Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 15, no. 10, October 2019, pp. 2608-2617
DOI: 10.23940/ijpe.19.10.p6.26082617

Performance Analysis on Features of Headlines and Media Organizations about Terrorist Attacks: based on a Combinatory Algorithm

Junchao Feng[a,b,*], Ping Liu[c], Jianjun Miao[a], Ruilun Liu[d], and Dongbo Wang[d]

[a]*School of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, 211100, China*
[b]*Foreign Language Department, Harbin University of Science and Technology, Harbin, 150080, China*
[c]*Library Information Consultation Department, Harbin University of Science and Technology, Harbin, 150080, China*
[d]*School of Information and Technology, Nanjing Agricultural University, Nanjing, 21005, China*

Abstract

This paper aims to study the relationship between the geographical distribution of terrorist attacks and features of headlines and media organizations reporting terrorist attacks. Data are derived from the Global Terrorism Database (GTD) and are established into terrorist attacks news corpus along One Belt One Road. This paper adopted a combinatory algorithm with the edit distance algorithm and the longest common subsequence algorithm to calculate the similarity headlines of media organizations across the world. Also, this paper further explores the relationship between the geographical distribution of terrorist attacks and that of global media organizations with co-occurrence analysis and social network analysis. The results show that the mainstream media organizations always imitate or copy news reports of media organizations with regional natures where terrorist attacks often happened. In regards to the geographical distribution of terrorist attacks, the results show that there is a positive correlation between the geographical distribution frequencies of terrorist attacks and that of media organizations in areas prone to suffering from terrorist attacks. The proposed combinatory algorithm for features of media coverage about terrorist attacks along One Belt One Road could provide a very significant performance increase in the terrorist events studies. The findings of this research could help relevant institutions early warn and guard against hazards of terrorist attacks.

Keywords: feature analysis; media organization and headline; terrorist attack; the longest common subsequence algorithm

(Submitted on August 22, 2019; Revised on September 26, 2019; Accepted on October 20, 2019)

1. Introduction

The terrorist attack that frequently happened along "One Belt One Road" has drawn worldwide attention. However, it is reasonable to believe that the ongoing implementation of the Initiative would promote the prevailing peace and prosperity of countries along the One Belt One Road [1-4]. Therefore, there is a growing interest in actively exploring safety problems of the terrorist attacks that happened in areas along the One Belt One Road. Research on rules of terrorist attacks' occurrence and development would be of great significance not only for the implementation of the Initiative, but also on security and emergency management at home and aboard.

Despite the extensive literature on terrorism, much of the existing research has used countries or conflicts as units to analyze terrorism attacks events [5-8]. There is existing research on feature analysis of terrorist-attack news reports [9-13] and characteristics of online news [14]. Nevertheless, the relationship between terrorist attacks and features of media organizations and headlines is rarely considered. What is the relationship between the geographical distribution of terrorist attacks and features of headlines and media organizations reporting terrorist attacks?

The headline is a kind of short text with sparse feature [15], which is a high-level overview of terrorist attacks when facing massive news data. In order to investigate the features of headlines and media organizations concerning terrorist

* Corresponding author.
E-mail address: fengjunchao@nuaa.edu.cn

attacks, the corpus-driven research paradigm is utilized to calculate the features with the edit distance algorithm and the longest common subsequence algorithm. The analysis on features of headlines and media organizations related to terrorist-attack news would help find the rule of a terrorist attack. Meanwhile, it would also help the public and the authorities know about the latest incident dynamic and learn of the current event. Consequently, the analysis on features of headlines and media organizations reporting terrorist attacks besides the relationship between the geographical distribution of terrorist attacks and that of media organizations would be a golden reference for early warning and guarding against hazards of terrorist attacks along One Belt One Road through features of media coverage about terrorist attacks.

2. Experiment Design

2.1. Subject

Data are derived from the Global Terrorism Database (GTD) from 2001 to 2015. Sixty-six "neighborhood" countries with high frequencies of terrorist attacks along the One Belt One Road are selected. The full texts of terrorist-attack news are obtained by Python software, including each event time, place, attack objective, and manner, etc. Then, news corpus of terrorist attacks along One Belt One Road is built up. This study collected a total of six fields including event number (EVENTID), year of events (IYEAR), country of events (COUNTRY_TXT) and three reference sources (SCITE1 SCITE2, SCITE3). Three reference sources refer to the data sources of each field for a specific event within news reporter, news headlines, media institutions or database (also known as the media institutions) and reported time. They are stored in the format of「reporters, "news headlines", media organizations, reports on time.」in a cell. In these three fields, this study mainly extracts the news headlines and media institutions as the research objects.

2.2. Methods

Our approach follows several recent papers focusing on similarities of headlines and media organizations that rely on edit distance algorithm [16] and the longest common subsequence algorithm [17].

