

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

11-2014

A first look at global news coverage of disasters by using the GDELT dataset

Haewoon KWAK

Singapore Management University, hkwak@smu.edu.sg

Jisun. AN

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Numerical Analysis and Scientific Computing Commons](#)

Citation

1

This Conference Proceeding Article is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylids@smu.edu.sg.

A First Look at Global News Coverage of Disasters by Using the GDELT Dataset

Haewoon Kwak* and Jisun An

Qatar Computing Research Institute, Doha, Qatar
{hkwak, jan}@qf.org.qa

Abstract. In this work, we reveal the structure of global news coverage of disasters and its determinants by using a large-scale news coverage dataset collected by the GDELT (Global Data on Events, Location, and Tone) project that monitors news media in over 100 languages from the whole world. Significant variables in our hierarchical (mixed-effect) regression model, such as population, political stability, damage, and more, are well aligned with a series of previous research. However, we find strong regionalism in news geography, highlighting the necessity of comprehensive datasets for the study of global news coverage.

Keywords: GDELT, global news coverage, news geography, regionalism, theory of newsworthiness, international news agency, foreign news, disaster.

1 Introduction

A wide news network woven by international news agencies helps to transform a remote disaster into an international crisis. Even though a majority of disasters remains unreported [9], the reported ones evoke compassion, and this potentially leads to various charitable acts, such as fund-raising to provide monetary support. These days it is not uncommon to expect help from the world when a tragic disaster happens. In this sense, global news coverage of a disaster is a sufficient condition for worldwide public action. Then, a central question naturally arises: which disasters are covered and which are not? The systematic approach to address this question requires a comprehensive dataset of news media sites in different countries over long period, which remains to be unexplored in traditional media research. We revisit previous research based on a single country or a region [11, 19] and examine whether it also holds globally.

In this work we use a large-scale news media coverage dataset collected by the GDELT (Global Data on Events, Location, and Tone) project. GDELT project monitors news media in over 100 languages from the whole world [14]. With large-scale data of 195 thousand disasters happening from April 2013 to July 2014 [1], we examine which disasters receive a great deal of attention from foreign news media.

* A long version of this work, “Understanding News Geography and Major Determinants of Global News Coverage of Disasters”, will be available on arXiv.

2 Theoretical Orientation

We clarify the theoretical orientation to address our research questions by reviewing previous literature. We begin with the theory of newsworthiness proposed in 1965 and proceed to following studies.

2.1 Theory of Newsworthiness

Foreign news coverage is the outcome of a news selection process [8]. Making a decision on which news items to report is essential because news is delivered through physically limited channels, such as pages in newspapers and minutes in TV news.

On one hand, one of the seminal studies that examines the factors of news selection was conducted by Galtung and Ruge [11]. They propose the theory of newsworthiness that is based on psychology of individual perception and explain which factors influence newsworthiness of an event. The suggested factors are frequency, intensity, unambiguity, meaningfulness, consonance, unexpectedness, continuity of an event, and some characteristics (e.g. identity) of an actor involved in the event. Some critics argue that applying their theory to the global news flow between nations is insufficient owing to the lack of systematic determinants based on the power structures of the world [21]. Nevertheless, the theory of newsworthiness has provided a foundation of subsequent news flow studies with a few variations of some factors [12]. We consider several factors among their suggestions, such as unexpectedness and intensity of an event and the identity of an actor extracted from the GDELT dataset.

The significance of the number of people killed by a disaster in predicting its news coverage is still debatable. Gaddy and Tanjong reported its importance [10], while others found no significance [5]. Yet, the number of victims is a good proxy to reflect the intensity of a disaster, especially when the other measures for the damage of a disaster is unavailable. We thus include the number of victims of a disaster in our study and validate its importance.

2.2 Effect of National Attributes on Foreign News Coverage

While determinants of global news flow, in terms of the amount and its direction, have been repeatedly investigated based on a single or few countries for decades [16], results are not consistent between countries, mainly due to cultural difference [17].

