City Disaster Susceptibility Comparisons Using Weighted Bipartite Graphs

Wubai Zhou, Florida International University, Email: wzhou005@cs.fiu.edu
and
Chao Shen, Florida International University, Email: cshen001@cs.fiu.edu
and
Tao Li, Florida International University, Email: taoli@cs.fiu.edu
and
Shu-Ching Chen, Florida International University, Email: chens@cs.fiu.edu
and
Ning Xie, Florida International University, Email: nxie@cs.fiu.edu
and
Jinpeng Wei, University of North Carolina at Charlotte, Email: jwei8@uncc.edu

Metropolises offer ample employment opportunities, convenient facilities and a wide array of entertainment options. However large cities are also more vulnerable to natural disasters, which have caused widespread destructions, claimed thousands of lives and left havoc for the survivors. Knowing which city is less susceptible to natural disasters is thus one of the most critical questions one faces when making decisions on travelling or job and business relocation. In this work, we propose a bipartite-graph based framework to compare the impacts of disasters on two cities by answering different queries using textual documents collected online. Besides intuitive simple comparisons using statistics, our system also generates textual comparative summaries to better describe the differences between the two cities in term of safety. Although a number of online services provide disaster events statistic information for cities, our framework compares the impacts of disasters on cities in a more straightforward and comprehensive way.

Keywords: Topic modelling, document summarization, disaster management, bipartite graphs.

1. INTRODUCTION

People are attracted to metropolis due to ample employment opportunities, convenient facilities and a wide array of entertainment options. However large cities can also be vulnerable to natural disasters, which tend to cause more damage in densely populated areas. For example, 80% of New Orleans was flooded in Hurricane Katrina 2005; New York City was seriously affected by Hurricane Irene and Hurricane Sandy in 2011 and 2012; The winter storm 2011 left 21 inches of snow in Chicago; shakes in the two major cities on the west coast of U.S., Los Angeles and San Francisco remind of the probability of an earthquake attack. Therefore, before making decisions on travelling or job and business relocation, one of the most critical questions people face is: which city is safer?

For city comparisons, a number of online services[1] provide statistic data about various aspects of cities or neighborhoods like crime rates, races, living expenses and house prices etc. However, to the best of our knowledge, none of them considers the impacts of natural disasters. On the other hand, although current and historical disaster data can be easily obtained (e.g., through National

[1]Examples include: `http://www.neighborhoodscout.com`,
`http://www.numbeo.com/`, and `http://www.city-data.com`

Hurricane Center[2] for hurricanes and U.S. geological survey[3] for earthquakes), information about how a disaster event affects a specific city is not readily available. In most cases, data on impacts of disasters on cities is stored by different government departments or organizations, so that extra effort is required to collect data or/and conduct integration into a unified database to support comparisons among different cities. Moreover, although statistics like damages and fatalities provide direct evidences for the safety comparison, it is still challenging to obtain an overview on historically how severe a city was affected by disasters, since some types of impacts from disasters – e.g. road close caused by a hurricane – are not reflected by the statistics in a straightforward way.

In this paper, we tackle this problem by aggregating easily acquired textual documents available online and providing comprehensive descriptions of different impacts under natural diasters of a city. Instead of answering the question "which city is safer?" directly, we provide straightforward and descriptive information for the following four types of queries to help users make their own decisions:

—What are the major impacts caused by a specific type of disasters for the two cities? For example, hurricanes in Miami are more likely to cause house damage, but more likely to cause rainfall and landslide in Los Angeles.

—What are the major types of disasters leading to a specific type of impacts for the two cities? For example, "house damage" is mainly caused by hurricanes in Miami, but by earthquakes in Los Angeles.

—What are the most likely disasters affecting the two cities? For example, hurricanes occur more frequently in Miami, and earthquakes in Los Angeles.

—What are the overall impacts caused by disasters for the two cities? For example, in Miami, there is more flooding and house damage, and in Philadelphia it is more likely to have rainfall and death.