2.2.1. Edit Distance Algorithm

Edit distance algorithm is first put forward by the Russian scientist, Levenshtein. It belongs to a kind of orderly matching algorithm. It means that one string into another string needs the minimum times of replacement, deletion, and insertion, and every operated object is a character. Suppose there are two strings, *A* and *B*. The collection with minimum the removal, insertion, and replacement when *B* is converted to *A*, is the edit distance between *A* and *B*. Set A string length is *m*, *B* string length is *n*, the edit distance of them is $dist(m,n) = 0$, the dynamic programming formula is:

- When $m = 0$ and $n > 0, dist(m,n) = n$
- When $m > 0$ and $n = 0, dist(m,n) = m$
- When $m > 0$ and $n > 0, dist(m,n) = \min\big((dist(m-1),n) + 1, dist(m,n-1) + 1 + f(m,n)\big)$

$f(m,n)$ is the required operating cost from the first m characters of A the first n characters of B, when the first m characters A is equal to the first n characters of B. If no operation is needed, $f(m,n) = 0$; otherwise, $f(m,n) = 1$ [10].

In general, the smaller the edit distance, the higher the similarity of two strings; otherwise, the larger the edit distance, the lower the similarity of two strings. Edit distance similarity algorithm uses 1 minus the ratio of the edit distance and the longer length of two strings. The specific formula is as follows:

$$S = 1 - \frac{E}{max(m,n)} \tag{1}$$

S is the similarity result, E is the edit distance, m and n the length of the two strings A and B respectively, and $\max(m,n)$ is to take a maximum of two numbers. The larger the similarity value, the more similar the two strings are. Similarity value ranges of $[0,1]$; the similarity of 0 mean nothing alike, and the similarity of 1 is entirely identical.

In this study, we calculate the similarity of headlines about terrorist-attack news along the One Belt One Road and compare with words of two headlines. This research calculates the edit distance in an array according to the improved edit distance algorithm by the Python programming.

2.2.2. The Longest Common Subsequence Algorithm

Longest Common sub-sequence (LCS) algorithm is to respectively remove zero or more characters of two strings, without changing the rest of the characters to get the same longest sequence of characters. For example, there are two strings $A = \{abcdefg\}$ and $B = \{cdefghi\}$. The sequence $\{cde\}$ is the longest common subsequence of two strings. Based on the similarity calculation between this array, this study will be an array as A sequence. We set $A = \{a_1, a_2, \cdots, a_3\}$, $B = \{b_1, b_2, \cdots, b_n\}$, and the longest common sequence A and B is $Z = \{z_1, z_2, \cdots, z_k\}$. The length of Sequence A is m, the length of Sequence B is n, $C_{m\times n}$ represents the LCS matrix A and B sequence. The dynamic programming model is [11]:

- When $a_m = b_n, z_k = a_m = b_n$ and z_{k-1} is the longest common sub-sequence of a_{m-1} and b_{n-1}
- When $a_m != b_n, z_k != a_m$ and z is the longest common sub-sequence of a_{m-1} and b
- When $a_m != b_n, z_k != b_n$ and z is the longest common sub-sequence of a and b_{n-1}

According to the nature of the recursive formula:

- When $m = 0, n = 0, C[m][n] = 0$
- When $m, n > 0$ and $a_m = b_n, C[m][n] = C[m-1][n-1] + 1$
- When $m, n > 0$ and $a_m = b_n, C[m][n] = \max\{C[m][n-1], C[m-1][n]\}$

The longest common sub-sequence length of A and B is recorded in $C[m][n]$ and Z is obtained as the longest common subsequence.

2.2.3. Co-Occurrence Analysis

Co-occurrence analysis is an analytical approach to quantify the co-occurrence information in all kinds of information carrier [18]. Through the study of co-occurrence, we could find the affinity and disaffinity relationship between the research objects to exploit the connotative or potentially useful knowledge. The co-occurrence refers to events/situations that happened simultaneously or events/circumstances related to each other. The interrelation of things is the internal reason for the co-occurrence, while the co-occurrence is the external performance of the interrelation of elements.

By building the co-occurrence matrix, this study can be more intuitive to show the degree of relationship between objects. Each number in the Matrix means the number of two objects' co-occurrence; the larger the number, the closer two objects' relations are. Besides, an object itself doesn't have the co-occurrence relation, so the number on the diagonal matrix is 0.

2.2.4. Social Network Analysis

Social network analysis is used widely in the social and behavioral sciences derived from the method of sociology. It is the process of investigating social structures through the use of networks and graph theory. It characterizes network structures in terms of nodes (individual actors, people, or things within the web) and the ties, edges, or links (relationships or interactions) that connect them. So, a relatively stable social structure for interaction networks is formed. The network nodes are people, groups, organizations, or other knowledge processing entity; the link is the relationship between nodes or flow. It will quantify the relationship between individual actors, focusing on the relationship between them rather than personal characteristics [19-20].

In social network analysis, "degree" refers to the number of nodes connected by edges. The larger the number of connections, the bigger the "degree" is. In this study, due to the complex and special data we collected, we only analyze and measure the "degree" of nodes.

2.3. Data Processing

Due to the existence of semi-structured or unstructured data, this study adopts the strategy of a combination of programming and artificial methods to undertake data cleaning and processing [21]. By statistical analysis of the errors produced, problem data mainly appear in the three reference sources. The specific sample errors are shown in Table 1.