On the other hand, one of the earliest studies reporting general patterns of global news flow was conducted with the 'Foreign News' dataset that contains news media of 46 countries for two weeks in 1995 [18]. The findings are relatively stable, and confirmed by subsequent studies. We note that these studies are usually conducted from the view of a guest and a host country relationship. The guest country is where the event happens, and the host country is where the news media exist. In this view, the problem of global news flow is transformed into the problem of dyadic relationship between countries, like whether an event in a certain guest country is covered in a particular host country. The general factors to affect the news coverage of a host country can be divided into two categories. One is the attributes of a guest country, and the other is proximity between two countries. We focus only on the former, the national attributes of the guest

country, because our aim is to capture the global view and thus does not necessarily assume the dyadic relationship led by a specific host country.

A wide range of national attributes affecting news coverage is found across the studies [6,16], such as GNP per capita, territorial size, GDP, defense budget, population density, share in world trade, press freedom index, number of scientific publications, and Internet use. We consider all these variables. We also include some variables such as world giving index, to consider the humanitarian view of the disasters.

3 The GDELT Project

GDELT (Global Data on Events, Location, and Tone) is a recently developed event dataset containing more than 200 millions geolocated events with global coverage since 1979 [14]. GDELT began with monitoring a wide range of international news sources, including AfricaNews, Agence France Presse, Associated Press Online, Associated Press Worldstream, BBC Monitoring, Christian Science Monitor, Facts on File, Foreign Broadcast Information Service, United Press International, and the Washington Post. Now in cooperation with Google, it has expanded its sources to cover non-English news media. Today it tracks news media in over 100 languages from the whole world. After the first release of GDELT, several studies have confirmed that the GDELT dataset performs as well or better than the previously widely-used datasets, such as ICEWS (Integrated Conflict Early Warning System) due to its large coverage and the improved automatic coding system [7].

GDELT provides two types of datasets. One is called the Event Database, coded by CAMEO taxonomy since 1979, and the other is the Global Knowledge Graph (GKG), an expanded dataset about ‘every person, organization, company, location, and over 230 themes and emotions from every news report’ since 2013. We use the GKG dataset because it offers various fields to describe the characteristics of natural and man-made disasters, such as the type of a disaster, the number of news articles reporting the disaster, the number of victims, the location where the disaster occurred, and etc.

4 News Geography of Disasters

We first concentrate on news geography, the extent to which countries are represented in international disaster news. We show how different their news geographies are and then move on to examine the representativeness of each region for *global* attention. It directly relates with the external validity of previous studies about foreign news rooted on a single or several countries.

We divide the world into seven regions according to the World Bank: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and Sub-Saharan Africa. The division mainly reflects geographical proximity. We map 10,009 news media into one of the seven regions according to the classification of Alexa, which is based on the nationality of website visitors. A list of news media falling in each region becomes a basis to construct news geography seen by each region.

We define the attention of a region, r_i , to a country, c_j , as the number of the disasters occurred in c_j covered by news media of r_i . We use the notation, $N_{r_i \Rightarrow c_j}$,

for representing the attention of r_i to c_j . Then, we define news geography seen by r_i as $\mathbf{N}_{r_i} = \{N_{r_i \Rightarrow c_1}, N_{r_i \Rightarrow c_2}, N_{r_i \Rightarrow c_3}, \dots, N_{r_i \Rightarrow c_K}\}$ where K is the number of the countries.

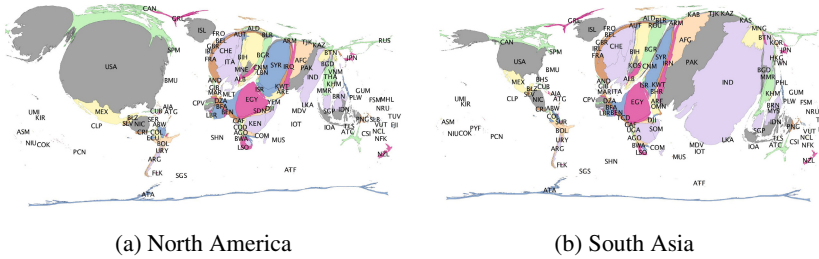


Fig. 1. News geography seen by each region

Figure 1 shows the news geography seen by North America and South Asia, respectively, through Cartogram, an intuitive visualization method of illustrating the territory of a country that is proportional to the assigned value. In the figure, the size of a territory is proportional to $N_{r_i \Rightarrow c_j}$ in the news geography seen by r_i .

