Risk Analysis of International Spreading in 2014 Ebola Outbreak to China Compared to Social Media

Xiao LongDeng
Beijing University of Post and Telecommunication, China
shannondeng@bupt.edu.cn

Hui Zhang
Tsinghua University
zhhui@tsinghua.edu.cn

Ya Qi Tang
Beijing University of Post and Telecommunication
tangyaqi_andrea@163.com

Le Yi Ren
Beijing University of Post and Telecommunication
renleyi@bupt.edu.cn

ABSTRACT

The 2014 West African EbolaOutbreak raisedfrom Sierra Leone, Guinea, and Liberia at December in 2013 has been reported to cause 21296 cases and 8429 deaths until now which became the deadliest recorded in history. In this paper, we proposed the riskanalysis to assess the international spread risk from mentioned three African countries to China by GEM(Global Epidemic Mobility) Model. As another part of analysis, we crawled related online social media data of Ebola from the most four favorite online social networks (including SINA, TENCENT) in China from June to November in 2014. By analyzing these attained social media data and airline data of GEM, we found some interesting results. For example, Beijing has the most importing risk of Ebola while it has the hottest discussion on social network.. Furthermore, we showed analysis of combining social network data with geographicaldemonstration and Chinese citizen sentiment towards this disaster.

Keywords

Riskanalysis, Ebola, Airline Data, Social Media, Online Opinion Analysis

INTRODUCTION

In the last decades, there have been much more public health crises in the world such as H1N1,H7N9 andEbola outbreak. And it has been proved that our world has come into the time when public crisis accidents number was growing fast.At the same time, with the fast increasing of Internet users and data scale, Internet is playing a more important role in information spreading. And nowadays, the online public opinionis becoming an important issue which is crucial to the country's stability and development of people's livelihood.

Facing the international spreading risk in 2014 West African Ebola (WHO, 2014) to China,We used the GEM Model(Global Epidemic and Mobility Model) (Balcan D, Hu H, Goncalves B, Bajardi P, Poletto C,2009;Balcan D, Colizza V, Goncalves B, Hu H, Ramasco JJ, 2009; Balcan D, Goncalves B, Hu H,Ramasco JJ, Colizza V, 2010;Legrand J, Grais R, Boelle P, Valleron A, Flahault A ,2007)to generate stochastic, individual based simulations ofepidemic spreading from Sierra Leone, Guinea, and Liberia to China.Our computing method is focused on the mobility model which integratesdaily airline passengers worldwideespecially because the epidemic is now affecting cities withmajor commercial airports and our computing method has found out the ranking riskamongmain cities in China.

In this paper, we proposed theriskanalysis to assess the international dissemination risk from the mentioned three West African countries to China by GEM Model. As another part of our analysis, we madeadetailed analysis and estimation of discussion data on main Chinese online social media data(including SINA Micro Blog, TENCENT Micro Blog, SOHU MicroBlog and NETEASE Micro Blog)from June to November in 2014.And our analysismethod combining airline with online opinion data in social network and geographical information has drawn some new light on risk analysis and estimation of global emergency events response management. In detail, our contributions are as follows:

1. We proposed a detailed introduction of riskanalysis to assess the international dissemination risk from the mentioned three West African countries to China by GEM Model;

2. We systematically collected all the airline data from Sierra Leone, Guinea, and Liberia to China to make the detailed Ebola international spreading risk analysis(IATA,2014; OAG,2014);

3. We systematically collected related online social media data of Ebola from the most four favorite online social networks(including SINA Micro Blog, TENCENT Micro Blog, SOHU MicroBlog and NETEASE Micro Blog) in China from June to November in 2014 which consists of more than one hundred and eighty thousand Micro Blog records.

4. We systematically analyzed the ranking risk of Ebola importing to main Chinese cities and the online opinion data from the mentioned four favorite online social networksin China, and we found it is very interesting that social network users in Beijing always havethe most hottestdiscussions of Ebola while Beijing is facing the most highest Ebola importing risk from Sierra Leone, Guinea, and Liberia.

The outline of the paper is as follows: Section 2 surveyed the related work of Ebola international spread riskand GEM Model (Global Epidemic and Mobility Model). In Section 3, we introduced our method. In Section 4, we proposed the analysis result and our findings.And Section 5 gives a conclusion.