After the twice cleaning and correction for the data, we built up media-organization-name word list by selecting appropriate and correct keywords from media organizations. Then, we use the Python program to conduct frequency statistics of media organizations. As a result, 3917 groups of data are obtained, and the cleaning effect is remarkable.

Table 1. Error data and expected data sample

Original data	Expected data
Iran News Agency, April 15, 2004.	Iran News Agency, April 15, 2004.
"Blasts in New Delhi kill 55", CNN, October 30, 2005	"Blasts in New Delhi kill 55", CNN, October 30, 2005
"South Africa", Toronto Star, August 15, 2001.	"South Africa", Toronto Star, August 15, 2001.
Xinhua News Agency (Beijing), Beijing Xinhua, Xinhua, etc.	Xinhua News Agency
BBC Urdu, BBC London, BBC Monitoring South Asia, etc.	BBC
http://news.xinhuanet.com/english/2009/07/13/content_11698085.htm	Xinhua News Agency
Agency France-Press, Agency French Press, etc.	Agency France-Press

2.4. Research Design

Based on the foundation of data processing, this study conducts the similarity analysis of news headlines of the same attacks provided by different media institutions. Our goal is to explore whether media organizations have a preference to mimic headlines with each other. Firstly, this research combines the edit distance similarity algorithm and the longest common subsequence algorithm. Edit distance similarity reflects the similarities of news headlines, and the longest common subsequence is used to represent the general contents of two news headlines, which are aid analysis. Secondly, given a terrorist attack with a reference ranging from one to three sources, this study builds a co-occurrence matrix for the media organizations through co-occurrence analysis. Through the co-occurrence matrix, we obtain co-occurrence frequency between two different media organizations and then list the media organizations in descending order by high co-occurrence frequencies. Finally, according to the above results, combined with geographic information provided by the event occurs from original data, we represent geographical positions for groups of media organizations to explore whether there is a positive correlation between media institutions and the distributions of the attacks.

3. Results and Discussion

3.1. Terrorist Attacks along the One Belt One Road

This study extracted 74186 terrorist-attack events that happened along the One Belt One Road in between 2001 to 2015 from GTD, and drew the frequency distributions and geographical distributions of countries where there are more than 100 terrorist attacks. Figure 1 shows that 66 countries along the One Belt One Road have almost happened terrorist attacks from 2001 to 2015. Besides, the high-frequency areas where terrorist attacks occurred are mainly distributed in west Asia, central Asia, and South Asia. Iraq is the largest number of terrorist attacks with a total of 18599 times, Pakistan is the second number of terrorist attacks with 10902, Afghanistan is the third with 9552 terrorist attacks, and India is the fourth with 6721 times.

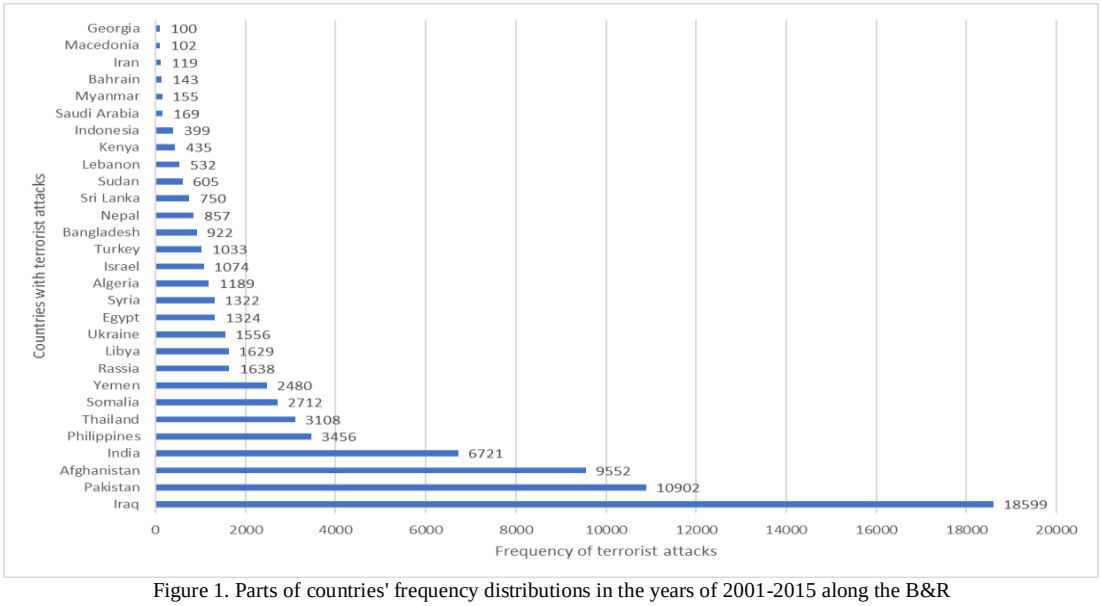

Figure 1. Parts of countries' frequency distributions in the years of 2001-2015 along the B&R

Iraq as a country with the largest number of terrorist attacks; however, the number of terrorist attacks in other countries around Iraq is relatively small, such as Syria with 1322 times, Iran with 119, and the United Arab Emirates with 169 times. This happened because of the establishment of religious extremist groups of the Islamic State of Iraq and Syria (ISIS) and

their activities. Similarly, Afghanistan in the second frequencies and Pakistan in the third as the cradle of AL-Qaeda are also influenced by ISIS. India ranked fourth may be due to the unsolved domestic religious conflicts after India Independence. Hindus accounted for over 80% of the total population, but there are still more than 150 million Muslims. Therefore, the hatred caused by religious opposition becomes an incubator for terrorism.