To answer these queries, we propose an interactive *weighted bipartite graph* to model impacts of disasters on cities. There are two types of nodes, **disaster** nodes and **impact** nodes, in the bipartite graph. **disaster** nodes represent hazards to city safety which cause significant damage and destruction and can be decided by domain experts or an ontology of disaster management. **impact** nodes represent consequences caused by **disaster** nodes, and are extracted from plain text via a topic modelling approach (Blei, Ng, and Jordan, 2003). A (weighted) edge from a **disaster** node to an **impact** node denote that the source node is responsible for the target node and its weight specifies to what extend the responsibility is. Triggered by users' queries, various comparative summaries will be generated from the filtered text to provide more textual descriptions of the differences between the two cities. A demonstration system can be visited at `http://bigdata-node01.cs.fiu.edu/textfilter/`.

In summary, our main contributions are the following:

—present a novel weighted bipartite graph based framework to model the problem of comparing city disaster susceptibilities, in which the casual relationship between different types of disasters and their impacts on a city is encoded in weighted edges;

—apply topic modelling to extract topics from documents representing different types of impacts of disasters;

—design a system which provides textual summaries about two cities upon various comparative queries;

—conduct a case study using Wikipedia documents on 4 cities of U.S to support 6 pairwise comparisons, and the results demonstrate the efficacy of our framework.

[2] `http://www.nhc.noaa.gov/data/`

[3] `http://earthquake.usgs.gov/earthquakes/map/`

The rest of the paper is organized as follows. After discussing related work in Section 2, we first give a brief overview of our framework in Section 3. Detailed descriptions of how to construct the bipartite graph and how to conduct city safety comparisons based on the bipartite graph is presented in Section 4 and Section 5, respectively. We present our case study results in Section 6 and finally conclude with discussions and outline of future extensions in Section 7.

2. RELATED WORK

City safety study has attracted much attention recently in computer science. Classical prediction methods such as ARIMA and artificial neural network (Chen, Yuan, and Shu, 2008; Chen, Chung, Xu, Wang, Qin, and Chau, 2004; Ballesteros, Rahman, Carbunar, and Rishe, 2012; Olligschlaeger, 1997) have been successfully applied in crime-related prediction, like drug market or other specially designed safety indices. Another direction is how to build up sensor networks that can quickly respond in an emergency event like fires and traffic accidents (Naphade, Banavar, Harrison, Paraszczak, and Morris, 2011; Karpiriski, Senart, and Cahill, 2006; Jung, Jeong, Lee, and Hong, 2009). While most existing studies focus on the safety of an individual city, our work provides a comparative view between different cities in term of their safety.

Many information systems and techniques have been proposed in disasters monitoring, relief and recovery. Commercial systems such as Web EOC and E-Team are usually used by Emergency Management departments located in urban areas (WebEOC, 2002; E-Teams, 2004). Ushahidi (Ushahidi, 2012) provides a platform to crowd source news stories and crisis information using multiple channels and prepares visualization and interactive maps. GeoVISTA (GeoVISTA, 2010) monitors tweets to form situation alerts on a map-based user interface according to the geo-locations associated with the tweets. These situation-specific tools provide query interfaces, GIS and visualization capabilities to support user interactions and queries to improve situation awareness (Zheng, Shen, Tang, Zeng, Li, Luis, and Chen, 2013) in a specific disaster event.

Multi-document summarization, especially comparative summarization, helps people understand what are the connections and differences between two document sets and has been studied with different applications. Kim and Zhai (Kim and Zhai, 2009) compare positive reviews and negative reviews for one product by extracting most related and representative sentence pairs for the two review sets, while Huang et al. (Huang, Wan, and Xiao, 2011) compare related news topics by extracting sentences covering most important related or representative concepts. Wang et al. (Wang, Zhu, Li, and Gong, 2012) model the comparative summary as a sentences set including most discriminative sentences from different document sets. Wan et al. (Wan, Jia, Huang, and Xiao, 2011) conduct comparative summarization on news from different regions (in different languages) on the same topic using random walk methods on a sentence graph. Instead of directly extracting sentences from different document set, this work utilizes weighted bipartite graph to model impacts of disasters, and filter documents for the comparative summarization.

3. FRAMEWORK OVERVIEW AND NOTATIONS

To capture the relationship between disasters and their impacts on a city, we propose a weighted bipartite graph based framework.

