Automatic Latent Street Type Discovery from Web Open Data

Yihong Zhang[a], Panote Siriarya[1], Yukiko Kawai[c], Adam Jatowt[d]

[a]*Department of Multimedia Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan*
[b]*Faculty of Information and Human Science, Kyoto Institute of Technology, Kyoto, 606-8585, Japan*
[c]*Division of Frontier Informatics, Kyoto Sangyo University, Kyoto, 603-8555, Japan*
[d]*Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan*

Abstract

Street categorization is an important topic in urban planning and in various applications such as routing and environment monitoring. Typically streets are classified as commercial, residential, and industrial. However, such broad categorization is insufficient to capture the rich properties a street may possess, and often cannot be used for specific applications. Previous works have proposed several advanced street categorization systems. However, most of these systems rely on manual analysis and design, which requires significant effort. In this paper, we propose a method for automatically discovering latent street types from multi-modal Web open data. We utilize data of different modalities including microblog tweets, Foursquare venues, and Google Street View images. The model we propose considers both coherence within each modality and association between modalities. Based on the San Francisco city data, our quantitative evaluation shows superiority of the proposed method in terms of coherence and association. In qualitative analysis, we show that the street types discovered by our method correspond to the official street plan. We also show an example application in which the discovered street types are used in crime prediction.

Email address: yhzhang7@gmail.com (Yihong Zhang)
Permanent URL of this article: https://doi.org/10.1016/j.is.2020.101536

Keywords: street classification, Web open data, latent topic analysis, urban computing

1. Introduction

Street classification has been an important topic given quick urbanization in recent years, and has attracted interests from city planners as well as general public [1]. Understanding street types can be beneficial in many applications. For example, it has been shown that improving public transportation on a special type of street called activity corridors will increase support for taller and denser housing [2]. It has also been shown that by understanding the functionality and position of streets in the street network, investment can be better guided to improve the living conditions of residents [3]. There are also specific applications such as customized navigation for fuel conservation [4] and noise monitoring and management [5]. Typically, the street types are manually designed. The categorizing systems include simple ones such as commercial, residential, and industrial streets [6], as well as more complex ones, such as a two-dimension system based on Place and Link [7]. To design such a categorizing system, researchers normally need to look at a number of street examples and provide categories accordingly. Manually designing street types, however, suffers some drawbacks. First, it requires significant effort to examine street examples. Second, some street aspects may still be overlooked by manual design. Third, manually designed street types often lack generality, and may not be used in applications other than the one considered in the design. Therefore, an automatic and unsupervised way to generate street types that covers a wide range of street aspects can be quite useful. These street types may not have been assigned a name, but they would represent *latent* aspects of a street that can be used as inputs in computational street analysis. In this article, we will investigate such a task.

Given the emergence of Web open data, we now have various kinds of geographical data at hand that can describe a street available online. The prominent examples include Twitter tweets, Foursquare, and Google Street View (GSV) images. *Twitter* is a micro-blogging platform that allows users to share short messages up to 140 characters. Commonly called tweets, these messages can include a wide range of topics about personal lives, news, opinions, and may also explicitly or implicitly reflect the character of locations they are sent from [8, 9]. *Foursquare* is a location-based online service

that serves as a business directory as well as an activity record platform. Foursquare venues are a popular descriptor of streets, and have often been used for pedestrian behavior analysis [10, 11]. Finally, GSV images are 360-degree panorama street images captured mostly using cameras installed on cars. They are good representations of street outlooks. Previous works have used them for street-based studies such as street perception analysis [12]. In our prior works, we have studied street attributes such as visual and facility diversities, pleasantness, and language densities using GSV images, tweets, and Foursquare venues [13, 14, 15]. We have shown that these data sources are very useful in discovering certain street aspects. However, these aspects have been all defined manually in the previous works. In this paper, we propose a method to automatically discover latent street types using descriptive Web open data. Our method does not rely on manual design, and can be applied easily in any cities where the data such as the one described above are available. To the best of our knowledge, this is the first work that proposes to use a computational data mining method to discover latent street types.

Our method follows the approach of topic modeling in text analysis, particularly Latent Dirichlet Allocation (LDA) [16]. LDA is a hierarchical probabilistic model that describes dependencies between documents and topics, as well as ones between topics and words. The model can be learned in an unsupervised fashion, and one of the learned parameters called document-topic distribution can be seen as the latent topics of documents. If we see streets and their descriptive data as documents, and types as latent topics, it will be easy to adopt LDA for our purpose. However, the original LDA does not support very well multi-modal data, especially when the vocabulary sizes of different modalities differ significantly. Consequently, we propose to extend LDA to consider multi-modal data.

To summarize, our main contributions with this paper include:

- We propose a novel task of using computational methods to automatically discover latent street types. This is one of the earliest work that deals with this problem. We also exploit Web open data that ensures practical applicability of our approach.

- We make a novel extension to LDA to consider multi-modal data. Our extension supports independence of different modalities, and is thus expected to produce high coherence and better latent type distribution, especially when the vocabulary sizes of different modalities differ significantly.

- We run extensive quantitative evaluation and qualitative analysis to test our method. With an example target city of San Francisco, we show that our approach can detect meaningful street types and our extension to LDA further improves the results.

- We show how the street types discovered by our approach can be used in an application scenario, namely, crime prediction. The experimental results show that the prediction with discovered street types outperforms traditional prediction signals such as demographic features.

The remainder of the paper is organized as the following: in Section 2, we will review related literature. In Section 3, we will present the details of our method for discovering topics from multi-modal data. The quantitative evaluation and qualitative analysis of the results will be presented in Section 4. In Section 5, we will show how discovered latent street types can be used in a practical application, namely, crime prediction. Finally, Section 6 will offer some concluding remarks.

2. Related Work

Several researches have showed the importance of identifying street types in urban planning. McLeod and Curtis study the impact of public transport improvement for a specific street type, called activity corridor, which is defined by land use [2]. The results of their research suggest that investments in public transport infrastructure along potential activity corridors are likely to result in increased support for taller or denser housing, especially for residents living within existing greyfields. Buhrgard compares the so-called boulevardisation in Stockholm and Helsinki, based on conditions of streets of similar types present in the two cities [17]. Since Helsinki has recorded successes in converting expressways to urban boulevards, identifying similar roads in Stockholm is beneficial as it allows city planner to borrow the experiences from the former city.