RELATEDWORK

The 2014 West African Ebola Outbreak is the largest ever observedepidemicdisaster in the new century, both by number of cases and geographical extension. Forthis reason, on 6 August 2014, an Emergency Committee of the WHO(WHO Statement on the Meeting of the International Health Regulations Emergency Committee , 2014)advised the 2014West African Ebola outbreak constitutesan ’extraordinary event’ and a public health risk to other States. For the reason that the outbreak region in Guinea has bus and train connected to big cities(Conakry, Freetown, Monrovia and Lagos), and there already have been Ebola cases in the four mentioned cities which have airlines connected to other countriesin the world. Theinternational spread risk is focused on the spreading of airline passengers. Gomes(Gomeset al., 2014)used the Global Epidemic and Mobility Model that integrates highresolutiondata on human demography and mobility on a worldwide scale by a metapopulation stochastic method.The method usesMonte Carlo likelihood analysis considering more than 1,000,000 simulations which sample the diseasemodel space and the data on the 2014 West African Ebola outbreak up to 9 August 2014. This approach selects the diseasedynamic model that we used to generate numerical stochastic simulations of an epidemic’s local (within WestAfrican countries) and global progression. The Global Epidemic and Mobility Model haveintegrated 3,362 subpopulations in 220 countries data on human demography and mobility.Dirk and Lars (Dirk and Lars,2014) used conditional probabilities method to analyze the relative import risk of Ebola to the other 1227 largest airports in the world to allow better rendering and avoid clutter. And furthermore, they have developed a visualization tool(http://rocs.hu-berlin.de/D3/ebola/) to show their research results. And there are also some other researcher’s findings.Towers and Patterson(Towers and Patterson,2014)researched temporal variations in the effective reproduction number of the 2014 West AfricanEbola Outbreak,and Nishiura (Nishiura and Chowel,2014)found some interesting result of early transmission dynamics of Ebola. But the mentioned research results above have not made a unified analysis of Ebola international spread risk and the online opinion data of the destination area or countries, so we will propose our method and make a unified analysis basing on Chinese online opinion data.

RISK ANALYSIS OF 2014 EBOLA TO CHINA

International SpreadRisk Estimation Model

The GEM(Global Epidemic and Mobility) model integrates social demographic and population mobility data in a spatially structured stochastic disease approach to simulate the spread of epidemics on the worldwide scale which includes local spreading(bus, train and other short journey transmissions) and remote spreading(mainly referring airlines).In our method, based on the basic Epidemic and Mobility method of GEM Model, we proposed our method to compute the possible airline passengers travelingto three biggest cities in China which are Beijing, Shanghai and Guangzhou in the following formula (1):

$$
\begin{aligned}
N_{estimate} = & \prod_{line1-of-N_{source1}} N_{source1} \times \alpha_{l_1\tau_1N_1} \times \alpha_{l_1\tau_2N_1} \times \alpha_{l_1\tau_3N_1} \times \alpha_{l_1\tau_iN_1} \\
& + \prod_{line2-of-N_{source1}} N_{source1} \times \alpha_{l_2\tau_1N_1} \times \alpha_{l_2\tau_2N_1} \times \alpha_{l_2\tau_3N_1} \times \alpha_{l_2\tau_jN_1} + \ldots \\
& + \prod_{line1-of-N_{source2}} N_{source2} \times \alpha_{l_1\tau_1N_2} \times \alpha_{l_1\tau_2N_2} \times \alpha_{l_1\tau_3N_2} \times \alpha_{l_1\tau_iN_2} \\
& + \prod_{line2-of-N_{source2}} N_{source2} \times \alpha_{l_2\tau_1N_2} \times \alpha_{l_2\tau_2N_2} \times \alpha_{l_2\tau_3N_2} \times \alpha_{l_2\tau_jN_2} + \ldots
\end{aligned}
\quad (1)
$$

In formula (1),For example, there are several airlines from Sierra Leone ($N_{source1}$), Guinea ($N_{source2}$), and Liberia ($N_{source3}$) to Beijing and $\prod_{line1-of-N_{source1}} N_{source1} \times \alpha_{l_1\tau_1N_1} \times \alpha_{l_1\tau_2N_1} \times \alpha_{l_1\tau_3N_1} \times \alpha_{l_1\tau_iN_1}$ determines thepossible number of airline passengers who would reach Beijingfrom airline $line1$ with the originalpassengers $N_{source1}$. $\alpha_{l_1\tau_1N_1}$, $\alpha_{l_1\tau_2N_1}$ and $\alpha_{l_1\tau_iN_1}$ determine the rest passenger percentageof the i relay transit airports in $line1$ from Sierra Leone to Beijing. In our estimation, they are all assigned value 50% for simplification. And $N_{estimate}$ is the total number of possible airline passengers from SierraLeone to Beijing in our estimation. As we can see that formula (1) indeed stands for the airline path matrix from the three West African countries to the destination cities.