3.2. The Frequency Distribution of Media Organizations Reporting Terrorist Attacks along the One Belt One Road

In this study, media organizations in three groups of cited sources were counted. Media organizations with an occurrence frequency of more than 1000 were intercepted for statistics and analysis. The frequency distribution diagram is shown in Figure 2. The top four most frequently used media outlets were Lexis Nexis 9304 times, OSC Summary 9297 times, Agence France Presse 8714 times, and Xinhua News Agency 8326 times. Figure 2 shows that the top of diversified media organizations, including world mainstream media such as The French News Agency, Reuters (Reuters), AP (Associated Press), Xinhua News Agency, etc. It also shows local representative media organizations such as The PNA (Philippines News Agency, namely The Philippines News Agency), Lahore Daily Times (Lahore), and AIP (Afghan Islamic Press). Combined with the geographical distribution of terrorist attacks, some extent, the figure also reflects the reasons for the high occurrence of these regional media organizations.

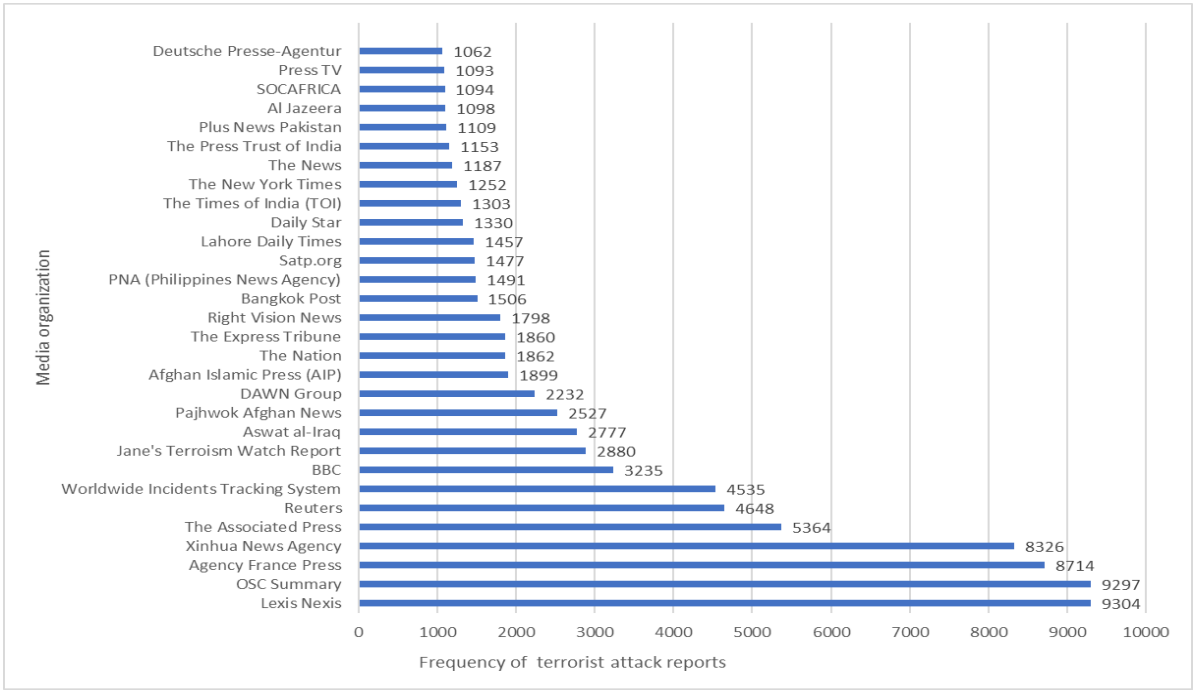

Figure 2. Frequency distribution of media organizations for terrorist attacks along the B&R

In addition, this paper calculated the similarity of each two headlines of the three groups of cited sources, respectively. By using the edit distance similarity algorithm and the longest common subsequence algorithm, the similarity histogram was obtained (see Figure 3).

According to the data, the similarity of 103891 news headlines was counted. The similarity range of editing distance of headlines is concentrated in the interval of [0,0.5], with a total of 97881pieces, accounting for 94.21% of the total. The similarity of editing distance is 1219 headlines in the interval of [0.95,1], accounting for 1.17%. 1219 headlines have the similarity of 1. For the interval of [0,0.5], it can be seen that most media organizations are relatively independent in the naming of titles, and the similarity range is mainly distributed in this interval. The reason is that each terrorist attack event has its keywords, which are necessary fields to describe a single event and cannot be replaced or omitted.