Given the importance of identifying street types, a number of researches attempt to systematically classify street types, with the well-known example being residential, commercial, and industry streets. Jones *et al.* propose a two-dimension street categorization based on *Link* and *Place* properties of the street, considering a street as both a connection for movement, and a destination in its own right [3]. This results in types such as urban center, urban retail, or suburban residential. Wu on other hand, introduces

a three-dimension street categorization, based on the hierarchy, land use, and position of the street [6]. By land use, the categories include mixed-use street, traffic road, commercial street, residential street, industrial street, the historical landmark protection streets. Targeting a more specific application, Moraes *et al.* propose a street categorization with regard to urban thermoacoustic analysis [5]. The categories are established by considering temperature, traffic, and building heights. Most of current work on street categorization, however, rely on manual design. In this paper, we make a novel contribution by discovering street types automatically from large-scale geographical Web data.

There is little existing work that proposes automatic computational methods for discovering latent street types from geographical open data. There are instead a number of works that aim at discovering latent geographical attributes for urban areas, though with different geographical units. For example, Graells-Garrido *et al.* propose to discover travel patterns in a city from mobile phone network data [18]. They use non-negative matrix factorization to approximate lower rank attribute tables for users and regions, and thus indirectly discover latent geographical attributes. This technique is similarly studied in text document clustering [19]. The travel patterns, however, are scarce, and difficult to use in street-level attribute inference. Vaca *et al.* propose to use Foursquare data to discover functional areas in a city [20]. They divide a city into grids and run a clustering algorithm to find areas with similar functions. Their assumption that adjacent areas tend to have similar functions, however, would miss many real-world cases. For example, instead of conforming to one type, many residential areas in US cities have commercial streets run through them. Celikten *et al.* similarly propose using Foursquare check-in data to discover region functions [21]. Their probabilistic model considers both spatial and temporal aspects of user behavior, and can be used to match similar regions across different cities. Again, their limitation is the assumption of similarity between geographically close regions, and their method is difficult to be applied to street-level analysis. Zambrano *et al.* propose to cluster resident activities from tweets and Foursquare data, where tweets are annotated by the closest Foursquare venue [22]. Their clustering method can discover activity clusters, such as "Film" and "Stadium" during a film festival and league matches. One of the limitations of their work is to rely on pre-defined Foursquare categories, and thus the method is unable to uncover other latent attributes not defined by Foursquare, such as an attribute that distinguishes roads for pedestrians and motor vehicles. We

can also view latent type discovery as a task of learning a distributed representation of documents or streets. Shoji *et al.* follow a distribution learning approach and propose location2vec, which generates semantic representation of locations from geo-tagged tweets [23]. Apart from the restriction imposed by only using tweets, the limitation, however, is that with this approach, it will become difficult to analyze features and interpret the meaning of discovered topics.

Our method follows closely topic modeling in text analysis. Latent Dirichlet Allocation (LDA) is a well-known technique for discovering latent topics or themes in text documents [16]. It is a graphical Bayesian model based on the probability dependencies between documents and topics, and topics and words. An advantage of this model is that the topic of a document is defined as a distribution instead of a fixed category, which is commonly presented in previous works [24]. Another advantage is that after learning the model, the top words of each topic can be extracted and analyzed. The original LDA paper uses a variational inference to learn the model parameters [16]. However, it is found that approximation techniques such as collapsed Gibbs sampling can learn the model faster and more accurately [25]. Fast algorithms such as the one proposed by Porteous *et al.* can quickly learn the parameters by counting the frequency of words appearing in documents and topics. A limitation of LDA is that it provides limited support for multi-modal data. Some works have extended LDA to more than one modalities. For example, Blei *et al.* propose to extend LDA for discovering association between text and images [26]. They add to the LDA model a dependency from images to text annotations. LDA is also used by Habibian *et al.* to extract text descriptions from video sequences [40]. In their work, text embeddings are first learned through LDA, and then used for annotating images. Andrews *et al.* propose a multi-modal LDA that incorporates two data types, namely, experiential and distributional data [27]. Based on this work, Roller *et al.* propose to integrate textual, cognitive, and visual modalities with LDA [28]. However, their model is learned through pairwise associations, as with the majority of existing works. In contrast to these approaches, our method considers all modalities together at the same time. Our novel extension to LDA considers both within-document coherence, and independence of modalities.

3. Discovering Latent Street Types from Web Open Data

Our method for discovering latent street types is based on topic modeling in text analysis. More specifically, we extend LDA to incorporate multi-modal data. Originally LDA is designed to run with bag-of-words (BOW) representation of text data. Our data, on the other hand, contains images and categorical data, and thus need transformation. In this section, we first present our data pre-processing step that converts image and categorical data into BOW data. Then we briefly review LDA framework and discuss how it can be applied in our study. Last we present our extension to LDA that incorporates multi-modal data.

3.1. Data Pre-processing

The geographical Web open data used in this study include Twitter short messages called tweets, Foursquare venue categories, and GSV images. Following the latent topic modeling approaches used in text analysis and document retrieval, we consider data bound to a street as a document, and the latent types as topics. Consequently, street types and topics will be used interchangeably in the rest of this article. Since all three considered data sources have geo-coordinate information, it is trivial to assign them to streets. We will discuss data collection and assignment to streets in a case study presented in Section 5.

Assuming that data have been assigned to streets, we then convert data into BOW representation that is suitable for processing with LDA. Specifically, we need to extract *words* from data. For tweets and venues, the conversion is straightforward. Tweets are text documents that can be easily tokenized into words. The category of each Foursquare venue can be considered as a word drawn from the list of venue categories. For GSV images, we convert them to words following a common practice for image analysis [29]. First we convert them into a vector using a pre-trained deep neural network built for recognizing objects in images. Inception [30] is an example of such network that we use. The latest version of this network, Inception-v3, has 42 deep layers, including convolution and fully connected layers. In the experimental evaluation for object classification, Inception-v3 reached a top-5 error rate of 3.46%, compared to 15.3% reached by AlexNet, and 6.67% by the original Inception network. Inception takes images of any size as the input, although our GSV images are JPEG images of the size 1920×640 pixels.

