A Bipartite-Graph Based Approach for Disaster Susceptibility Comparisons among Cities

Wubai Zhou, Chao Shen, Tao Li, Shu-Ching Chen, Ning Xie, Jinpeng Wei
School of Computing and Information Sciences
Florida International University
Miami, FL 33199, U.S.A.
Email: {wzhou005,cshen001,taoli,chens,nxie,weijp}@cs.fiu.edu

Abstract—**People are attracted to large cities because of more employment opportunities, convenient facilities, and rich cultural activities. However, large cities are also more vulnerable to natural disasters, which have caused widespread physical destructions, great loss of life and property, and immense havoc. "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 comparison using statistics, our system also generates textual comparative summaries to better describe the differences between the two cities in terms 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: Disaster Susceptibility Comparison, Disaster-Impact Bipartite Graph, Comparative Summarization

I. Introduction

People are attracted to metropolitan areas due to more employment opportunities, convenient facilities, and rich cultural activies. However, large cities are also vulnerable to natural disasters, which tend to cause more damage in densely populated areas. For example, about 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; lots of earthquakes have happened in the two major cities on the west coast of U.S., Los Angeles and San Francisco; and frequent hurricane hits in Miami area. Therefore, before making decisions on traveling or job and business relocation, one of the most critical questions people face is: which city is safer?

For city safety comparison, a number of online services[1] provide statistic data about various aspects of cities or neighborhoods such as crime rates, races, living expenses and house prices. 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 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 and archived by different government agencies or organizations. Extra efforts are often required to collect data or/and perform data integration into a unified database to support comparisons among different cities. Moreover, although statistics about damages and fatalities can provide direct evidences for the safety comparison, it is still quite challenging to obtain an overview on historically how severe a city was affected by disasters, since many types of impacts from disasters – e.g., road closure caused by a hurricane – are not reflected by the statistics.

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 about a pair of cities 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 impact 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 or frequent 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 the disaster impacts on cities. There are two types of nodes, **disaster** nodes and **impact** nodes, in the bipartite graph. **Disaster** nodes represent hazards to the city safety which can cause significant damages and destructions. The hazards can be decided by domain experts

[1] Examples include: http://www.neighborhoodscout.com, http://www.numbeo.com/, and http://www.city-data.com
[2] http://www.nhc.noaa.gov/data/
[3] http://earthquake.usgs.gov/earthquakes/map/

or using an ontology of disaster management. **Impact** nodes represent consequences caused by **disaster** nodes, and they are extracted from plain texts via a topic modeling approach [1]. A weighted edge from a **disaster** node to an **impact** node denotes that the source node is responsible for the target node and its weight specifies to what extent the responsibility is. Triggered by users' queries, various comparative summaries will be generated from the filtered text to provide detailed textual descriptions of the differences between the two cities. A demonstration system can be visited at http://bigdata-node01.cs.fiu.edu/CitySafetyComparison/.

In summary, our main contributions are listed below:

- We present a 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;
- We apply topic modeling to extract topics from documents to represent different types of disaster impacts;
- We design a prototype system which provides textual summaries about two chosen cities for various comparative queries;
- We conduct a case study using Wikipedia documents on 4 different U.S. cities with 6 pairwise comparisons to demonstrate the effectiveness of our proposed framework.

The rest of the paper is organized as follows. After discussing related work in Section II, we first give a brief overview of our framework in Section III. Detailed descriptions of how to construct the bipartite graph and how to conduct city safety comparisons based on the bipartite graph are presented in Section IV and Section V, respectively. We present our case study results in Section VI and finally conclude with discussions and outlines for future extensions in Section VII.