Based on the statistical similarity of editing distance, the longest common subsequence algorithm was used to work out the parts with the same title of each group of news within the similarity interval of editing distance [0,0.5]. Table 2 shows partial results. When edit distance similarity is relatively more significant, the content of the longest common subsequence displays much more detailed, and parts of speech are more diversified, including nouns, verbs, adjectives and so on. Nouns mainly represent time, location, targets and relative figures, such as "27 May", "Moscow", "Kashmir", "suicide attack", 4

Afghan" and so on. Verbs mainly mean attack modes, such as "killed", "kills"; adjectives refer to directions, such as "southern", etc. When edit distance similarity is relatively smaller, the content of the longest common sub-sequence mainly represents geographical nouns and directional prepositions, such as "Iraq", "Somalia", "on" and "in", etc. These characteristics are concise with definite requirements.

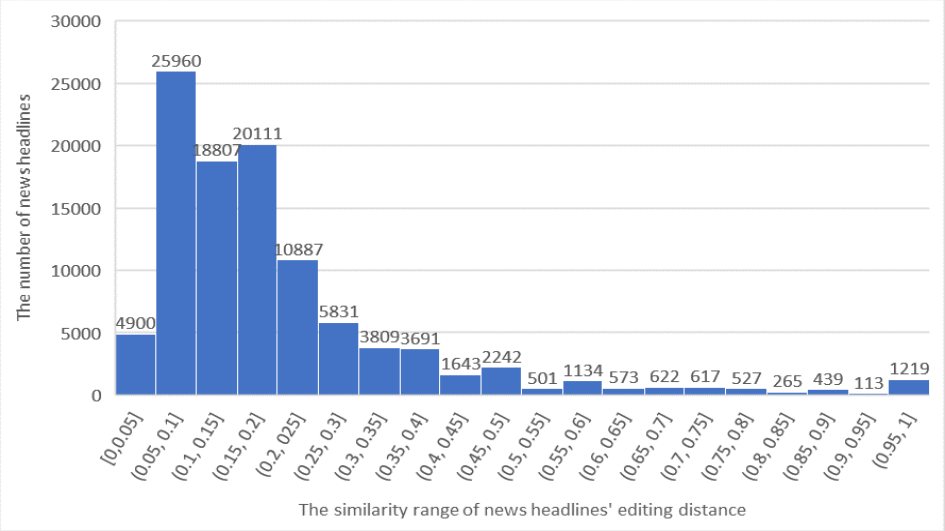

Figure 3. News headlines' similarity histogram based on edit distance algorithm

Table 2. The correlation between the similarity in [0, 0.05] interval and the longest common subsequence

No.	Edit distance similarity	LCS	The content of the longest common subsequence
1	0.5	4	Russian deputy killed Moscow
2	0.5	7	Police officers Killed in Suicide Attack in Kashmir
3	0.4444	3	kills 4 Afghan
4	0.3	2	civilians 16
5	0.25	3	in southern Thailand
6	0.25	4	Iraq Events 27 May
7	0.1818	1	on
8	0.0769	1	Somalia
9	0.0714	1	in
10	0	1	Chechen

According to the results of the longest common subsequence algorithm, there are 1219 news headlines with a similarity of 1. Some news headlines with the similarity of 1 are extracted, as shown in Table 3. For the interval of [0.95,1], we found that the headlines' contents are identical with each other, which may be because of the following two reasons: the phenomenon of direct copying in the naming of the headlines or the use of news report body, or the naming habits of journalists. It's a coincidence that the two headlines are the same.

Table 3. Partial results of the longest common sub-sequence with similarity

Edit distance similarity	LCS	The content of the longest common sub-sequence
1	9	Five Policemen Killed in Landmine Blast in Southern India
1	8	Insurgents kill 17 people including VP s guards
1	11	PWG on Rampage in AP Blasts Factories of CM Federal Minister
1	5	Shooters kill five in Baramulla
1	9	Chechen Web Site Reports Fierce Clashes Southeast of Groznyy
1	7	Police Avalanche Kills Four Rebels in Kashmir
1	6	Twelve injured in Pakistan bomb explosion
1	7	Two killed in the attack in western Afghanistan
1	8	MSF compound attacked in southern Sudan four killed

3.3. Co-Occurrence Analysis of Media Organizations

The name of processed media organizations would be correlated according to whether they report on the same terrorist

attack. If the same terrorist attack was reported by multiple media organizations, the incident would be included in the GTD database. Then, we obtain co-occurrence graph of media organizations. The co-occurrence matrix of media organizations with a length of 3917×3917 was constructed based on the list of the processed media organizations and the associated data. The degree of media organizations in the matrix was calculated by using the software Gephi (see Table 4), and the co-occurrence graph of some media organizations was drawn, as shown in Figure 4.