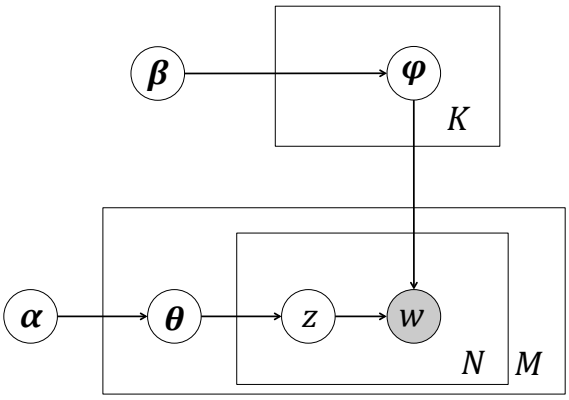


Figure 1: Graphical model for LDA

We add short code to the Inception program[1] for extracting the output of the third pooling layer, which is a vector of 2,048 dimensions representing the semantics of the input image. Each dimension represents strength of certain semantic aspects of the image. We then set a threshold δ, such that values larger than δ and smaller than $-\delta$ in the vector are picked to be the words of the image. As the result, each street is now described by a number of words from three modalities.

3.2. Latent Street Type Discovery Based on Topic Model

LDA is a widely used technique for discovering latent topics in text documents [16]. LDA uses two multinomial distributions, namely, θ and ϕ, to represent the distribution of topics in a document, and the distribution of words in a topic. Both distributions are assumed to be generated from Dirichlet distribution, controlled by hyper parameters α and β, which are non-zero numbers. The observation w, which are the words appearing in documents, is generated from θ and ϕ, using two steps. First, a topic z_{nm} indicating the topic of n-th word in document m is selected using θ. Then, based on z and ϕ, a word is selected. This generative process is depicted in Figure 1.

While direct calculation of the two distributions from the conditional probability $p(w|\theta, \phi)$ are intractable, approximation techniques such as Gibbs-

[1] https://www.tensorflow.org/tutorials/image_recognition

sampling have been developed for LDA. Particularly, a fast collapsed Gibbs sampling algorithm is commonly used to efficiently learn LDA models [31]. Through this technique, θ and ϕ are marginalized out, and z becomes the only parameter to be learned, which can be done by counting words. More specifically, let us denote the count of word w in document m and topic k as C_{wkm}, and $C_{km} = \sum_w C_{wkm}$, $C_{wk} = \sum_m C_{wkm}$. In other words, C_{km} is count of words assigned to topic k in document m, and C_{wk} is the count of word w assigned to topic k in all documents. Using Gibbs sampling, in each learning iteration:

$$p(z_{nm} = k|\mathbf{z}^{\neg nm}, \mathbf{w}, \alpha, \beta) = a_{km} b_{wk} \tag{1}$$

where

$$a_{km} = \frac{C_{km}^{\neg nm} + \alpha}{C_{m}^{\neg nm} + K\alpha} \quad b_{wk} = \frac{C_{wk}^{\neg nm} + \beta}{C_{k}^{\neg nm} + W\beta}$$

and $C^{\neg nm}$ means the count excluding the n-th word in document m. After obtaining $p(z_{nm} = k)$ for all $k \in K$, a value is sampled from this distribution and assigned to z_{nm}.

In the case of L-modal data where a document consists of L parts of words, a straightforward solution of applying LDA is to combine all L parts together. More specifically, if V_l is the vocabulary size of modality l, we make combined dictionary with size $V = \sum_l V_l$. Consequently, each document now has $N = \sum_l N_l$ words, where N_l is the number of words in the document for modality l. The distinction of modality is invisible in this solution, which we call combined LDA (cLDA).

While being simple, this approach has some drawbacks. An immediate problem is that modality of different vocabulary sizes will have imbalanced influence on the result. As we will show in the case study in Section 4, we often have tweets with vocabulary of thousands of words and Foursquare venues of hundreds of categories. Since they are combined together indifferently, and each update of parameter is taking into account all other parameters, the modality of larger vocabulary size will have more chance to be trained. In other words, modality of different sizes will have different influence on the topic-word distribution ϕ. To mitigate this problem, we propose Weighted Multi-modal LDA (WM-LDA), which we now describe.

3.3. Weighted Multi-modal LDA

One way to deal with the imbalanced influence from different modalities is to use a separate word-topic distribution for each of the modality. This

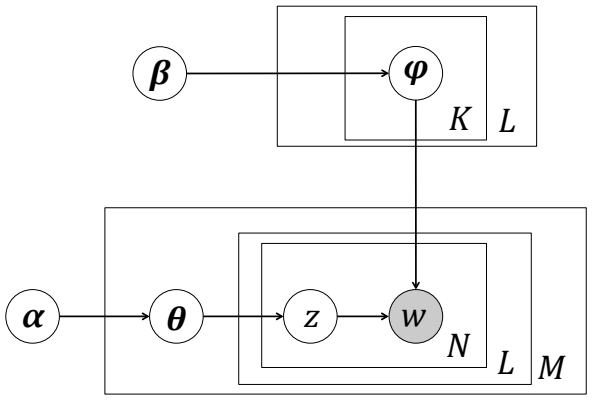


Figure 2: Graphical model for WM-LDA

concept is illustrated in Figure 2. In this graphical model, we have a ϕ for each of L modalities, and the counting of words is also based on modalities instead of the whole document. A problem of this solution is that there is no interaction between modalities. Even though each modality is trained by itself, and may have a higher coherence, the topics they represent may not associate to each other. To tackle this problem we propose weighted multi-modal LDA (WM-LDA) that considers both the individuality of modalities, as well as the interaction among them.

One way to interpret the factors a and b for determining $p(z_{nm})$, shown in Equation (1), is to consider a as the influence of document coherence, and b as the vocabulary consistency across different topics. With an additional weighting parameter λ, WM-LDA uses an alternative calculation of a and b so that a conveys the coherence of the entire document, and b conveys vocabulary consistency of individual modality. More specifically, let us similarly define C_{wklm} as the count of word w in topic k, modality l in document m. We extend $p(z_{nm})$ in Equation (1) as $p(z_{nlm})$, where $l \in L$ is the current modality, and calculate a_{klm} and b_{wlk} as:

$$a_{klm} = \sum_{l} \lambda_l \frac{C_{klm}^{\neg nlm} + \alpha}{C_{lm}^{\neg nlm} + K\alpha} \quad b_{wlk} = \frac{C_{wlk}^{\neg nlm} + \beta}{C_{lk}^{\neg nlm} + W_l\beta} \tag{2}$$

where W_l is the vocabulary size of modality l. Given this formula, the interaction between modalities is thus achieved by computing a considering all

modalities, which then defines each $p(z_{nlm})$. Here λ_l controls the influence of each modality on the distribution. Normally we set the same value for all λ_l, $l \in L$, unless we have prior knowledge of which modality is more important. Having the same λ_l for all modalities will give a balanced influence to each modality, thus mitigating the problem stated in the previous section. Algorithm 1 shows a fast procedure for updating z_{nlm}. Note that we precompute p_{klm}, components of a_{klm}, when starting to proceed document m and modality l (line 3), and avoid calculating them in the inner-most loop of the algorithm, which can significantly reduce the computation time.