Definition 1. *A* weighted bipartite graph *is a graph $G = (U, V, E, w)$ whose vertices can be divided into two disjoint set U and V such that every edge connects a vertex in U to a vertex in V, i.e. $E \subseteq U \times V$, and $w : E \to \mathcal{R}^+$ is a weight function which assigns a non-negative weight to each edge $e \in E$.*

In our framework, U is the set of **disaster nodes**, V is the set of **impact nodes**, and every edge is associated with a triple (c, S, w), where c is the label of a city, S is a sentence set related to the edge and w is the weight of the edge.

Definition 2. Disaster nodes *are the (left) vertices in the bipartite graph that represent city hazards, such as hurricane, storm, tornado.*

Definition 3. Impact nodes *are the (right) vertices in the bipartite graph that represent consequences caused by the* ***disaster*** *nodes, such as death, house damage, economic loss.*

Definition 4. *An* impact topic of disasters *is a bag-of-words mostly commonly used to describe a type of impacts of disasters. For example, death, died, killed, fatalities, injuries are commonly used words to describe the impact "human life loss" caused by disasters.*

Figure 1 shows our framework architecture and Table I summarizes the notation used in this paper.

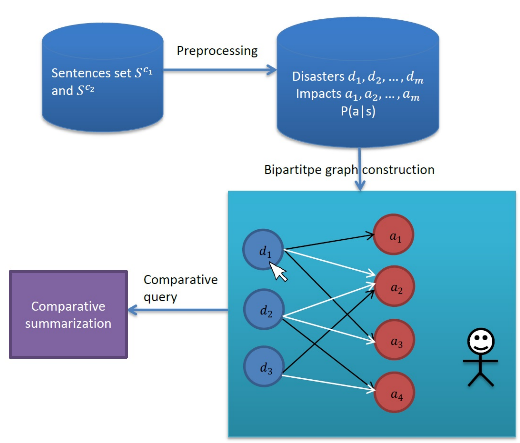
The input of our framework is several sets of sentences, $S^c, c \in \{c_1, c_2, ..., c_n\}$, and the sentence set S^c for city c is collected from online disaster-related documents (e.g., Wikipedia pages of disaster events in our case study in Section 6). Every sentence $s \in S^c$ depicts some aspect of the city c in a disaster event. Following is one sentence instance about *Chicago*:

Only two people died in the fire but 10,000 were made homeless and 1,800 buildings were burned to the ground.

In the above sentence, *fire* is a disaster type and impacts it causes include *death, homeless, building burned.*

To process the sentences, words describing disaster damages are extracted from sentences and grouped into *impact topics* in our framework. Then for each impact topic a, we assign a probability $p(a|s)$ for each sentence s (details to be described later), indicating the weight of impact topic a discussed in the sentence. For instance, in the above example, "homeless, building, burned" will be assigned higher weight than "died" for the disaster fire in Chicago.

The vertex set of the bipartite graph includes *disaster nodes* and *impact nodes*, representing disasters and impact topics, respectively. Edges between *disaster nodes* and *impact nodes* indicate causal relationship between them and the weight on an edge specifies the strength of the causal relationship. The bipartite graph encodes all the information about the queries mentioned in section 1 for city safety comparisons. Users can interact with this bipartite graph and submit a comparative query by clicking a node. The default query without clicking any nodes is: *what are the overall differences between city c_1 and c_2?* By clicking disaster node d_i, the query becomes: *what are the differences between city c_1 and c_2 on disaster d_i?* By further clicking impact node a_j, the query becomes: *what are the differences between city c_1 and c_2 on impact a_j caused by disaster d_i?*



An overview of the system framework.