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i\tau_1N_i}$	Result
Guinea	Casablanca-(Doha,Frankfurt,Paris Charles De Gaulle,ParisAudiocodes,Dubai)-Beijing	AT526	B737	145	0.5*0.5	36.25
	Paris Charles De Gaulle- Beijing	AF727	A330	295	0.5	147.5
	Paris Charles De Gaulle- Amsterdam - Beijing	KL2209	A330	295	0.5*0.5	73.75
	Paris Charles De Gaulle- Zurich -Beijing	AF755	A330	295	0.5*0.5	73.75
	Brussels- Beijing	SN1255	A330	295	0.5	147.5
	Lufthansa- 2 unfixedAirport- Beijing	LH5549	A330	295	0.5*0.5	73.75
	KLM- 2 unfixedAirport - Beijing	KL2293	A330	295	0.5*0.5	73.75
	Air France- 2 unfixedAirport- Beijing	AF727	A330	295	0.5*0.5	73.75
	Delta Airlines- 2 unfixedAirport- Beijing	DL8625	A330	295	0.5*0.5	73.75
					SUM	**773.75**

Table 1. Airline Passengers from Guinea to Beijing

(*.Result: the estimated number of passengers;KLM:Netherlands Airline

Company)

The last four row of Table1 stands for the airlines with unfixed airports and other rows of Table1 stands for the airlines with fixed airports. If there is only one transit airport in the airline to Beijing, $\prod_{line_i} \alpha_{l_i \tau_1 N_i}$ will be 0.5.If there are two transit airports, this number will be multiplied by another 0.5. Because we made an assumptionthat when the flight reached a transit airport,there will be about 50% passengers continue to choose next destination airport.The maximal number of seats of Boeing737(B737) is 160, while the minimal number of seats is 130. So we usedthe average number 145.The maximal number of seats ofAirbus330(A330) is 440 while the minimal number of seats is 295.Since the most common flight has 295 seats so we used this number.

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Liberia	Brussels- Beijing	SN1247	A330	295	0.5	147.5
	Casablanca-Paris-Beijing	AT598	B737	145	0.5*0.5	36.25
					SUM	**183.75**

Table 2. Airline Passengers from Liberia to Beijing

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Sierra Leone	Brussels- Beijing	SN1255	A330	295	0.5	147.5
	Casablanca-New York-Beijing	AT596	B737	145	0.5*0.5	36.25
					SUM	**183.75**

Table 3. Airline Passengers from Sierra Leone to Beijing

Also based on analysisof Table 1 to Table 9(Table 4 to 9 can be found in APPENDIX of this article), we can find that Beijing is facing the most maximumimporting passengers of 1141.25which is SUM of Table 1 to 3,and Shanghai has the possible number of 846.25 which is SUM of Table 4 to 6, while the number in Guangzhou is 920.625 which is SUM of Table 7 to 9.

Online Social Media Data Analysis

We collected about 350 thousands records written in Chinese from four Chinese online social network including SINA Micro Blog, TENCENT Micro Blog, SOHU MicroBlog and NETEASE Micro Blog from June to November in 2014.In addition,we analyzed the developing trend of Ebola discussion in the period which can be found in the following figures:

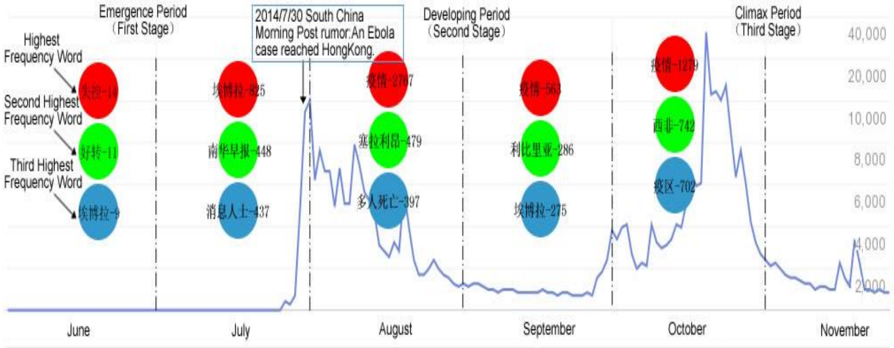

Figure 1. Hot Words Frequency from June to November in 2014

The x axis in Figure1 stands for time and y axis stands for the tweets number. It can be found in Figure 1 that the online social network opinion data of Ebola in China can be divided into three stages which are Emergence Period including June and July,Developing Period including August and September triggered by the rumor reported by the South China Morning Post which told that an Ebola case has reached HongKongon July 30th,and Climax Period including October and November.We also have counted the hot Chinese words frequency from June to November and we found that after July 30th, the rumor has caused that the Chinese word “疫情” which means epidemic situation suddenly rising to 2767 times in August and 1279 times in October with Chinese word "塞拉利昂" which means Sierra Leone and Chinese word "利比里亚" which means Liberia followedin the second place.

Figure 2. Geographical Distributionof Discussion in August2014

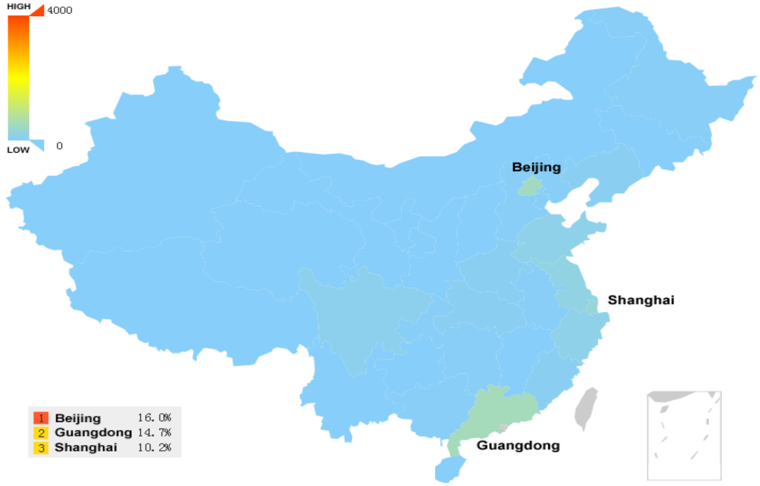

Figure 3. Geographical Distributionof Discussion in September2014

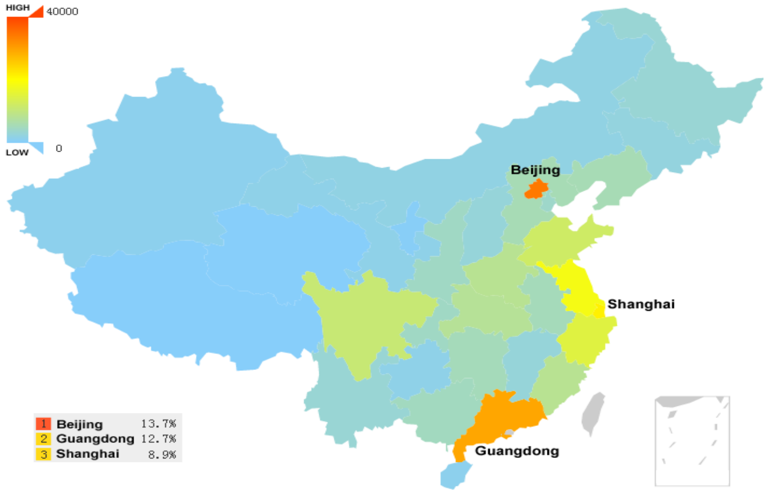

Figure 4. Geographical Distributionof Discussion in October2014

Based on the Figure 1 to 4 of geographical distribution of discussion in August to October,we can find that Beijing is always the hottest discussion city in Ebola having 14.5%,16.0% and 13.7% proportionof the whole discussion in the country with Guangdong province followed.

We also used the ICTCLAS2013 (Zhang Huaping , 2013) which segmentsChinese wordsto software. It helped us to split the collected discussion tweets from the four online social networks into single words and analyze user's sentiment towards the Ebola event.The user's sentiment is divided into three categories which are Positive, Negative and Indifferent. In detail, it is composedof different eight tendencies which are Happy, Good, Angry, Sorrow, Fear, Hate, Surprise and Indifferent. It can be seen from Figure 5 that there are 37.4% positive discussions,34.0%Indifferent discussions and 28.6%Negative discussions in China.The purple Fear curve stands for the fear of Ebola which reached the maximum in October.We can find that in spite of the climax fear in October,Chinese always hold the positive attitudes to Ebola and trust that it can be controlled.