II. Related Work

City safety study has attracted much attention recently in computer science. Classical prediction methods such as ARIMA models and artificial neural networks [2], [3], [4], [5] 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 [6], [7], [8]. While most existing studies focus on the safety of an individual city, our work provides a comparative view between different cities in terms 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 [9], [10]. Recently many disaster situation-specific tools provide query interfaces, GIS and visualization capabilities to support user interactions and queries to improve situation awareness [11] in a specific disaster event.For example, Ushahidi [12] provides a platform with visualization and interactive maps to crowd source news stories and crisis information using multiple channels and GeoVISTA [13] monitors tweets to form situation alerts according to the geo-locations associated with the tweets. However, these tools do not answer the comparative queries about all disaster related data of different cities.

Multi-document summarization has been used to provide concise summaries about large document collections and many different approaches have been developed including centroid-based [14], graph-based [15], [16], clustering-based [17], [18], knowledge-based [19], [20], etc. Comparative summarization, as a special class of summarization tasks, helps people understand what are the connections and differences between two document collections and has been studied with different applications. Kim and Zhai [21] compare positive reviews and negative reviews for one product by extracting the most related and representative sentence pairs for the two review sets, while Huang et al. [22] compare related news topics by extracting sentences covering the most important related or representative concepts. Wang et al. [23] model the comparative summary as a sentence set including the most discriminative sentences from different document sets. Wan et al. [24] 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 sets, this work utilizes the weighted bipartite graph to model impacts of disasters and filter documents for comparative summarization.

Graph-based approaches have also been used to generate event storylines that describe how an event evolves over time. Wang et al. [25] developed a multi-view graph based framework for integrating text, image, and temporal information to generate storylines to reflect the evolution of the given topic. Wu et al. [26] proposed a two-layer storyline generation framework which provides global storylines for cross-location disaster events on the first layer and location-specific storylines for individual events on the second layer. Shahaf et al. [27] developed *metro map* for creating structural summaries of documents by optimizing several objectives (e.g., relevance, coherence, coverage and connectivity) simultaneously. Unlike existing studies, in this work, we utilize a weighted bipartite graph based framework to perform city safety comparison.

III. The 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 and 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 and economic loss.

Definition 4: An *impact topic of disasters* is a bag-of-words which are 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 VI). Every sentence $s \in S^c$ depicts some aspect of the city c in a disaster event.

The following is a 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 its impacts 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 (the details will be described in Section IV-A), indicating the weight of impact topic a discussed in the sentence. For instance, in the above example, "homeless, building, burned" will be assigned higher weights 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 the causal relationship between them and the weight on an edge specifies the strength of the relationship. The bipartite graph encodes all the information about the queries mentioned in Section I for city safety comparison. 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 a 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 a 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?*

IV. Bipartite Graph Construction

The weighted bipartite graph is constructed as follows. First, we pre-define some disaster types like *hurricane, tornado, storm and earthquake*. We then apply a domain ontology of disaster management [19], [20] to extract sentences from the input sentence sets which contains concepts belonging 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".

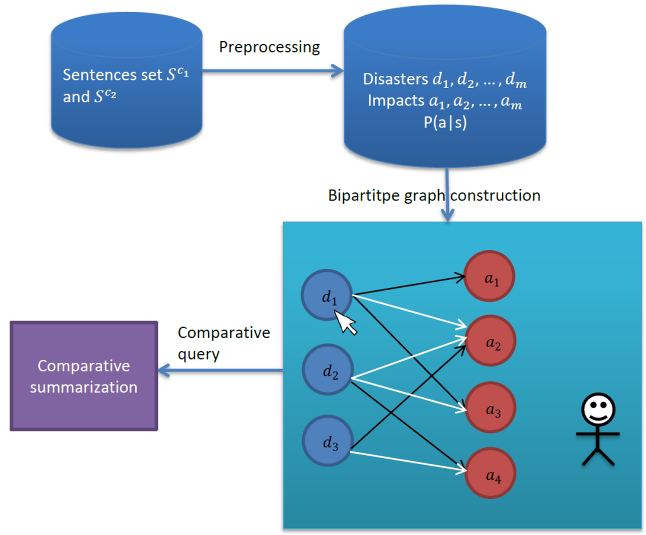

Fig. 1. An Overview of The System Framework.