Table I: A summary of notations

c	city
d	**disaster** node j
a	**impact** node k
S^c	sentences set of city c
$\hat{S}^c$	sentences set of city c after removing **impact** unrelated words
S_i^c	sentences subset of city c filtered out on **disaster** node i
$S_{i,j}^c$	sentences subset of city c filtered out on **disaster** node i and **impact** node j
$p(a\|s)$	probability of an impact topic a in sentence s, which comes from output of LDA. Here document-topic distribution in LDA model represents sentence-impact distribution.
$R_n(s)$	the most likely top n impact topics according to sentence-impact distribution of s
$e_{i,j}^c$	edge between **disaster** node i and **impact** node j on city c
$w_{i,j}^c$	weight of edge $e_{i,j}^c$

4. BIPARTITE GRAPH CONSTRUCTION

The the weighted bipartite graph is constructed as follows. First, we pre-define some disaster types like *hurricane, tornado, storm and earthquake*. We then apply domain knowledge of disaster management in the form of ontology, to extract sentences from the input sentence sets which contains concepts belong to those disasters. For instance, sentences containing the phrase "tropical cyclone" are extracted as sentences about "hurricane", since "tropical cyclone" is considered as a sub-concept of "hurricane".

4.1 Impact Node Extraction

According to definition 3 and definition 4, impact nodes encode negative consequences caused by disasters and is associated with a representative bag-of-words. However, unlike disaster nodes, it is difficult to enumerate or predefine all possible type of impacts and it is even more difficult to associate predefined types of impacts to actual description in real text documents. To overcome this difficulty, we extract impacts directly from text using information extraction and text mining techniques. Consider the ideal case in which input sentence set is on impacts of disasters on cities, and each sentence is a textual description of a tuple $(disaster, where, when, impact)$. Based on this intuition, we use a topic modelling tool, *latent Dirichlet allocation* (LDA) (Blei et al., 2003) to cluster words about impacts into several groups, each corresponds to an impact node. To exclude other words in sentences, we preprocess the original sentence set S^c as follows: (1) remove words related to **disaster** nodes using the same approach as in identifying disaster related sentences; (2) remove words explaining when, where, who using entity recognition techniques (Finkel, Grenager, and Manning, 2005); and (3) remove stop words.

After the preprocessing, we get a sentences set $\hat{S}^c$ for every city c. To compare two cities c_1 and c_2, we apply LDA on the preprocessed sentence set $\hat{S}^{c_1} \cup \hat{S}^{c_2}$ together with the impact number k, which specifies the number of **impact** nodes. The LDA will generate k topic with words distribution respectively, as well as a conditional probability $p(a|s)$ for every impact topics a on a given sentence s, which is then used to calculate the weights of edges between **disaster** nodes and **impact** nodes.

4.2 Weight Calculation for Disaster-Impact Edges

We calculate weight of an edge based on the sentence set of the city related to the disaster node and the impact node.

Let S_i^c be the set of sentences related to disaster i in S^c, which is extracted using a disaster ontology as

$$S_i^c = \{s \in S^c \mid s \text{ contains } d_i \text{ or a sub-type of } d_i\}. \tag{1}$$

Let $S_{i,j}^c \subset S_i^c$ be the sentence set about city c, **disaster** node i and **impact** node j, which, roughly speaking, is the set of sentences containing impact topic j:

$$S^c_{i,j} = \{s \in S^c_i \mid p(a_j|s) > \varepsilon\}, \tag{2}$$

where ε is a threshold parameter.

However we find it is difficult in practice to choose a proper parameter ε value, as it is very sensitive to the input data set. Small ε will lead to too many connections, while large ε will rule out too many sentences and result in very sparse bipartite graphs. Instead, in our framework, for every sentence s in S^c_i, we only consider its top n most likely impact topics $R_n(s)$[4], and use the following to define $S^c_{i,j}$ in place of Eq.(2):

$$S^c_{i,j} = \{s \in S^c_i \mid a_j \in R_n(s)\} \tag{3}$$

Finally, the weight of edge $e^c_{i,j}$, $w^c_{i,j}$, is defined as

$$w^c_{i,j} = \sum_{s \in S^c_{i,j}} p(a_j|s). \tag{4}$$

If $w^c_{i,j}$ is 0, then we remove the edge between d_i and a_j and assume there is no connection between the disaster and the impact.

5. CITY COMPARISONS BASED ON BIPARTITE GRAPH

Our framework provides city comparisons through two perspective views: simple comparisons and textual comparisons. Simple comparisons through bipartite graph provide general and direct discrepancies so that users can grasp quickly the differences between two cities but offer no detail textual descriptions. Textual comparisons remedy this by providing comparative summaries according to users' comparative queries.