By visual inspection, we observe clear differences of the news geography across the region. Although we omit other regions due to lack of space, as similar with two visualizations in Figure 1, every region is overrepresented in the news geography seen by the corresponding region. For example, disasters occurring in Latin America & Caribbean are not frequently reported in other regions. Similarly, Indonesia is well-recognized in $\mathbf{N}_{East Asia}$, Serbia in \mathbf{N}_{Europe} , Saudi Arabia in $\mathbf{N}_{Middle East}$, USA in $\mathbf{N}_{North America}$, India in $\mathbf{N}_{South Asia}$, and Kenya in \mathbf{N}_{Africa} . This strong regionalism raises concerns about the external validity of studies of foreign news coverage based on a single country or region.

At the same time, Figure 1 poses an interesting question about the over-representation of a certain country (e.g. Syria in news geography seen by North America) that cannot be explained by regionalism. This relatively regionalism-free country could be explained by the proximity in another layer, such as politics, economy or culture, instead of geographic proximity. Since the scope of this work is investigating global attention to disasters, rather than attention of a certain region or country, we do not study this further here.

5 Determinants of Global News Coverage of Disasters

5.1 Methods

We build a hierarchical (mixed-effect) multiple regression model to examine what affects global news coverage of disasters. We choose this model to control a random effect driven by variation rooted on country-level differences. Previous studies have shown that international news coverage varies significantly by country [20].

We define global news coverage of a disaster as the number of countries reporting the disaster, and use it as the dependent variable in our model. For 10,009 news media appearing in the GKG dataset, we extract the origin country of each news media from Alexa. In our data, the range of global news coverage lies between 1 and 34. The median

is 1 (mean: 1.78), indicating that a large fraction of disaster news is consumed within a single country.

We consider 26 independent variables as candidates that might exert influence on the global news coverage of a disaster according to our theoretical orientations. The variables can be organized into three broad categories: (1) the attributes of a nation which measures political and economical status of a nation; (2) the attributes of a disaster; and (3) logistics of news gathering.

We obtain national attributes for all the countries listed in the GDELT dataset from various sources, including the World Bank Open Data, which provides a wide range of up-to-date measures of 254 countries. Fifteen national variables are driven from it: GDP (gross domestic product) per capita, GNI (gross national income) per capita, military expenditure, population, land size, population density, merchandise exports (US\$), merchandise imports (US\$), number of scientific journal publications, unemployment rate, foreign aid received (US\$), Internet use (per 100 people), mobile cellular subscriptions (per 100 people), and homicide rate (per 100,000 people). In addition, we create a trade variable as the sum of the magnitude of exports and imports. While some information is not up-to-date, most of variables are reported for 2013. We additionally consider the index of press freedom [2], the world giving index [4], and the political stability index [3].

We obtain attributes of disasters directly from the GKG dataset. It marks a disaster as either man-made, natural, or both. Also, a fine-grained subtype of a disaster, such as 'Flooding' and 'Landslide', is available. We select one representative subtype for the disaster by considering the frequency of subtypes across all the disasters, while GDELT can assign multiple subtypes to one disaster. For the matter of categorical coding, we select top 30 subtypes out of 274 subtypes, which account for 31.5% of all disasters. All the other types are coded as 'other'. To measure unexpectedness of a disaster (denoted as UE_disaster), we use the inverse of the frequency of the same-subtype disaster occurring in the corresponding country. In other words, the more frequently a disaster occurs, the more expected the disaster is. We add variables about the impact of disasters: the number of people involved in the disaster, the type of people's involvement in the disaster (denoted as count type), and the type of people affected by the disaster (denoted as object type). We also consider the country where the disaster occurs.