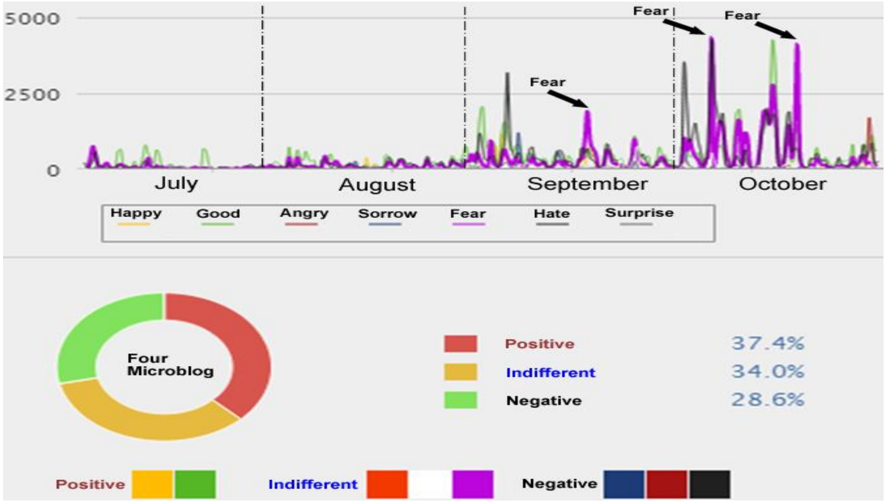

Figure 5. Sentiment Analysis from July to Octoberin 2014

In addition, the negative tweets of our disaster are extracted out and we counted and ranked the number of negative tweets in different provinces in China to study the difference sentiment of people from different provinces.Table 10 shows the results in the following

Time	The Top 4 Provinces		Negative Tweets Count
July 2014	1	Beijing	84
	2	Guangdong	86
	3	Shanghai	72
	4	Jiangsu	58
August 2014	1	Beijing	151
	2	Guangdong	128
	3	Jiangsu	88
	4	Shanghai	65
September2014	1	Beijing	164
	2	Guangdong	164
	3	Shanghai	134
	4	Jiangsu	104
October 2014	1	Beijing	9960
	2	Guangdong	9359
	3	Shanghai	6482
	4	Jiangsu	6121
November 2014	1	Beijing	4801
	2	Shanghai	3841
	3	Jiangsu	2520
	4	Guangdong	2400

Table 10. Top 4 Hottest Provinces in Negative Discussion from July to November

(*.Province Jiangsu is adjacent to Shanghai and Passenger for Jiangsu always landed in Shanghai)

It can be found in Table 10 that the top 4 provinces with negative tweets are still focused on the cities with the highest airline importing risk. Cities are includedBeijing,Guangzhou(Guangzhou is the capital city of Guangdong province),Shanghai and its adjacent province Jiangsu.And the ranking result devoted to the ranking result of Table 1 to 9.