Table 4 shows media organizations when the degree is $\geq$200. Figure 4 represents the co-occurrence graph of media organizations when the degree is $\geq$100. The higher the degree of the node, the larger the node is rendered; the bigger the label font of the node, the darker the color of the node. The line between node and node represents the frequency of co-occurrence. As a result, we found that the top 4 degrees are Agence France Presse, Xinhua News Agency, the Associated Press, Reuters, and the British Broadcasting Corporation respectively. The number of connected edges is 1028, 834, 793, 588 and 522. We found that the rest of the media organizations apart from Lexis Nexis, the OSC Summary and Worldwide Incidents Tracking System are all listed in the top 8 in terms of the number of terrorist attacks reported. Therefore, the media organizations with more reporting terrorist-attack frequencies have some correlation with the frequencies of other media agencies which reported the same attack.

Table 4. The degree of media organization node

Media Organization	Degree	Media Organization	Degree
Agence France Presse	1028	AFP	248
Xinhua News Agency	834	Gulf News	246
The Associated Press	793	Deutsche Presse-Agentur	239
Reuters	588	Daily Star	232
BBC	552	Telegraph	228
OSC Summary	482	The Nation	225
Lexis Nexis	371	Al Jazeera	211
The New York Times	347	Daily News	209
Jane's Terrorism Watch Report	284	Pajhwok Afghan News	207
Worldwide Incidents Tracking System	283	UPI	204
CNN	257	Afghan Islamic Press (AIP)	201
PNA (Philippines News Agency)	251	The Independent	200

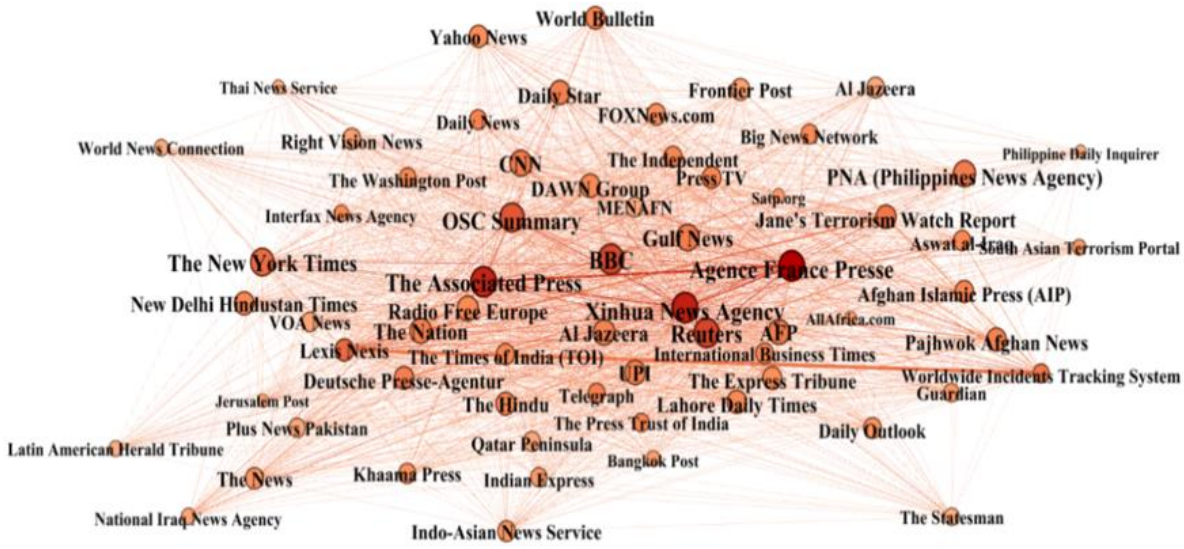

Figure 4. Co-occurrence graph of partial media organizations

3.4. The Co-Occurrence Analysis of Media Organizations' High Similarity

In this study, we took advantage of the co-occurrence analysis on the correlation between 1219 headlines with a similarity of 1 and their media organizations. We obtained the co-occurrence matrix of 350×350 for media organizations with the high similarity of headlines. Then, Gephi software was used to draw the co-occurrence graph of media organizations, and the degree of each node was counted (see Figure 5). In the high similarity interval, six media organizations have the most co-occurrence frequencies (see Table 5), namely Agence France Presse, the Associated Press, Plus News Pakistan, Xinhua

News Agency, Reuters and the Daily Star. Their degree is 32, 24, 18, 16 and 14 respectively. In addition to the co-occurrence of the world's mainstream media, some regional media organizations co-occurred, such as the Philippines News Agency, the Press Trust of India and News Track India, etc. Some of the low-frequency media agencies also appeared, for example Big News Network, the Frontier Star, Yahoo News, Jerusalem Post, Daily Pak Banker, and FOXNews.com. However, their degree is relatively lower than that of the mainstream media institutions.

According to Table 5 and Figure 5, this study found that media organizations in high similarity interval have the following features. (1) The co-occurrence frequency of single world mainstream media organization is more than that of regional media organizations, (2) Regional media organizations are basically located in areas with high incidence of terrorist attacks, such as Plus News Pakistan located in Pakistan, PNA in the Philippines and the Press Trust of India, News Track India and Hindu in India. (3) The co-occurrence frequency of relatively small media organizations with regional characteristics is lower than that of mainstream media organizations. (4) The co-occurrence frequency between the world's mainstream media organizations is rare. However, there is a copying phenomenon from the regional media organizations in the reports between the world's mainstream media with high similarity.