We finally add one binary variable to show whether a disaster is reported by any of international news agencies, denoted as INAs covered. This reflects logistics of news gathering, determining the news flow by gate-keeping. We focus on three global news agencies, as many previous literature does: Agence France-Presse (AFP), Associated Press (AP) and Reuters.

After considering the above variables, elimination of multicollinearity is a crucial step because multicollinearity distorts estimated coefficients of variables. We take three steps of analysis to select relevant variables. First, we build a linear regression for each of the independent variables to see its predictive power for global news coverage of disasters. In this step, we discard the homicide rate variable as it shows low significance in predicting the global coverage. Then, we compute the (Pearson) correlation coefficient between each pair of variables. We find that a few national variables are correlated with each other (i.e., there are high positive correlations (where $r > 0.60$) among GDP, GNI,

GNI, Internet use, life expectancy, the number of scientific journal articles, political stability, and index of press freedom). We retain “political stability” as it is the most predictive factor among those eight variables. By a variance inflation factor (VIF) test, we additionally remove the trade variable. Lastly, through stepwise variable selection using AIC, we get the final regression model with 14 independent variables and an additional control variable (location). We confirm no collinearity by a VIF test; all the remaining variables have VIF below than 5.3.

For the analysis, we use the GKG dataset provided by the GDELT project that incorporates 3,574,627 events happening in 205 countries from April 2013 to July 2014. We extract 666,150 natural or man-made disasters and filter out disasters if any variable is missing. As a result, we have 195,513 disasters to build the model.

5.2 Results

We move on to examine the explanatory power of variables in determining global news coverage of a disaster with our hierarchical multiple regression model. Seven national variables are entered in the first model. Then, six disaster attributes are added to the second model for determining their unique contribution while controlling for the national characteristics. Finally, the full model is tested, including the simultaneous examination of all variables. We discuss only the contribution of individual variables in the full model due to lack of space. We use a 0.05 level of statistical significance to evaluate the results of the regression analysis.

Table 1 shows the regression result with estimated coefficient and its statistical significance. In the first model, three national variables, which are $\log(\text{population})$, mobile subscription, political stability, have a significant effect on the dependent variable and explain 3.1% of its variance. The figures for the second model, in which disaster variables are added to the first model, indicate that the characteristics of a disaster themselves explain an additional 4.3% of the variance. Together, the national and disaster variables explain 7.9% of the variance in global news coverage.

In the final step, the newly added variable, INAs covered, explains an additional 18% of the variance, resulting in a total of 25.4% of the variance in global news coverage. Its gain is much more than the amount of variance explained by the national and disaster variables. The explanatory power of our model is comparable with previous studies in this research area; in the study of finding systemic determinants of news coverage of 38 countries [22], 20 out of 38 models have lower R^2 than ours.

National Attributes. We find that out of 7 variables considered, only two are statistically significant. Population has a positive coefficient, but its effect is marginal. High population of a country commonly leads to more emigrants to other countries. For example, China and India are major nationalities of US immigrants, in spite of their remoteness to US. As Lacy et al. point out, news coverage of newspapers is influenced by audience demand [13]. Higher number of immigrants possibly explain more news coverage for them. This relationship would be clarified when a guest, a host country framework for foreign news coverage is adopted, when the number of immigrants can be counted.

Table 1. Hierarchical multiple regression predicting global news coverage (N = 195,513)