TABLE I. A Summary of Notation.

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^c_i	sentences subset of city c filtered out on **disaster** node i
$S^c_{i,j}$	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
$e^c_{i,j}$	edge between **disaster** node i and **impact** node j on city c
$w^c_{i,j}$	weight of edge $e^c_{i,j}$

A. Impact Node Extraction

According to Definition 3 and Definition 4, impact nodes encode negative consequences caused by disasters and are associated with a representative bag-of-words. However, unlike disaster nodes, it is difficult to enumerate or predefine all possible impact types and it is even more difficult to associate predefined impact types to the actual textual descriptions in given documents. To overcome this difficulty, we extract impacts directly from texts using information extraction and text mining techniques. Consider the ideal case in which the input sentence set is about disaster impacts on cities, then each sentence is a textual description of a tuple $(disaster, where, when, impact)$. Therefore, if each impact node represents an impact topic, we need to identify different impacts that have less overlap with each other. Based on this intuition, we use a topic modeling tool, *latent Dirichlet allocation* (LDA) [1] to cluster words about impacts into several groups, where each group corresponds to an impact topic. To exclude other unnecessary words in the sentences, we preprocess the original sentence set S^c as follows: (1) remove words related to **disaster** nodes; (2) remove words explaining when, where, who using entity recognition techniques [28]; and (3) remove the stop words.

After the preprocessing, we obtain 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 approach 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.

B. Weight Calculation for Disaster-Impact Edges

We calculate the 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

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

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:

$$S_{i,j}^c = \{s \in S_i^c \mid p(a_j|s) > \epsilon\}, \quad (2)$$

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. A small ϵ will lead to too many connections, while a 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_i^c, we only consider its top n most likely impact topics $R_n(s)$ (n is set to 2 in our case study), and use the following to define $S_{i,j}^c$ in place of Eq.(2):

$$S_{i,j}^c = \{s \in S_i^c \mid a_j \in R_n(s)\} \quad (3)$$

Finally, the weight of edge $e_{i,j}^c$, $w_{i,j}^c$, is defined as

$$w_{i,j}^c = \sum_{s \in S_{i,j}^c} p(a_j|s). \quad (4)$$

If $w_{i,j}^c$ 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.

V. City Comparison Based on The Bipartite Graph

Our framework provides city comparisons through two perspective views: simple comparison and textual comparison. Simple comparison through bipartite graph gives general and direct discrepancies so that users can quickly grasp the differences between two cities but it does not provide detailed textual description. Textual comparison remedies this by providing comparative summaries according to users' comparative queries.

A. Simple Comparison

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 observe that earthquakes occur more frequently in Los Angles, while in Miami hurricanes happen much more often.

More generally, the four types of queries of city safety comparisons described in Section I can now be addressed using the information stored in the bipartite graph (in particular, the edge weight $w_{i,j}^c$ indicating the 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 $w_{i,1}^{c_1}, \ldots, w_{i,k}^{c_1}$ and $w_{i,1}^{c_2}, \ldots, w_{i,k}^{c_2}$, which are visualized in the bipartite graph as the line thickness of highlighted edges with 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 $e_{1,j}^{c_1}, \ldots, e_{m,j}^{c_1}$ and $e_{1,j}^{c_2}, \ldots, e_{m,j}^{c_2}$, which are visualized in the bipartite graph as the line thickness of highlighted edges with 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 $w_{i*}^c = \sum_j w_{i,j}^c$. Then we can compare the two aggregated weight vectors $w_{1*}^{c_1}, \ldots, w_{n*}^{c_1}$ and $w_{1*}^{c_2}, \ldots, w_{n*}^{c_2}$, which are visualized as the 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 $w_{*j}^c = \sum_i w_{i,j}^c$. Then we can compare two weight vectors $w_{*1}^{c^1}, \cdots, w_{*m}^{c^1}$ and $w_{*1}^{c^2}, \cdots, w_{*m}^{c^2}$, which are visualized as the length of bars along with the impact nodes.