GERGRAPHICAL DISTRIBUTION OF RISK ANALYSIS

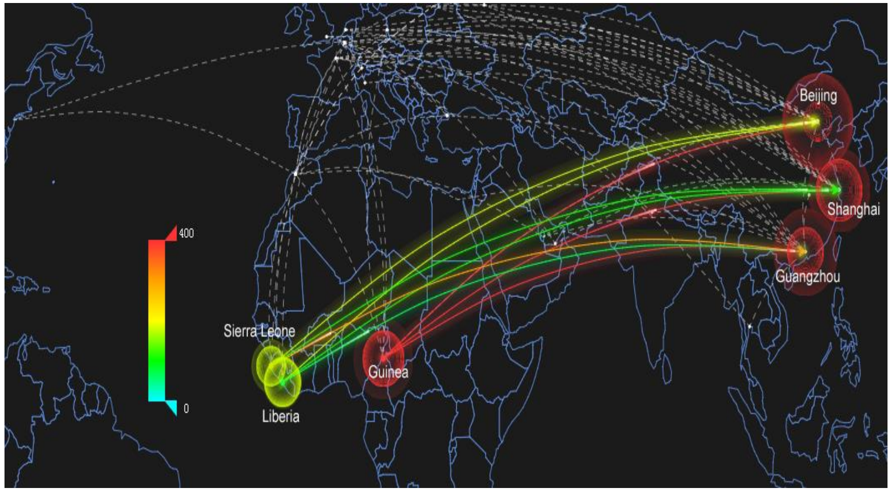

Figure 6. International Spreading Risk in Airlines to China from West Africa

Finally we drew out the importing risk of Beijing,Guangzhou and Shanghai in Echarts(http://echarts.baidu.com/)with colored airlines to stand for the number of passengers from different airlines in Figure 6.And the white line stands for the airlines started from the three West African Countries.

To our surprise, the aggregate risk analysis in this paper has showed that the city which has the highest international disease importing risk always has the hottest discussions in online social network and people in these cities may have the most fear than people in other cities and provinces. The result showed in this paper has done a very interesting research combining dynamic model(e.g.The Global Epidemic and Mobility Model)in nature science area with online opinion in social science area.

CONCLUSION

In this paper, we firstly proposed a new method based on Global Epidemic and Mobility Model to compute the Ebola international outbreak risk to China in 2014 by collectingall the airline data from Sierra Leone, Guinea, and Liberia to China.By the computation, we ranked the international importation risks in Chinese main cities.Secondly,we systematically collected related online social media data of Ebola from the most four favorite online social networks in China from June to November in 2014 which consists of more than one hundred and eighty thousand Micro Blog records.And then,we made an aggregate analysis on international importation risk and online discussion data for the same event.We found it is very interesting that social network users in Beijing always have the hottest discussions of Ebola while Beijing is facing the highest Ebola importing risk from Sierra Leone, Guinea, and Liberia.

ACKNOWLEDGMENTS

We thank all authors for their hard work and contributions to this paper.Also thanks to the support of National Natural Science Foundation of China (NNSF) (Grants No. 91124002), National Culture Support Foundation Project of China (2013BAH43F01),and National 973 Program Foundation Project of China (2013CB329600) in social network analysis.We appreciate professor Hui Zhang’s direction and his aid form Joint-Operated project from National Natural Science Foundation of China (NNSF) (Grants No. 91224008-14) from Tsinghua University.

REFERENCES

1. WHO (2014) Epidemic Pandemic Alert and Response. World Health

Organization, Regional Office for Africa. *http://www.who.int/csr/en/*

2. Balcan D, Hu H, Goncalves B, Bajardi P, Poletto C, et al. (2009) Seasonal transmission potential and activitypeaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC Medicine*7: 45.
3. Balcan D, Colizza V, Goncalves B, Hu H, Ramasco JJ, et al. (2009) Multiscale mobility networks and the spatialspreading of infectious diseases. *Proceedings of the National Academy of Sciences* 106: 21484-21489.
4. Balcan D, Goncalves B, Hu H, Ramasco JJ, Colizza V, et al. (2010) Modeling the spatial spread of infectiousdiseases: The Global Epidemic and Mobility computational model. *Journal of Computational Science* 1: 132-145.
5. Legrand J, Grais R, Boelle P, Valleron A, Flahault A (2007) Understanding the dynamics of Ebola epidemics.*Epidemiol Infect 135: 610.*
6. IATA (2014) International Air Transport Association.
7. OAG (2014) Official Airline Guide.
8. WHO (2014) WHO Statement on the Meeting of the International Health Regulations Emergency Committeeregarding the 2014 Ebola outbreak in West Africa.
9. Gomes MFC, Pastore y Piontti A, Rossi L, Chao D, Longini I, Halloran ME, Vespignani A. Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLOS Currents Outbreaks. 2014 Sep 2.doi:10.1371/currents.outbreaks.cd818f63d40e24aef769dda7df9e0da5*.
10. Dirk Brockmann, Lars Schaade, LuzieVerbeek.2014 Ebola Outbreak: Worldwide Air-Transportation, Relative Import Risk and Most Probable Spreading Routes.*http://rocs.hu-berlin.de/publications/ebola/index.html*
11. Towers S, PattersonLomba O, Castillo-Chavez C. Temporal Variations in the Effective Reproduction Number of the 2014 West Africa Ebola Outbreak. *PLOS Currents Outbreaks. 2014 Sep 18.doi:10.1371/currents.outbreaks.9e4c4294ec8ce1adad283172b16bc908*
12. Nishiura H, Chowell G. Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014. *EuroSurveill.2014; 19(36): pii=20894.*
13. Zhang Huaping.*http://ictclas.nlpir.org/*
14. Althaus CL. Estimating the Reproduction Number of Ebola Virus (EBOV) During the 2014 Outbreak in West Africa. *PLOS Currents Outbreaks. 2014 Sep 2. doi: 10.1371/currents.outbreaks.91afb5e0f279e7f29e7056095255b288*