5.1 Simple Comparisons

Figure 2 shows a simple comparison result of two cities, Miami and Los Angeles. From the thickness of edges between disaster nodes and impact nodes (used to denotes the weights of edges) in the bipartite graph, one can see that earthquakes occur more frequently in Los Angles, while in Miami hurricanes and tornadoes happen much more often.

More generally, the four types of queries of city safety comparisons described in Section 1 can now be addressed using information stored in the bipartite graph, in particular the edge weight $w^c_{i,j}$, which indicates causal strength between disaster d_i and a_j in city c, as follows:

- *What are the impact differences caused by the specific disaster d_i for city c_1 and c_2?* Such a query can be answered by comparing two weight vectors $w^{c_1}_{i,1}, \ldots, w^{c_1}_{i,k}$ and $w^{c_2}_{i,1}, \ldots, w^{c_2}_{i,k}$, which are visualized in the bipartite graph as line thickness of highlighted edges of different colors.
- *What are the disasters differences leading to specific impact a_j for city c_1 and c_2?* Such a query can be answered by comparing two vectors $e^{c_1}_{1,j}, \ldots, e^{c_1}_{m,j}$ and $e^{c_1}_{1,j}, \ldots, e^{c_1}_{m,j}$, which are visualized in the bipartite graph as line thickness of highlighted edges of different colors as well.
- *What are the overall disaster differences for city c_1 and c_2?* To answer such a query, for a city c and a disaster d_i, we aggregate weights of edges from d_i to all impact nodes as $w^c_{i*} = \sum_j w^c_{i,j}$. Then we can compare the two aggregated weight vectors $w^{c_1}_{1*}, \ldots, w^{c_1}_{n*}$ and $w^{c_2}_{1*}, \ldots, w^{c_2}_{n*}$, which are visualized as length of bars along with the disaster nodes.
- *What are the overall impact differences caused by disasters for city c_1 and c_2?* To answer such a query, for a city c and impact a_j, we accumulate weights of edges originating from all disaster nodes to a_j as $w^c_{*j} = \sum_i w^c_{i,j}$. Then we can compare two weight vectors $w^{c^1}_{*1}, \cdots, w^{c^1}_{*m}$ and $w^{c^2}_{*1}, \cdots, w^{c^2}_{*m}$, which are visualized as length of bars along with the impact nodes.

[4] n is set to 2 in our case study.

5.2 Textual Summarization for Comparative Queries

The bipartite graph provides simple comparisons using visualized weights induced from topic modelling, but it lacks detailed textual description, which can be remedied by textual comparative summarization. In this work, we apply the comparative summarization method in (Wang et al., 2012) on two sentences sets for two cities according to different comparative queries.

For two cities c_1 and c_2, our framework performs comparative summarization on sentence sets S^{c_1} and S^{c_2}. Different comparative queries from user interaction by clicking bipartite graph nodes decide S^{c_1} and S^{c_2} of comparative summarization. S^{c_1} and S^{c_2} are set to $S_i^{c_1}$ and $S_i^{c_2}$ by clicking disaster node d_i and $S_{i,j}^{c_1}$ and $S_{i,j}^{c_2}$ by sequentially clicking disaster node d_i and impact node a_j.

In (Wang et al., 2012) comparative summarization is modeled as a discriminative sentence selection process based on a multivariate normal generative model to extract sentences best describing the unique characteristics of each document group.

Problem 1. Suppose we have f sentences of the document collection, denoted by $\{X_i \mid i \in F\}$, where F is an index set of sentences with $|F| = f$. We are also given the group variable, Y, which is represented by multiple group indicator variables. The problem of *sentence selection* is to select a subset of sentences, $S \subset F$, to accurately discriminate a group of documents from other groups, i.e. to predict the group identity variable Y, given that the cardinality of S is m $(m < f)$. Let us denote $\{X_i \mid i \in S\}$ by X_s, for any set S. The prediction capability of Y given X_s can be measured by the entropy of Y given X_s, which is defined as

$$H(Y|X_s) \stackrel{def}{=} -E_{p(Y,X_s)} \log p(Y|X_s), \tag{5}$$

where $E_p(\cdot)$ is the expectation given the distribution p, and p stands for the underlying document distribution, i.e. the joint distribution $p(Y, X_s)$. The sentence selection problem using the mutual information criterion is

$$\operatorname*{argmin}_S H(Y|X_S). \tag{6}$$

Selecting an optimal subset of sentences known to be a NP-hard problem. A greedy approach is proposed in (Wang et al., 2012) by sequentially selecting features to obtain a sub-optimal solution.