Table 5. Media organizations' node degree of headlines with high similarity

Media Organizations	Degree	Media Organizations	Degree
Agence France Press	32	Yahoo News	8
The Associated Press	24	Jerusalem Post	7
Plus News Pakistan	18	Lexis Nexis	7
Xinhua News Agency	16	Daily The Pak Banker	7
Reuters	14	FOXNews.com	7
Daily Star	14	News Track India	7
Big News Network	13	All Voices	7
The News	11	ZeeNews.com	6
PNA (Philippines News Agency)	10	The Hindu	6
The Frontier Star	10	DAWN Group	6

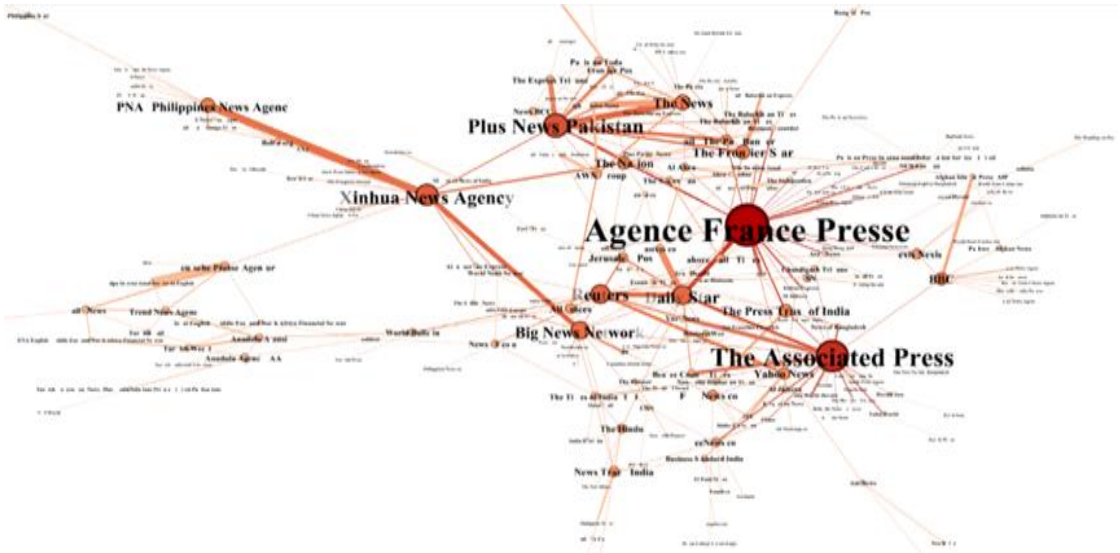

Figure 5. Media organizations' co-occurrence graph of news headlines with high similarity

The reason for these characteristics may be that media organizations with a regional nature are close to the place where terrorist attacks take place, and the speed of news transmission is much faster than that of foreign mainstream media organizations, with high authenticity and strong reliability. To timely report a terrorist attack, some mainstream media organizations will directly carry the news of some smaller media organizations of a regional nature. However, mainstream media organizations mainly report events on a global scale, which leads to the phenomenon that the co-occurrence frequency of mainstream media in the world is higher than that of regional media organizations.

4. Conclusions

According to the research results, we know that the reports of various media organization about terrorist attacks covered 66

countries along the One Belt One Road. Terrorist attacks frequently occur in west Asia, central Asia, and south Asia. In addition, the geographical distribution frequency of events reported by media organizations is positively correlated with the geographical frequency of events, that is, the more times terrorist attacks occur, the more times they are reported. Among them, the world's mainstream media reported more on terrorist attacks, such as Lexis Nexis, OSC summary, Agence France Presse, Xinhua News Agency, Reuters and so forth. However, some small media organizations also reported much more on terrorist attacks, for example Philippines News Agency, Lahore Daily times and Afghan Islamic Press. The reason for this may be that most of these media organizations with regional features are in areas with a high incidence of terrorist attacks.

In the presence of headlines' similarity, most media organizations tend to avoid repetition when naming an event. Therefore, most of the similarity of headlines would concentrate on $[0, 0.5]$ interval. The main reason may be that the headlines of terrorist attacks basically independent and due to that each event has its own keywords, such as time, place, the object of attack, etc. These keywords are the necessary fields to depict a single event and cannot be substituted or omitted. While the headlines' similarity is 1, the reason may be that a media organization directly copies the reports from another media organization, or news reporters with similar naming rules and habits use as few keywords as possible to describe an event. So, it's a coincidence that the two headlines are identical.

In regards to the co-occurrence analysis on media organizations, the features of the degree distribution are similar to the frequency distribution of media organizations' reports on events. Overall, the degree of the world's mainstream media organizations is high, but the number of mainstream media organization is small. While the degree of media organizations with regional nature ranks in the middle and bottom, the number of them is more than that of mainstream media organizations in the world. This distribution features also exist in the interval with the similarity of 1.