	Model 1	Model 2	Full model
Intercept	-0.87 (0.63)	0.33 (0.64)	0.19 (0.54)
National attributes			
Mobile subscriptions	0.00 (0.00)**	0.00 (0.00)**	0.00 (0.00)**
log(Population)	0.08 (0.03)*	0.07 (0.03)*	0.06 (0.03)*
Political stability	-0.17 (0.06)**	-0.17 (0.05)**	-0.13 (0.04)**
Disaster attributes			
Manmade disaster		-0.95 (0.02)***	-0.76 (0.02)***
Natural disaster		-1.06 (0.01)***	-0.83 (0.01)***
# of affected people		0.00 (0.00)***	0.00 (0.00)***
UE of disaster		-0.28 (0.07)***	-0.17 (0.07)**
Count type			
Kill		-0.33 (0.04)***	-0.22 (0.04)***
Other		-0.73 (0.04)***	-0.52 (0.04)***
Protest		-0.56 (0.04)***	-0.45 (0.04)***
Wound		-0.40 (0.04)***	-0.26 (0.04)***
Object type			
Victims		-0.47 (0.22)*	-0.44 (0.20)*
Disaster subtype			
Radiation leak		0.58 (0.37)	0.65 (0.33)*
Toxic waste		0.37 (0.17)*	0.06 (0.15)
Aftershocks		1.37 (0.37)***	1.08 (0.34)**
Flooding		0.56 (0.14)***	0.33 (0.13)**
Heat wave		0.33 (0.17)*	0.18 (0.15)
Ice		0.58 (0.14)***	0.31 (0.13)*
Landslide		0.67 (0.14)***	0.30 (0.13)*
Monsoon		0.29 (0.15)*	0.14 (0.13)
Severe weatehr		1.10 (0.17)***	0.07 (0.15)
Tsunami		0.33 (0.15)*	0.20 (0.13)
Other		0.38 (0.14)**	0.20 (0.13)
Logistics of news gathering			
INAs covered			3.74 (0.02)***
R^2	0.03090539	0.07369887	0.2536145
AIC	802417.49	793681.70	750501.26
BIC	802519.33	794241.77	751071.51
Log Likelihood	-401198.75	-396785.85	-375194.63
Num. groups	74	74	74

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

The political stability is negatively correlated with global news coverage. In other words, disasters happening in politically unstable countries receive more global attention. This finding is aligned with the study by Masmoudi that reveals the Western news media intentionally cover crisis or conflicts of unstable countries so that the stereotype of them can be reinforced [15]. Even though we find the possibility of such a tendency from the aggregated news media, we do not attempt to quantify the contribution of Western news media to this. In-depth analysis is required to assess the universality of reinforcing stereotype.

Disaster Attributes. The negative coefficients of man-made disaster and natural disaster mean that a disaster tagged as both man-made and natural disasters is more likely to get global attention than when only one theme is tagged. It implies that when a natural disaster happens, a complex situation where human factors are involved is likely to be covered by news media.

We find that the number of affected people is statistically significant with a marginal effect. This supports the previous finding that the number of killed people is an important factor for the natural disaster news coverage [10]. In our study, not only killed people but also affected people have been taken into account and we find that it has significant explanatory power.

The number of affected people greatly varies across disasters. For example, sometimes a few tens of thousands of people are evacuated when serious flooding occurs. Given that the significance of the number of killed people is still debatable [5], we believe that follow-up investigations in various settings are vital for assessing the significance of the affected people in global news coverage.

Unexpectedness of disaster subtype is negatively correlated with global news coverage. This finding counters what had been found in the theory of newsworthiness [11]. We explain a possible mechanism in the discussion section.

The count type is a categorical variable. The beta coefficients of category show the relative predictive power of each type compared to the base type, which is Kidnap in our model. Among the count types we consider, we find that the Kidnap type tends to get the most global attention compared to other type: Kill (-0.22), Wound (-0.26), Protest (-0.45), Other (-0.52). The negative coefficient of the Other type means that fewer countries report the Other type than they report the Kidnap type.

Lastly, some disaster types are more favorable globally than others. Although marginal, five disaster types are statistically significant and positively related to the number of countries covering the disaster: Aftershocks (1.08), Radiation leak (0.65), Flooding (0.33), Ice(0.31), and Landslide(0.30).

Logistic of News Gathering. INAs covered has the largest beta coefficients (3.74). Its positive sign indicates that a disaster covered by INA is more likely to get global attention. This is expected; however, the extent of the effect is not expected. INAs covered alone explains 18% of the variance in global news coverage. This result is in line with previous work reporting that the presence of INAs is a primary predictor of the amount of news coverage about the country [22]. We also agree with Wu's argument that the most news media sites are dependents of INAs because the cost of managing correspondents to investigate foreign issues is higher than using news copy of INAs. We discover that the INAs still play a prominent role in expanding news coverage of a foreign disaster.