Algorithm 1 Updating z with WM-LDA

```
1:  for each document m, modality l do
2:      for each topic k do
3:          p_klm ← λ_l (C^¬nlm_klm + α) / (C^¬nlm_lm + Kα)
4:      end for
5:      for each z_nlm of n-th word do
6:          remove z_nlm from counters
7:          for each topic k do
8:              calculate a_klm using Equation (2) and p_klm
9:              calculate b_wlk using Equation (2)
10:             p(z_nlm = k) ← a_klm b_wlk
11:             sample z_nlm ~ p(z_nlm)
12:         end for
13:     end for
14: end for
```

Based on z, the document-topic distribution θ and topic-word distribution ϕ are obtained with the following:

$$\theta_{km} = \sum_l \lambda_l \frac{C_{klm} + \alpha}{C_{lm} + K\alpha} \qquad \phi_{lwk} = \frac{C_{wlk} + \beta}{C_{lk} + W_l\beta}$$

Top words of each modality that are most likely to appear in a topic can thus be obtained by ranking ϕ values.

4. A Case Study: San Francisco

In the remaining studies, we will use San Francisco as our target. San Francisco is one of the largest cities in the US[2], with well-known regions such as Financial District, Union Square, and Fisherman's Wharf. We choose San Francisco because of the diverse street types it contains. It is also highly popular among tourists for whom street-level information can be very useful.

4.1. Data Collection

We first identify street segments in San Francisco using OSM data. OSM contained over 4 billion nodes over the world, where each node represents a geographical point of interest (POI)[3]. A street segment is defined in OSM as a series of points (particularly, it can have a high number of points if the street segment is not straight). OSM also provides the starting coordinates, ending coordinates, length and the name of the street that the segment belongs to. We collect all street segments and their data in San Francisco city using OSM public API[4]. In total, the collected data contains 252,537 street segments. We then define the distance between a geographical point and a street segment as the shortest distance between the point and the segment. If the segment is straight, the distance is the length of the perpendicular line from the point to the segment. We use the QGIS software[5] to perform this calculation. This measurement of distance is our basis for assigning geographical data to street segments.

Tweets are collected by monitoring live stream using Twitter Filter API[6] from May 2016 to April 2017, resulting in 751,628 geo-tagged tweets. We then collect Foursquare data using Venue Search API[7]. We gather information for 41,515 venues in SF. Tweets and venues are then assigned to streets that are within 20 meters from them. Multiple streets can be assigned to one data point. We collect GSV images using Google Street View Image API[8]

[2]https://www.census.gov/quickfacts/fact/table/sanfranciscocountycalifornia,US/
[3]https://wiki.openstreetmap.org/wiki/Stats
[4]https://wiki.openstreetmap.org/wiki/API
[5]https://www.qgis.org/en/site/
[6]https://developer.twitter.com/en/docs/tweets/filter-realtime/overview
[7]https://developer.foursquare.com/docs/venues/search
[8]https://developers.google.com/maps/documentation/streetview/

within the specified bounding box covering the San Francisco city[9]. A total of 98,874 panorama view images are collected. We then assign an image to a street segment if it is within 10 meters from the segment. An image can be assigned to multiple street segments.

The tweets, Foursquare venues, and GSV images are then converted to BOW representations following the method discussed in Section 3.1. The vocabulary sizes of these three modalities are 8,755, 528, and 1,237, respectively. Note that in order to preserve meaningful representation, we only keep street segments that contain at least three words in each of the modalities in our dataset.

4.2. Baseline Methods

In the following experiments, we compare our WM-LDA model with two baselines. The first is combined LDA (cLDA), which has been described in Section 3.2. For this baseline, the content of three modalities are concatenated into one document, before applying the standard LDA.

The second baseline is the method proposed by Roller and Im Walde [28], called 3D-LDA. This method is an extension to the mLDA, which allows the addition of a second modality to the standard LDA model [27]. Similarly, 3D-LDA allows two additional modalities, provided that the associations between the primary and additional modalities are established. Roller and Walde use manual annotation to create association between two modalities, for example, a word and a visual clue. In this paper, we create association between modalities based on the co-occurrence of the words. Using tweet words as the primary modality, we calculate the term frequency - inverse document frequency (TFIDF) score of Foursquare venues and GSV image words. More specifically, for each word w we select street segments S that contain w, and the score of a venue v is calculated as

$$\text{TFIDF}(v, S) = tf(v, S) \cdot \log \frac{|D|}{|\{d \in D : v \in d\}|}$$

where $tf(v, S)$ is the number of times v appears in S, and D is the set of all street segments. In this way, the venue more likely to appear together with the word will have a high score. We calculated the TFIDF scores for

[9]the bounding box is defined by a pair of coordinates (-122.523057, 37.813163) and (-122.354814, 37.708275)

all venues, and the venue that has the highest score will be associated with the word w. We create association between words and GSV image words the same way. After establishing the associations, we can now run the 3D-LDA using tweet words as the primary modality. We use the implementation made available by the author. Running 3D-LDA will generate the document-topic distribution *theta* and three ϕs for three modalities,.