The geographical distribution of terrorist attacks reported high similarity. The results show that the mainstream media organizations always imitate or copy news reports of media organizations with regional natures where terrorist attacks often happened. The geographical distribution of terrorist attacks shows that there is a positive correlation between the geographical distribution frequencies of terrorist attacks and that of media organizations in areas prone to suffering from terrorist attacks. This paper can be a golden reference for early warning and guarding against hazards of terrorist attacks along One Belt One Road through features of media coverage about terrorist attacks.

Acknowledgments

The authors thank Wang Dongbo for helpful feedback on this project. The lead author also thanks to the National Natural Science Foundation of China (No.11705089) and the National Social Science Foundation of China for grants (No.15BTQ058). At last, the authors are grateful to all the anonymous reviewers for their valuable comments and suggestions.

References

1. A. Basu, "Social Network Analysis: A Methodology for Studying Terrorism," *Intelligent Systems Reference Library*, Vol. 65, pp. 215-242, 2014
2. Q. J. Liu and J. C. Fang, "The New Development of Terrorism and Its Impact on China," *International Problem Study*, Vol. 4, pp. 114-126, 2015
3. Y. T. Gong, "The Analysis of the New Dynamics of Terrorist Attacks along the Belt and Road," *Party Policy Study*, Vol. 2, pp. 18-26, 2016
4. X. L. Zhang and E. Zhang, "The Risk of Terrorist Attacks and Chinese Countermeasure for the Initiative of the Belt and Road," *International Trade*, Vol. 3, pp. 27-32, 2016
5. J. Falkheimer and E. Olsson, "Depoliticizing Terror: The News Framing of the Terrorist Attacks in Norway, 22 July 2011," *Media, War and Conflict*, Vol. 8, pp. 70-85, 2015
6. H. Zhang and D. Y. Sun, "Study on the Terrorist Cover Networks from the Perspective of Social Network Analysis," *Journal of Safety and Environment*, Vol. 3, pp. 259-264, 2011
7. M. Figueres-Esteban, P. Hughes, and C. V. Gulijk, "Visual Analytics for Text-based Railway Incident Reports," *Safety Science*, Vol. 89, pp. 72-76, 2016
8. H. Lida, J. G. Chen, H. Y. Yuan, and Y. Wang, "Terrorist Attack Vulnerability Analysis of a Natural Gas Network based on the Attacker - Defender Game Model," *Tsinghua Science and Technology*, Vol. 6, pp. 609-613, 2017
9. N. Ruigrok and W. V. Atteveldt, "Global Angling with a Local Angle: How U.S., British, and Dutch Newspapers Frame Global and Local Terrorist Attacks," *The International Journal of Press/Politics*, Vol. 1, pp. 68-90, 2007
10. Z. C. Wang, N. Wang, and Y. Wang, "Research of Title Party News Identification Technology based on Topic Sentence Similarity," *New Technology of Library and Information Service*, Vol. 11, pp. 48-53, 2011
11. L. Xin, R. Gao, and L. Song, "Internet News Headlines Classification Method based on the N-Gram Language Model," in *Proceedings of International Conference on Computer Science and Information Processing*, pp. 826-828, 2012
12. B. Han, P. Cook, and T. Baldwin, "Lexical Normalization for Social Media Text," *ACM Transactions on Intelligent Systems*

and Technology, Vol. 1, pp. 1-27, 2013
13. B. Erik, N. Hasher, and A. Rana, "The Effect of Terrorist Events on Media Portrayals of Islam and Muslims: Evidence from New York Times Headlines, 1985-2013," *Ethnic and Racial Studies*, Vol. 9, pp. 1-19, 2015
14. T. Einar and J. Daniel, "Seven Characteristics Defining Online News Formats," *Digital Journalism*, Vol. 7, pp. 1-22, 2018
15. M. Grandjean, "A Social Network Analysis of Twitter: Mapping the Digital Humanities Community," *Cogent Arts and Humanities*, Vol. 1, pp. 117-128, 2016
16. T. T. A Nguyen and S. Conrad, "An Improved String Similarity Measure based on Combining Information-Theoretic and Edit Distance Methods," in *Proceedings of International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, *Springer International Publishing*, Vol. 51, pp. 228-239, 2014
17. M. I. Nasab and R. Javidan," A New Approach for Finding Semantic Similar Scientific Articles," *Journal of Advanced Computer Science and Technology*, Vol. 1, pp. 53-59, 2015
18. L. L. Rong, Y. Y. Cai, and D. Wang, "Relations of Emergency Events Occurring in China based on the Co-Occurrence Analysis," *Systems Engineering*, Vol. 6, pp. 1-7, 2011
19. E. Otte and R. Rousseau, "Social Network Analysis: A Powerful Strategy, also for the Information Sciences," *Journal of Information Science*, Vol. 6, pp. 441-453, 2002
20. B. X. Fang, Y. Jia, and Y. Han, "Social Network Analysis — Key Research Problems, Related Work, and Future Prospects," *Bulletin of Chinese Academy of Sciences*, Vol. 2, pp. 187-198, 2015
21. A. F. R. Al-Masoudi and H. S. R. Al-Obeidi, "Smoothing Techniques Evaluation of N-Gram Language Model for Arabic OCR Postprocessing," *Journal of Theoretical and Applied Information Technology*, Vol. 3, pp. 432-439, 2015