6. CASE STUDY

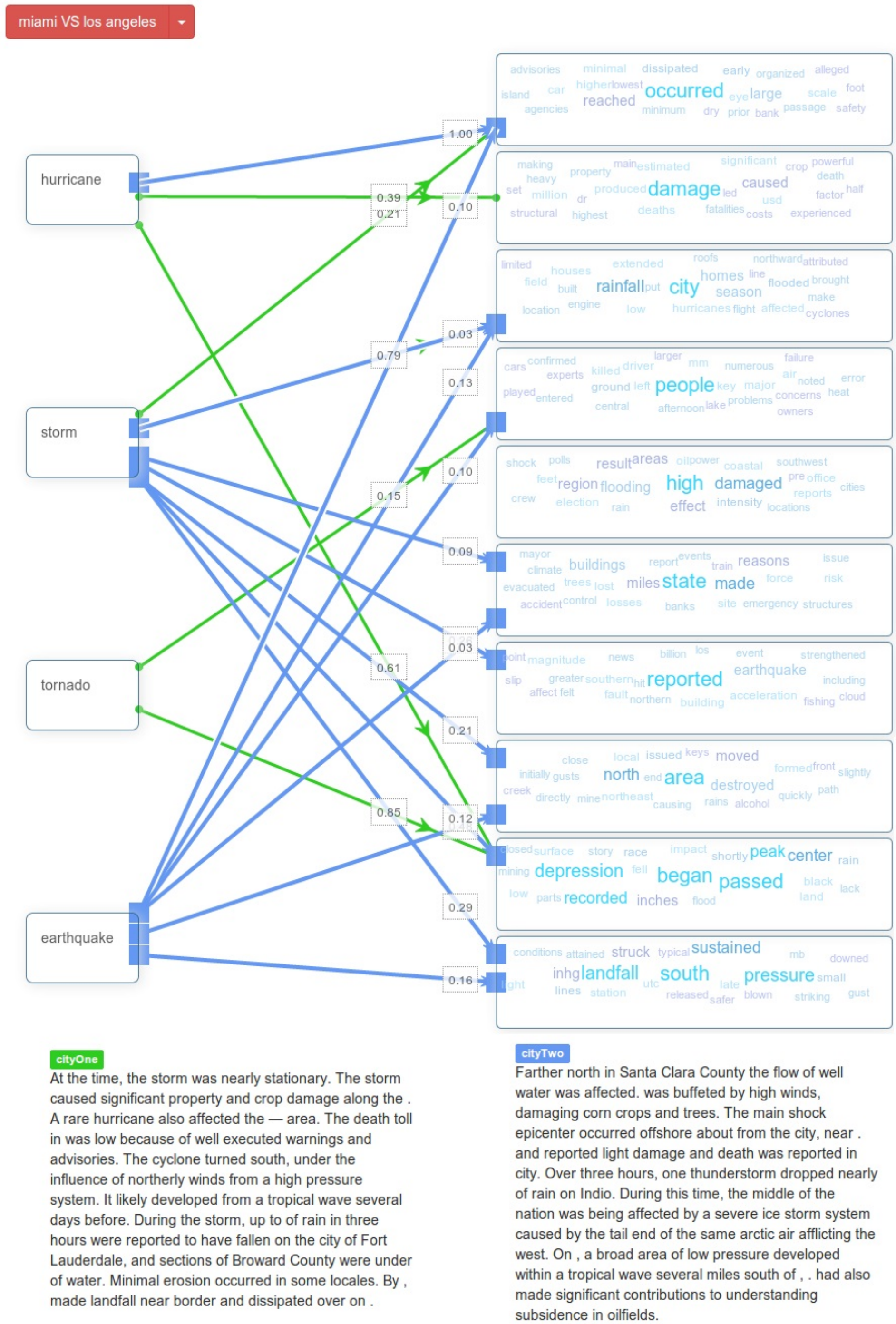
To show the effectiveness of the proposed framework, a case study is conducted to compare city safety among four U.S. cities Miami, Chicago, Los Angeles and Philadelphia, using the impacts of four types of disasters – hurricane, storm, tornado and earthquake.