4.3. Quantitative Evaluation

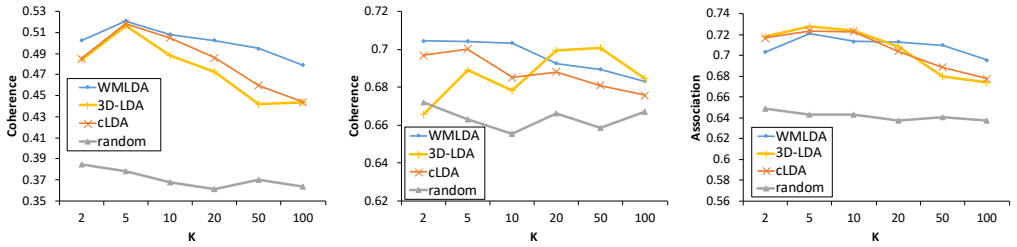
We use automatic methods to quantitatively evaluate the quality of discovered topics. Specifically, we measure the *coherence* of topic words, and the *association* between different modalities. In recent years it has been a common practice to evaluate topic models through a reference source [33, 34]. For example, Newman *et al.* propose to measure topic coherence based on the position of words in WordNet, a linked dictionary of English words [33]. In this paper, we use a reference source called GloVe, which recently has shown considerable impact in text analysis [35]. GloVe is an algorithm for learning distributed representations of words. The authors have released several versions of pre-trained word representations online[10], and we use the one trained on two billion tweets that contains 1.2 million words, each of which is represented as a vector indicating semantics of the word in a latent dimension. Words with similar meaning or context (e.g., *coffee* and *tea*) will be represented as similar vectors.

To measure the coherence of a modality of a topic, we calculate the average pairwise cosine similarity of top T words for each topic. In other words, the coherence of top T words $\mathbf{w}$ of a modality is calculated as:

$$coherence(\mathbf{w}) = \frac{1}{T(T-1)/2} \sum_{i=1}^{T-1} \sum_{j=i+1}^{T} cosSim(w_i, w_j)$$

Normally we set $T = 10$. The coherence of a model is subsequently calculated as the mean coherence of all topics. This is straightforward for tweet data. For Foursquare venues data, we can use the venue name as words matched in GloVe, but there are some cases where venue names are composed of multiple words (e.g., *seafood restaurant*. In such cases, we use the last word in the venue name, which in general defines the function and

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



(a) Coherence of top tweet words (b) Coherence of top venues (c) Association between tweets and venues

Figure 3: Quantitative evaluation results for latent street type discovery

meaning of the venue. For GSV image words, however, we could not find a reference source, and will leave it for future work.

To measure the association between modalities, we again exploit GloVe-based similarity, which is available only for tweets and Foursquare venues. Thus to calculate the association between tweets and venues, for each topic, we first take a word from tweets, and find the most similar venue, and use it as the association of this word to venues. In other words, the association of a tweet word to venues $\mathbf{v}$ is calculated as:

$$assoc(w, \mathbf{v}) = \arg\max_i cosSim(w, v_i)$$

The association of tweets and venues of this topic is then calculated as the average association of top T tweet words to venues. The association between tweets and venues of a model is thus the average association of all topics.

Based on the data collected, we run experiments to quantitatively evaluate latent street types discovered by different models. We tested different numbers of topics, $K = \{2, 5, 10, 20, 50, 100\}$. For each evaluation, we run the experiment ten times, and take the average coherence and association results. Figure 3 shows the evaluation results. In addition to cLDA and 3D-LDA, we also evaluate a random baseline, which randomly generates the topic-word distribution ϕ. The random method achieves a low tweet coherence because tweet words are of a wide range of semantics. But for venue coherence, the random method performs relatively well, because most venue words are related to locations or businesses.

Comparing WM-LDA with cLDA and 3D-LDA, we can see that WM-LDA tends to provide more coherent topics in terms of tweet words and

venues, in most cases. Particularly for venues, even though the room to improve is smaller, considering the high coherence achieved by the random method, WM-LDA still has a large improvement over cLDA. The reason seems to be that for cLDA, the venue dictionary is much smaller than tweet word dictionary, and according to the imbalanced influence we explained, venues in cLDA could not be properly learned. However, by using the same weighting in WM-LDA, venues can be learned equally well as the tweets, thus achieving much higher coherence. At the same time, tweet words in WM-LDA also achieve better coherence than cLDA, albeit smaller improvement, because they are excluded from the influence of other modalities. In regards to the association, we can see that WM-LDA achieves about the same level of association as cLDA, and in some cases higher, especially when K is larger. 3D-LDA performs similar to cLDA, except for venue coherence with higher K, for which it achieves better results. This is perhaps because of the way the tweet words are mapped to venues, which makes more popular venues to appear more frequently in the model, resulting in a smaller vocabulary, and consequently higher coherence. But this also causes association between tweet words and venues to weaken, and as we can see, it achieves lower association results with higher K.

4.4. Qualitative Analysis

In this section, we try to examine the quality of discovered latent street types through manual analysis. We first compare discovered street types with an official street plan. Then we check a number of typical streets by searching relevant information on the Web.

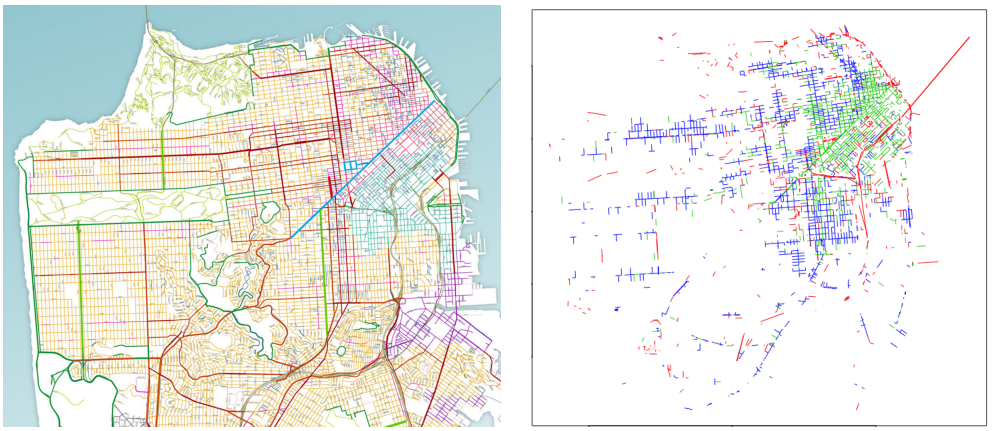
4.4.1. Comparison with Street Plan

In this qualitative analysis, we compare the street plan of San Francisco with the discovered street types. We obtain the official street plan from SF Better Streets website[11]. Shown in Figure 4 (a), this street plan map is defined by land use context and transportation characteristics, and contains street types like neighborhood commercial streets (purple), neighborhood residential streets (yellow), and park edges (dark green)[12].