B. Textual Summarization for Comparative Queries

The bipartite graph provides simple comparisons using weights induced from topic modeling, but it lacks detailed textual descriptions, which can be remedied by textual comparative summarization. In this work, we apply the comparative summarization method in [23] on two sentence sets according to different comparative queries.

For two cities c_1 and c_2, our framework performs comparative summarization on two sentence sets S^{c_1} and S^{c_2}. Different comparative queries (resulted from user interactions via clicking bipartite graph nodes) will generate differentS^{c_1} and S^{c_2} for comparative summarization. For example, S^{c_1} and S^{c_2} are set to be $S_i^{c_1}$ and $S_i^{c_2}$ when a user clicks the disaster node d_i; they are set to be $S_{i,j}^{c_1}$ and $S_{i,j}^{c_2}$ when the user sequentially clicks the disaster node d_i and the impact node a_j.

TABLE III. Most likely disaster types and impact types for the 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 impacts caused by each disaster type for cities of Miami and Los Angeles.

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

TABLE II. The size of city sentence set

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

In [23], the 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 an NP-hard problem. A greedy approach is proposed in [23], which sequentially selects sentences to obtain a sub-optimal solution.

VI. The Case Study

To demonstrate 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.

A. Dataset Description

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. Table II shows the basic statistics of the dataset.

B. Results Analysis

Figure 2 demonstrates the comparative result of pairwise city comparison between Miami and Los Angeles, in which green components encode the information for city Miami and blue components encode the information for city Los Angeles. Furthermore, Table III shows the general differences in pairwise city comparison. The third column in Table III lists the most likely disaster types, and the fourth column in Table III lists the most likely effects/impacts. For each entry, 3 representative words are manually selected among the 15 top-ranked words, according to the word probability in the corresponding impact topic generated from LDA. Similar to Section V-A, we can answer the following queries in Section I from the 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 the most likely disaster for Los Angels is earthquake. 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 observe that tornadoes barely happen in Los Angeles.

What are the overall impact differences caused by disasters for city Miami and Los Angeles? Table III shows that the most likely impact types for city Miami are *depression, inches, rain*, which is regarded as rainfall, but for city Los Angeles they are *ground,kill,dropped*, which can be interpreted as life loss and house collapse. This observation can be easily explained since frequently occurred storms in Miami cause plentiful rainfall while earthquakes in Los Angeles cause life loss and house collapse. Here, we only illustrate results of pairwise