6.1 Dataset

We collect the dataset from Wikipedia. For each city, we first extract all the paragraphs of Wikipedia page containing the city name, and further extract sentences containing phrases about one of the four disaster types from. Table II shows the basic statistics of the dataset.

Table II: The sentence set sizes of cities

city	# of sentences
Miami	772
Chicago	618
Los Angeles	607
Philadelphia	685



Bipartite graph of city pair Miami and Los Angeles

Table III: Most likely disaster types and impact types for two cities in pairwise comparison.

City Pair	$\text{argmax}_d\, p_1(d)$	$\text{argmax}_d\, p_2(d)$	$\text{argmax}_e\, p_1(e)$	$\text{argmax}_e\, p_2(e)$
Miami Chicago	storm	storm	landfall, fatalities, weather	damaged, struck, collapse
Miami Los Angeles	storm	earthquake	depression, inches, rain	ground, killed, dropped
Miami Philadelphia	storm	storm	killed, flooded, streets	destroyed, accident, fatalities
Chicago Los Angeles	storm	earthquake	rain, temperatures, flood	flight, killed, billion
Chicago Philadelphia	storm	storm	fire, flight, alarm	death, rain, attack
Los Angeles Philadelphia	earthquake	storm	adventures, destroyed, discovery	fire, killed, weather

Table IV: Most likely effect caused by each disaster for cities Miami and Los Angeles.

City Pair	hurricane	storm	tornado	earthquake
Miami	crash, bodies, dropped	landfall, fatalities, weather	damage, million, buildings	$\emptyset$
Chicago	landfall, fatalities, weather	flooding, homes, killed	warning, pressure, tides	depression, quickly, evaluated
Miami	depression, inches, rain	rainfall, flooded, houses	depression, inches, rain	$\emptyset$
Los Angeles	occurred, large, reached	landfall, pressure, struck	$\emptyset$	warnings, destroyed, moved
Miami	killed, flooded, streets	landfall death warnings	reported, pressure, force	$\emptyset$
Philadelphia	peak, inches, power	damage, rainfall, million	$\emptyset$	reported, pressure, force

6.2 Results

The aforementioned Figure 2 demonstrates an experimental result visualizing pairwise city comparison between Miami and Los Angeles, in which green components encodes information for city Miami and blue for city Los Angeles. Furthermore, Table III shows the general differences in pairwise city comparison. The third column shows the most likely disaster types, and the fourth is the most likely effects. For each entry, 3 representative words are manually selected among 15 top-ranked words, according to the word probability in the corresponding impact topic generated from LDA. Similar to section 5.1, we answer the queries in section 1 from our user case study results.

What are the overall disaster differences for city Miami and Los Angeles? From Table III, one can see that the most likely disaster for Miami is storm, and earthquake for Los Angele. This reflects the real difference between these two cities, since Miami is a city located on the Atlantic coast in south-eastern Florida which has a tropical monsoon climate and Los Angeles is subject to earthquakes due to its location on the Pacific Ring of Fire. In addition, from figure 2, one can see that tornadoes barely happen in Los Angeles.

What are the overall impact differences caused by disasters for city c_1 and c_2? Table III shows that the most likely impact types are *depression, inches, rain*, which is regarded as rainfall, but for city Los Angeles it is *ground,kill,dropped*, which is interpreted as life loss and house collapse. This observation can be easily explained since frequently occurred storm in Miami cause plentiful rainfall while earthquake in Los Angeles causes life loss and house collapse. Here, we only illustrate results of pairwise city comparison between Miami and Los Angeles; results from other five pairwise city comparisons are listed in Table III.