APPENDIX

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{l\,i\,ne_i} \alpha_{l_i \tau_1 N_i}$	Result
Guinea	Casablanca-(Doha,Frankfurt,Paris Charles De Gaulle,ParisAudiocodes,Dubai,Munich)-Pudong	AT526	B737	145	0.5*0.5	36.25
	Paris Charles De Gaulle- Pudong	AF727	A330	295	0.5	147.5
	Paris Audiocodes - Pudong	AF755	A330	295	0.5	147.5
	Lufthansa- 2 unfixedAirport- Pudong	LH5549	A330	295	0.5*0.5	73.75
	KLM- 2 unfixedAirport -Pudong	KL2293	A330	295	0.5*0.5	73.75
	Delta Airlines- 2 unfixedAirport- Pudong	DL8625	A330	295	0.5*0.5	73.75
	Brussels Airlines-- 2 unfixedAirport-Pudong	SN1255	A330	295	0.5*0.5	73.75
					SUM	**626.25**

Table 4. Airline Passengers from Guinea to Shanghai

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Liberia	Brussels- (Beijing, Frankfurt,Paris Charles De Gaulle, Milan, Abu Dhabi,Amsterdam, London, Moscow, Roma)-Hongqiao,Pudong	SN1247	A330	295	0.5*0.5	73.75
	Casablanca-Paris-Pudong	AT598	B737	145	0.5*0.5	36.25
					SUM	**110**

Table 5. Airline Passengers from Liberia toShanghai

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Sierra Leone	Brussels-(Beijing,Milan)-Hongqiao,Pudong	SN1255	A330	295	0.5*0.5	73.75
	Casablanca-Paris-Pudong	AT596	B737	145	0.5*0.5	36.25
					SUM	**110**

Table 6. Airline Passengers from Sierra Leone to Shanghai

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Guinea	Casablanca-(Doha,Istanbul,Dubai)-Baiyun	AT526	B737	145	0.5*0.5	36.25
	Paris- Amsterdam-Baiyun	KL2209	A330	295	0.5*0.5	73.75
	Paris-(Istanbul,Wuhan)-Baiyun	AF755	A330	295	0.5*0.5	73.75
	Paris -Baiyun	AF727	A330	295	0.5	147.5
	KLM- 2 unfixed Airport-Pudong	KL2293	A330	295	0.5*0.5	73.75
	Delta Airlines- 2 unfixed Airport-Pudong	DL8625	A330	295	0.5*0.5	73.75
	Brussels Airlines-- 2 unfixedAirport-Pudong	SN1255	A330	295	0.5*0.5	73.75
					SUM	**552.5**

Table 7. Airline Passengers from Guinea to Guangzhou

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{line_i} \alpha_{l_i \tau_1 N_i}$	Result
Liberia	Brussels-(Beijing,Bangkok,Istanbul, Amsterdam, Moscow, New Delhi)-Baiyun	SN1247	A330	295	0.5*0.5	73.75
					SUM	**73.75**

Table 8. Airline Passengers from Liberia toGuangzhou

Source	Main Airline	Flight	Aircraft Type	Seats	$\prod_{l\,i\,ne_i} \alpha_{l_i \tau_1 N_i}$	Result
Sierra Leone	Brussels- (Beijing Bangkok,Istanbul, Amsterdam, Moscow, New Delhi)- Baiyun	SN1255	A330	295	0.5*0.5	73.75
	Casablanca- (Doha,Cairo,Istanbul)- Baiyun	AT596	B737	145	0.5*0.5	36.25
	Lufthansa- 3 unfixedAirport- Baiyun	LH5549	A330	295	0.5*0.5*0.5	36.875
	Kenya Airways- 1 unfixedAirport- Baiyun	KQ511	A330	295	0.5	147.5
					SUM	**294.375**

Table 9. Airline Passengers from Sierra Leone to Guangzhou