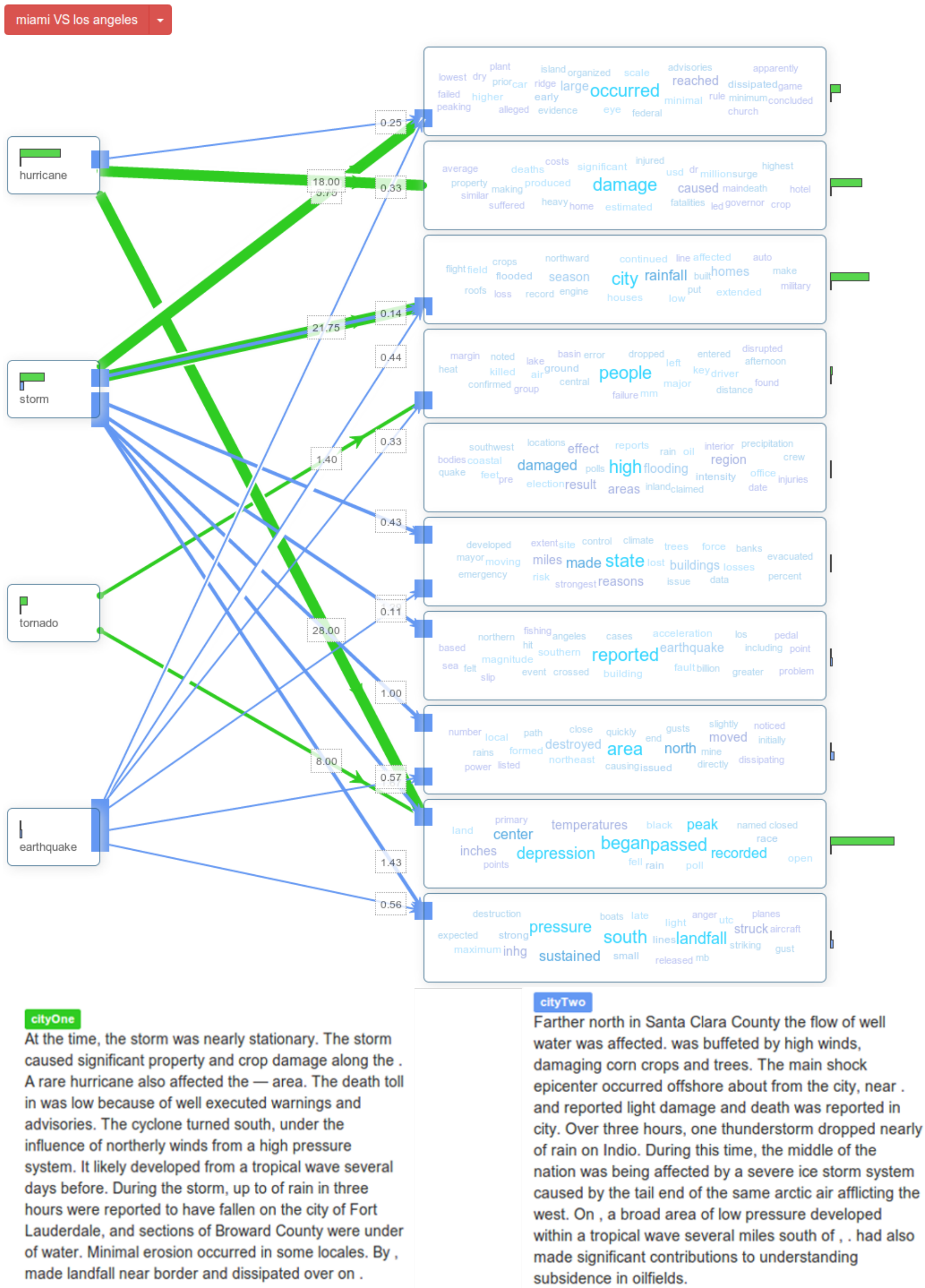

Fig. 2. Bipartite graph of city pair Miami and Los Angeles

city comparison between Miami and Los Angeles; results from other five pairwise city comparisons are also listed in Table III.

What are the impact differences caused by the specific disaster d_i for city Miami and Los Angeles? Table IV highlights the comparison between Miami and Los Angeles for the most likely effects/impacts given a disaster type, which provides answers for this type of query. The most likely impact types caused by hurricane in Miami are *depression, inches, rain*, but the impact types for Los Angeles are *occurred, large, reached*. 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 is more mountainous, which obstructs 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*.

VII. Conclusion

In this paper, we study the problem of comparing cities' disaster susceptibilities and propose a weighted bipartite graph based framework. Using our framework, direct city comparison can be performed on the bipartite graph and additional textual comparative summaries for different queries can be generated through user interactions via clicking the bipartite graph nodes.

For the future work, we plan to extend our framework in the following aspects: (1) We will improve the impact node extraction to extract more accurate impact topics; (2) We will include more safety issues like crime and man-made disasters; (3) We will employ more efficient graph algorithms (e.g., random walk) to utilize the bipartite graph structure in our framework.