What are the impact differences caused by the specific disaster d_i for city c_1 and c_2? Table IV highlights the comparison between Miami and Los Angeles for the most likely effect given a disaster type, which provides answer for this query. Most likely type of impacts caused by hurricane in Miami is *depression, inches, rain*, but *occurred, large, reached* for Los Angeles. Besides, storms in Miami most likely cause *rainfall, flooded, houses*, but in Los Angeles they mainly cause much more peaceful type of impacts *landfall, pressure, struck*. These differences can be explained by that Miami is more geographically flat but Los Angeles mountainous which obstruct further evolution of strong rainfall. For the other two disasters, tornadoes only occur in Miami and mainly lead to impacts *depression, inches, rain*, meanwhile earthquakes only occur in Los Angeles and mainly lead to impacts *warning, destroyed, moved*.

7. CONCLUSION

In this paper, we study the problem of comparing cities' disaster susceptibilities, and propose a novel bipartite graph based framework. Using our framework, direct comparisons can be performed on the bipartite graph and additional textual comparative summaries for different queries can be generated through interactions between users' and the bipartite graph.

For future work, we plan to extend our framework in following aspects: (1) We will improve the impact node extraction to extract more accurate impact topics; (2) More safety issues like crime and man-made disasters will be included. (3) More sophisticated graph algorithms like random walk will be employed to utilize the bipartite graph structure in our framework in order to get more accurate results.

Acknowledgment

The work was supported in part by the National Science Foundation under grants HRD-0833093, CNS-1126619, and IIS-1213026, the U.S. Department of Homeland Security under grant Award Number 2010-ST-062000039, and Army Research Ofce under grant number W911NF-1010366 and W911NF-12-1-0431.

References

Ballesteros, J., Rahman, M., Carbunar, B., and Rishe, N. 2012. Safe cities. a participatory sensing approach. In *Local Computer Networks (LCN), 2012 IEEE 37th Conference on*. IEEE, 626–634.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation. *the Journal of machine Learning research 3*, 993–1022.

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., and Chau, M. 2004. Crime data mining: a general framework and some examples. *Computer 37,* 4, 50–56.

Chen, P., Yuan, H., and Shu, X. 2008. Forecasting crime using the arima model. In *Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on*. Vol. 5. IEEE, 627–630.

E-Teams. 2004. `http://www.nc4.us/ETeam.php`.

Finkel, J. R., Grenager, T., and Manning, C. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 363–370.

GeoVISTA. 2010. http://www.geovista.psu.edu.

Huang, X., Wan, X., and Xiao, J. 2011. Comparative news summarization using linear programming. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*. Association for Computational Linguistics, 648–653.

Jung, H.-S., Jeong, C.-S., Lee, Y.-W., and Hong, P.-D. 2009. An intelligent ubiquitous middleware for u-city: Smartum. *Journal of Information Science & Engineering 25,* 2.

Karpiriski, M., Senart, A., and Cahill, V. 2006. Sensor networks for smart roads. In *Pervasive Computing and Communications Workshops, 2006. PerCom Workshops 2006. Fourth Annual IEEE International Conference on*. IEEE, 5–pp.

Kim, H. and Zhai, C. 2009. Generating comparative summaries of contradictory opinions in text. In *Proceeding of the 18th ACM conference on Information and knowledge management*. ACM, 385–394.

Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., and Morris, R. 2011. Smarter cities and their innovation challenges. *Computer 44,* 6, 32–39.

Olligschlaeger, A. M. 1997. Artificial neural networks and crime mapping. *Crime mapping and crime prevention*, 313–348.

Ushahidi. 2012. http://www.ushahidi.com/.

Wan, X., Jia, H., Huang, S., and Xiao, J. 2011. Summarizing the differences in multilingual news. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information*. ACM, 735–744.

Wang, D., Zhu, S., Li, T., and Gong, Y. 2012. Comparative document summarization via discriminative sentence selection. *ACM Transactions on Knowledge Discovery from Data (TKDD) 6,* 3, 12.

WebEOC. 2002. `http://www.esi911.com/home`.

Zheng, L., Shen, C., Tang, L., Zeng, C., Li, T., Luis, S., and Chen, S.-C. 2013. Data mining meets the needs of disaster information management. *IEEE Transactions on Human-Machine Systems 43,* 5, 451–464.