Although we do not have the street-by-street labels used for producing this map in order to run a quantitative analysis, we can nevertheless compare

[11] https://www.sfbetterstreets.org/design-guidelines/street-types/

[12] Full color codes can be viewed on the above website



(a) Street Plan (b) WM-LDA Street Types K=3

(c) cLDA residential (d) cLDA commercial (e) 3D-LDA residential

(f) 3D-LDA commercial (g) WM-LDA residential (h) WM-LDA commercial

Figure 4: Comparison of the street plan with discovered street types

our results with this map visually. By comparing with this plan map, we can see whether automatically discovered types capture types defined in realistic street plan. For this study, we set K=3 for all models. The discovered street types using WM-LDA is shown in Figure 4 (b), from which we can see roughly residential streets, commercial streets, and highways, corresponding to color, green, blue, and red, respectively. In Figure 4 (c) to (h), we separate the residential streets and commercial streets discovered by different models. Compared to the commercial streets defined in the plan, which are concentrated in the east and northeast corner, colored in purple and light blue, we can see from (h) that WM-LDA capture more fully the commercial areas. In (d) and (f), we can see cLDA and 3D-LDA are not very capable of capturing the commercial area, and they sometimes mix commercial with highways. In (c) and (e) we also see that these two models assigned most of the streets as the residential streets, showing their weakness in distinguishing different types of streets. On the other hand, in (b), (g), and (h), we see a much clearer separation between residential streets and commercial streets, produced by WM-LDA.

Please be reminded that here we set $K = 3$ so that it is easy to compare with street plan. However, the latent topic models are more powerful when K is set to a higher number, with which detailed streets types that are overlooked in street plans can be represented.

4.4.2. Cardinal Type Streets

In this qualitative analysis, we would like to find out if our method produce coherent street types through examples. We first pick some typical streets from the results. We determine if a street is typical of a type based on entropy [32]. Recall that the parameter θ is the document-topic distribution that shows probability of a street belonging to each type. The entropy is calculated as $-\sum_i P_i \log P_i$. We calculate the entropy for each street, and the lower the entropy, the more the street is considered pure of a type. We run WM-LDA with $k = 50$ to generate θ. From the low entropy streets we pick two streets that have different cardinal types for qualitative analysis. The top tweets, venues, and images for the type the street is associated with are shown in Table 1. The images are picked so that they have the highest sum for the top image words. We also compare these data with the information we found in search engines about the street.

The first street is a special cultural street with restaurants and shops reflecting a specific culture group. Identified as Japantown by search engines,

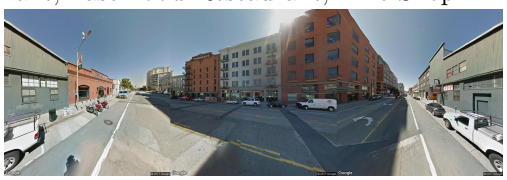
Typical Street 1	Typical Street 2
Japantown, a Japanese-themed street	Civic Center, surrounded by public facilities
Tweets: Ramen, Sushi, Regency, Ballroom, ramen, sushi, Hinodeya, Legend, Madrone, Ink	Tweets: Center/UN, lol, Mr., Jewish, Contemporary, break, Lunch, Bakehouse, Holmes, shift
Venue: Sushi Restaurant, Japanese Restaurant, Korean Restaurant, General Entertainment, Ramen Restaurant, Karaoke Bar, Spa, Gift Shop, Hardware Store, Chiropractor	Venue: Government Building, Bike Rental / Bike Share, Intersection, Gym / Fitness Center, General Travel, Convenience Store, General College & University, Sushi Restaurant, Fast Food Restaurant, Bike Shop
	

Table 1: Detected Typical Streets

we can see that the associated tweets and venues reflect this information, with both referring to Japanese food. The image also shows a residential-business area with Asian food outlets and grocery stores. The second street contains a government facilities, and we see this is reflected in tweets and venues. The images show an official building in a less commercialized area.While there are more potential analyses can be done, for now we conclude that our method is able to discover typical coherent streets.

5. Application: Crime Prediction with Latent Street Types

The discovered latent types can be considered as a novel vector representation of a street, and can thus be useful in many applications of computational street analysis, from customized routing to housing price estimation. In this section, we study one of such applications, namely, street crime prediction. Crime is a critical social phenomenon, and crime prediction is an important research topic beneficial for both government and residents [36]. In existing literature, crime is often associated with demographic data, such as income, education level, and employment rate [37]. While largely overlooked in existing literature, it is recognized that street outlooks can have an acute influence on crime rate, which leads to theories such as "broken window" [38]. In this study, we will show that the latent street types discovered from street outlook images, venue categories, and conversations, can be effectively

used in predicting crime.

5.1. Data and Evaluation Setup

We still use San Francisco as our target city. In addition to the descriptive Web data, we also collect crime data for training and testing prediction models. Particularly, we collect data on assault and burglary, the two most common offenses against person and property, from the "Police Department Incidents Data Set"[13] provided by DataSF[14]. Unlike speeding or domestic violence, these crimes are considered highly related to the characteristics of the street. The dataset dates range from Jan 1, 2003 to Jun 25, 2017. We assign the crimes to streets in the similar way as was with the other data. For the valid streets, there are a total of 36,741 assaults and 21,335 burglaries. The distribution statistics about crimes per edge are shown in Table 2.

Table 2: Distributions of the crimes per edge

	min	median	mean	max
assault	0	10	14.4	131
burglary	0	6	8.3	73

We compare latent street types discovered by cLDA, 3D-LDA, and WM-LDA. For all models, we set four K values, 5, 10, 20, and 50. Accordingly, the latent type vectors obtained from θ to represent streets are of 5, 10, 20, and 50 dimensions. We also compare latent street types against demographic features, the common predictor in existing literature. We use demographic features proposed in [39], which include the following demographic features effective for crime prediction: total population, population density, poverty, disadvantage index, residential stability, ethnic diversity, race distribution. We obtain these demographic information from a government website[15], and assign them to streets. Furthermore, we consider the prediction made by a spatial aggregation method, called kNN (k nearest neighbors). This prediction does not involved demographic or any other features, and the crime number in one street is predicted as the average of crime numbers in nearest k streets. We set $k = 3$ as it generally produce good predictions.