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, the U.S. Department of Homeland Security's VACCINE Center under Award Number 2009-ST-061-CI0001, and Army Research Ofce under grant number W911NF-1010366 and W911NF-12-1-0431.

References

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

[2] P. Chen, H. Yuan, and X. Shu, "Forecasting crime using the arima model," in *Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on*, vol. 5. IEEE, 2008, pp. 627–630.

[3] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," *Computer*, vol. 37, no. 4, pp. 50–56, 2004.

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

[5] A. M. Olligschlaeger, "Artificial neural networks and crime mapping," *Crime mapping and crime prevention*, pp. 313–348, 1997.

[6] M. Naphade, G. Banavar, C. Harrison, J. Paraszczak, and R. Morris, "Smarter cities and their innovation challenges," *Computer*, vol. 44, no. 6, pp. 32–39, 2011.

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

[8] H.-S. Jung, C.-S. Jeong, Y.-W. Lee, and P.-D. Hong, "An intelligent ubiquitous middleware for u-city: Smartum." *Journal of Information Science & Engineering*, vol. 25, no. 2, 2009.

[9] E. A. Inc, "Webeoc," http://www.esi911.com/home.

[10] NC4, "E-teams," http://www.nc4.us/ETeam.php.

[11] L. Zheng, C. Shen, L. Tang, C. Zeng, T. Li, S. Luis, and S.-C. Chen, "Data mining meets the needs of disaster information management," *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 5, pp. 451–464, 2013.

[12] Ushahidi, "http://www.ushahidi.com/," 2012.

[13] GeoVISTA, http://www.geovista.psu.edu.

[14] D. R. Radev, H. Jing, and M. Budzikowska, "Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation, and user studies," in *Proceedings of the 2000 NAACL-ANLP Workshop on Automatic Summarization*. Association for Computational Linguistics, 2000, pp. 21–30.

[15] G. Erkan and D. R. Radev, "Lexpagerank: Prestige in multi-document text summarization." in *EMNLP*, vol. 4, 2004, pp. 365–371.

[16] C. Shen and T. Li, "Multi-document summarization via the minimum dominating set," in *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 984–992.

[17] C. Shen, T. Li, and C. H. Ding, "Integrating clustering and multi-document summarization by bi-mixture probabilistic latent semantic analysis (plsa) with sentence bases." in *AAAI*, 2011.

[18] D. Wang, T. Li, S. Zhu, and C. Ding, "Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization," in *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008, pp. 307–314.

[19] L. Li and T. Li, "An empirical study of ontology-based multi-document summarization in disaster management," *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, vol. 44, no. 2, 2014.

[20] L. Li, D. Wang, C. Shen, and T. Li, "Ontology-enriched multi-document summarization in disaster management," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010, pp. 819–820.

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

[22] X. Huang, X. Wan, and J. Xiao, "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*, 2011, pp. 648–653.

[23] D. Wang, S. Zhu, T. Li, and Y. Gong, "Comparative document summarization via discriminative sentence selection," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 6, no. 3, p. 12, 2012.

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

[25] D. Wang, T. Li, and M. Ogihara, "Generating pictorial storylines via minimum-weight connected dominating set approximation in multi-view graphs." in *AAAI*, 2012.

[26] W. Zhou, C. Shen, T. Li, S. Chen, N. Xie, and J. Wei, "Generating textual storyline to improve situation awareness in disaster management," in *In Proceedings of the 15th IEEE International Conference on Information Reuse and Integration (IRI 2014)*, 2014.

[27] D. Shahaf, C. Guestrin, and E. Horvitz, "Trains of thought: Generating information maps," in *Proceedings of the 21st international conference on World Wide Web*. ACM, 2012, pp. 899–908.

[28] J. R. Finkel, T. Grenager, and C. Manning, "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, 2005, pp. 363–370.