New York, NY, USA, pp 473–481. https://doi.org/10.1145/2487575.2487693

Publisher's Note