[13]https://catalog.data.gov/dataset/sfpd-incidents-from-1-january-2014
[14]https://datasf.org/opendata/
[15]https://www.census.gov/

We use 10-fold cross validation, which takes 90 percent data for training and 10 percent data for testing. We use Support Vector Regression (SVR) as our learning model. We use the SVR implemented in the e1071 R package[16], and use default parameters and the radial kernel. We also consider neural network models. However, it has been previously shown that simple neural networks with fully connected layers do not perform as well as SVR, and designing an advanced neural network is beyond the scope of this paper. After training the model, we compare prediction and testing data and measure Root Mean Square Error (RMSE) and Spearman Rank Correlation, two common metrics for measuring prediction accuracy. RMSE measures the accuracy on the absolute crime number. Rank Correlation measures the ranking similarity of the prediction and actual crime number, which is useful when the relative crime rate is important.

5.2. Evaluation Results

The experimental evaluation results are shown in Table 3. We can see from the results that for both assault and burglary, WM-LDA with K=50 achieves the lowest error compared with other approaches, especially compared to demographic features It means that the proposed WM-LDA street types can be used to effectively represent streets, better than demographic features. On the other hand, cLDA and 3D-LDA are unable to achieve better results than demographic features for assault. For burglary they are better, perhaps due to burglary being more associated to street types than assault. WM-LDA nevertheless achieves better results than cLDA and 3D-LDA in all the tests. The spatial model kNN, though generates larger errors than other methods, produces better rank correlations, most likely due to the spatial correlation of crimes.

We are also interested to know whether WM-LDA street types complement the demographic features and the spatial model, and can improve the results they achieve. Table 4 shows the accuracy results of combining demographic features, kNN predictions and WM-LDA street types for prediction. Comparing with Table 3, we can see that the combined features achieve better results than using kNN prediction or demographic features individually. We thus conclude that WM-LDA street types can be used as complement to the spatial model and demographic features in prediction tasks, although

[16]https://cran.r-project.org/web/packages/e1071/index.html

	assault		burglary	
	RMSE	Rank Corr	RMSE	Rank Corr
kNN	16.280	**0.348**	9.931	**0.255**
demographic	15.941	0.234	9.238	0.110
cLDA K=5	16.207	0.142	9.245	0.117
cLDA K=10	16.182	0.162	9.273	0.131
cLDA K=20	16.228	0.155	9.249	0.144
cLDA K=50	15.960	0.206	9.149	0.184
3D-LDA K=5	16.251	0.122	9.215	0.123
3D-LDA K=10	16.070	0.156	9.258	0.111
3D-LDA K=20	16.165	0.156	9.321	0.075
3D-LDA K=50	15.877	0.205	9.177	0.137
WMLDA K=5	16.007	0.210	9.261	0.131
WMLDA K=10	16.055	0.201	9.224	0.148
WMLDA K=20	15.754	0.262	9.150	0.192
WMLDA K=50	**15.628**	0.256	**9.115**	0.179

Table 3: Prediction accuracy comparison of kNN, demographic features and latent street types

we cannot claim that increasing the number of types can always bring better prediction.

	assault		buglary	
	RMSE	Rank Corr	RMSE	Rank Corr
kNN + demo. + WMLDA K=5	15.054	0.365	9.030	0.220
kNN + demo. + WMLDA K=10	15.253	0.334	8.918	**0.274**
kNN + demo. + WMLDA K=20	14.899	**0.379**	**8.899**	0.265
kNN + demo. + WMLDA K=50	**14.797**	0.369	8.934	0.255

Table 4: Prediction accuracy of combining kNN predictions, demographic features and latent street types discovered by WMLDA

6. Conclusion

In this paper, we tackle the problem of automatically discovering latent street types using Web open data. Our learning method extends LDA with

multi-modal capacities. We run experiments to test the method using real-world data of San Francisco. In quantitative evaluation, we show that types discovered by our method are superior in terms of coherence and association between modalities. In qualitative analysis, we show that the discovered street types correspond closely to the types defined in the street plan. One reason that our modified version of LDA performs better than the original version is that modalities are considered separately before they are combined, and thus discovered street types are better representations taking information equally from all modalities. We also demonstrate an example application of crime prediction that can be improved by incorporating detected latent street types.

We need to note that our approach has some limitations. First, learning the model is an iterative process that requires time, making it unsuitable as an online method. Second, this approach cannot be used for streets with little or no data. Finally, even though we can make rough guesses through the top topic words and visualizations, confidently claiming the meaning of a discovered street type still needs human effort, ideally from domain experts. Nevertheless, our method of discovering latent street types can lead to many interesting street-based studies in the future. We plan to investigate the impact of discovered street types in more applications such as pedestrian flow prediction and land price estimation. We are also interested in designing an online algorithm for learning the model that can be used in a real-time routing system.

References

[1] P. Jones, R. Thoreau, Involving the public in redesigning urban street layouts in the uk, in: TRB Urban Streets Symposium, Seattle, 2007 (2007).

[2] S. McLeod, C. Curtis, Contested urban streets: Place, traffic and governance conflicts of potential activity corridors, Cities (2018).

[3] P. Jones, N. Boujenko, S. Marshall, A comprehensive approach to planning and designing urban streets, in: Proceedings of European Transport Conference 2008, 2008 (2008).

[4] Larsson, Hanna, Brundellfrei, Karin, Optimizing route choice for lowest fuel consumption : Potential effects of a new driver support tool,

Transportation Research Part C Emerging Technologies 14 (6) (2006) 369–383 (2006).

[5] E. L. Moraes, I. M. Franco, M. V. Silva, I. A. Rocha, D. F. Pinheiro, M. P. Freitas, Categorization of street types in urban thermoacoustic analysis, Journal of the Acoustical Society of America 133 (5) (2013) 3453 (2013).

[6] X. H. Wu, Study on street functional classification of small town, in: International Workshop on Energy & Environment in the Development of Sustainable Asphalt Pavements, 2010 (2010).

[7] P. Jones, S. Marshall, N. Boujenko, Creating more people-friendly urban streets through 'link and place'street planning and design, IATSS research 32 (1) (2008) 14–25 (2008).

[8] S. Wu, J. M. Hofman, W. A. Mason, D. J. Watts, Who says what to whom on Twitter, in: Proceedings of the 20th International World Wide Web Conference, 2011, pp. 705–714 (2011).