6 Discussion

A lot of US media are tracked by GDELT, as the strongest player in the media industry. Thus, the number of news articles about a disaster is readily influenced by US news media. To neutralize this possible bias, we define global attention to a disaster as the number of countries covering the disaster. This equalizes the contribution of news media in each country to global attention. While this simplifies the international power relationship, it also captures well the news geography and news flow among countries, which is essential to our research question.

References

1. All GDELT Global Knowledge Graph (GKG) Files, <http://data.gdeltproject.org/gkg/index.html> (accessed August, 8 2014)
2. Index of press freedom, <http://rsf.org/index2014/en-index2014.php> (accessed August 8, 2014)
3. Political Stability and Absence of Violence, The Worldwide Governance Indicators, <http://www.govindicators.org> (accessed August 8, 2014)
4. World giving index, <http://www.theguardian.com/news/datablog/2010/sep/08/charitable-giving-country#data> (accessed august 8, 2014)
5. Adams, W.C.: Whose lives count? tv coverage of natural disasters. *Journal of Communication* 36(2), 113–122 (1986)
6. Ahern, T.J.: Determinants of foreign coverage in us newspapers. In: *Foreign News and The New World Information Order*, pp. 217–236 (1984)
7. Arva, B., Beiler, J., Fisher, B., Lara, G., Schrodt, P.A., Song, W., Sowell, M., Stehle, S.: Improving forecasts of international events of interest. In: *The 3rd Annual Meeting of the European Political Science Association, Barcelona, Spain (June 2013)*
8. Chang, T.-K., Lee, J.-W.: Factors affecting gatekeepers' selection of foreign news: A national survey of newspaper editors. *Journalism & Mass Communication Quarterly* 69(3), 554–561 (1992)
9. Franks, S.: The carma report: western media coverage of humanitarian disasters. *The Political Quarterly* 77(2), 281–284 (2006)
10. Gaddy, G.D., Tanjong, E.: Earthquake coverage by the western press. *Journal of Communication* 36(2), 105–112 (1986)
11. Galtung, J., Ruge, M.H.: The structure of foreign news the presentation of the congo, cuba and cyprus crises in four norwegian newspapers. *Journal of Peace Research* 2(1), 64–90 (1965)
12. Harcup, T., O'neill, D.: What is news? galtung and ruge revisited. *Journalism Studies* 2(2), 261–280 (2001)
13. Lacy, S., Chang, T.-K., Lau, T.-Y.: Impact of allocation decisions and market factors on foreign news coverage. *Newspaper Research Journal* 10(4) (1989)
14. Leetaru, K., Schrodt, P.A.: Gdelt: Global data on events, location, and tone, 1979–2012. Paper presented at the ISA Annual Convention, vol. 2, p. 4 (2013)
15. Masmoudi, M.: The new world information order. *Journal of Communication* 29(2), 172–179 (1979)
16. Östgaard, E.: Factors influencing the flow of news. *Journal of Peace Research* 2(1), 39–63 (1965)
17. Peterson, S.: Foreign news gatekeepers and criteria of newsworthiness. *Journalism Quarterly* 56(1), 116–125 (1979)
18. Sreberny, A., Stevenson, R.: Comparative analysis of international news flow: an example of global media monitoring. In: *International Media Monitoring*, pp. 55–72 (1999)
19. Westerståhl, J., Johansson, F.: Foreign news: News values and ideologies. *European Journal of Communication* 9(1), 71–89 (1994)
20. Wilke, J., Heimprecht, C., Cohen, A.: The geography of foreign news on television a comparative study of 17 countries. *International Communication Gazette* 74(4), 301–322 (2012)
21. Wu, H.D.: Investigating the determinants of international news flow a meta-analysis. *International Communication Gazette* 60(6), 493–512 (1998)
22. Wu, H.D.: Systemic determinants of international news coverage: A comparison of 38 countries. *Journal of Communication* 50(2), 110–130 (2000)