[9] Y. Zhang, C. Szabo, Q. Z. Sheng, Sense and focus: Towards effective location inference and event detection on twitter, in: Proceedings of the 16th International Conference on Web Information Systems Engineering Part I, 2015, pp. 463–477 (2015).

[10] S. Wakamiya, H. Kawasaki, Y. Kawai, A. Jatowt, E. Aramaki, T. Akiyama, Lets not stare at smartphones while walking: memorable route recommendation by detecting effective landmarks, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016, pp. 1136–1146 (2016).

[11] A. Noulas, C. Mascolo, E. Frías-Martínez, Exploiting foursquare and cellular data to infer user activity in urban environments, in: Proceedings of the 14th International Conference on Mobile Data Management, 2013, pp. 167–176 (2013).

[12] M. De Nadai, R. L. Vieriu, G. Zen, S. Dragicevic, N. Naik, M. Caraviello, C. A. Hidalgo, N. Sebe, B. Lepri, Are safer looking neighborhoods more lively?: A multimodal investigation into urban life, in: Proceedings of the 2016 ACM on Multimedia Conference, ACM, 2016, pp. 1127–1135 (2016).

[13] S. Wakamiya, P. Siriaraya, Y. Zhang, Y. Kawai, E. Aramaki, A. Jatowt, Pleasant route suggestion based on color and object rates, in: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, ACM, 2019, pp. 786–789 (2019).

[14] Y. Wang, Y. Zhang, P. Siriaraya, Y. Kawai, A. Jatowt, Language density driven route navigation system for pedestrians based on twitter data, in: Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion, ACM, 2018, p. 26 (2018).

[15] Y. Zhang, P. Siriaraya, Y. Wang, S. Wakamiya, Y. Kawai, A. Jatowt, Walking down a different path: Route recommendation based on visual and facility based diversity, in: Companion of the The Web Conference 2018 on The Web Conference 2018, International World Wide Web Conferences Steering Committee, 2018, pp. 171–174 (2018).

[16] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, Journal of Machine Learning Research 3 (Jan) (2003) 993–1022 (2003).

[17] S. Buhrgard, From expressways to boulevards : The compared conditions for boulevardisation in stockholm and helsinki, Master's thesis, KTH, Urban Planning and Environment (2015).

[18] E. Graells-Garrido, D. Caro, D. Parra, Toward finding latent cities, in: Joint Proceedings of the ACM IUI 2018 Workshops co-located with the 23rd ACM Conference on Intelligent User Interfaces, 2018 (2018).

[19] W. Xu, X. Liu, Y. Gong, Document clustering based on non-negative matrix factorization, in: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, ACM, 2003, pp. 267–273 (2003).

[20] C. K. Vaca, D. Quercia, F. Bonchi, P. Fraternali, Taxonomy-based discovery and annotation of functional areas in the city, in: Ninth International AAAI Conference on Web and Social Media, 2015 (2015).

[21] E. Çelikten, G. Le Falher, M. Mathioudakis, Modeling urban behavior by mining geotagged social data, IEEE Transactions on Big Data 3 (2) (2017) 220–233 (2017).

[22] O. Zambrano, A. Avendaño, W. Yanez, C. Vaca, Milano, città d'arte: Urban residents preferences clusters from tweets, in: 2017 Fourth International Conference on eDemocracy & eGovernment (ICEDEG), IEEE, 2017, pp. 210–215 (2017).

[23] Y. Shoji, K. Takahashi, M. J. Dürst, Y. Yamamoto, H. Ohshima, Location2vec: Generating distributed representation of location by using geo-tagged microblog posts, in: International Conference on Social Informatics, Springer, 2018, pp. 261–270 (2018).

[24] D. Gildea, T. Hofmann, Topic-based language models using em, in: Sixth European Conference on Speech Communication and Technology, 1999 (1999).

[25] T. L. Griffiths, M. Steyvers, Finding scientific topics, Proceedings of the National academy of Sciences 101 (suppl 1) (2004) 5228–5235 (2004).

[26] K. Barnard, P. Duygulu, D. Forsyth, N. d. Freitas, D. M. Blei, M. I. Jordan, Matching words and pictures, Journal of machine learning research 3 (Feb) (2003) 1107–1135 (2003).

[27] M. Andrews, G. Vigliocco, D. Vinson, Integrating experiential and distributional data to learn semantic representations., Psychological review 116 (3) (2009) 463 (2009).

[28] S. Roller, S. S. Im Walde, A multimodal lda model integrating textual, cognitive and visual modalities, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 2013, pp. 1146–1157 (2013).

[29] A. Ramisa, F. Yan, F. Moreno-Noguer, K. Mikolajczyk, Breakingnews: Article annotation by image and text processing, IEEE transactions on pattern analysis and machine intelligence (2017).

[30] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826 (2016).

[31] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, M. Welling, Fast collapsed gibbs sampling for latent dirichlet allocation, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2008, pp. 569–577 (2008).

[32] A. Rosenberg, J. Hirschberg, V-measure: A conditional entropy-based external cluster evaluation measure, in: Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL), 2007 (2007).

[33] D. Newman, J. H. Lau, K. Grieser, T. Baldwin, Automatic evaluation of topic coherence, in: Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2010, pp. 100–108 (2010).

[34] M. Röder, A. Both, A. Hinneburg, Exploring the space of topic coherence measures, in: Proceedings of the eighth ACM international conference on Web search and data mining, ACM, 2015, pp. 399–408 (2015).

[35] J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1532–1543 (2014).

[36] C. Graif, A. S. Gladfelter, S. A. Matthews, Urban poverty and neighborhood effects on crime: Incorporating spatial and network perspectives, Sociology compass 8 (9) (2014) 1140–1155 (2014).

[37] L. G. Alves, H. V. Ribeiro, F. A. Rodrigues, Crime prediction through urban metrics and statistical learning, Physica A: Statistical Mechanics and its Applications 505 (2018) 435–443 (2018).

[38] J. Q. Wilson, G. L. Kelling, Broken windows, Atlantic monthly 249 (3) (1982) 29–38 (1982).

[39] H. Wang, D. Kifer, C. Graif, Z. Li, Crime rate inference with big data, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 635–644 (2016).

[40] A. Habibian, T. Mensink, C. Snoek, Discovering semantic vocabularies for cross-media retrieval, in: Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, ACM, 2015, pp.131–138 (2015).