

# A Study of the Relationship between the Geographic Locations of the User and Participation in Twitter during Different Types of News Events

<sup>1</sup>Ghada Amoudi, <sup>2</sup>Carolyn Watters

<sup>1</sup>*K Faculty of Computing and Information Technology, King Abdulaziz University,*

<sup>2</sup>*Computer Science Department, Dalhousie University*

gaamoudi@kau.edu.sa; cwatters@dal.ca

## ABSTRACT

Twitter is one of the most active social networks in news sharing. People report local events sometimes faster than news agencies. During major events, such as earthquakes and presidential elections, people share tweets, retweets, images and links related to the event, creating an overwhelming number of posts. The amount of data generated by social media provides a shared resource for the discovery of information for individuals, organizations and governments, but the large stream of tweets makes tracking interesting posts a challenging task. Understanding user behavior during different contexts provides substantial insight to get the most out of the social media. The objective of this work is to investigate the relationship between user participation, news type and geographic locations of Twitter's users. The results show that financial news tweets have distinct user behavior compared to tweets about political events and tweets about disasters. Tweets about financial events tend to have more original tweets and more links, than tweets about political events, which had larger numbers of retweets. The investigation of the relationship between the style of user's participation and the type of news provides insight for social media management, specifically, for using information diffusion proactively, real time filtering, and location-aware news recommendation systems. We found that there are relationships between the country of the user, the type of news and user tweeting behavior for news related tweets. News type is strongly correlated to users' behavior and was found to have a stronger relationship to users' behavior than geographic location.

**Keywords:** Twitter, Social media, News sharing, geocoding, logistic regression.

## 1 Introduction

News sharing and reporting is one of the most popular social media activities. Social media sites, including Twitter, experience heavy traffic during major events, e.g., earthquakes and presidential elections, as people tweet, retweet and reply to each other. This makes tracking important posts from around the world a challenging task. Individuals often use multiple social media accounts in addition to Twitter, such as Facebook, Snapchat, and Pinterest. Posts are generated continuously in such applications, resulting in vast amounts of data that the user must deal with and much important information could be overlooked. While access to news is now nearly global, the influence of local differences remains important. Social

media allows people to follow like-minded friends from around the globe, which may introduce bias in the coverage, however, the desire to learn about local events is still important [1].

Twitter, with 340 million users generating 6000 tweets per second on average [2]. [3] Found that over 85% of topics shared in Twitter are headline news or comments about news. People report news sometimes before official news channels; for example, the US Airways plane crash that took place in 2009 was reported in Twitter by pictures before any local news agency appeared in the scene [4]. Additionally, the retweet feature provided by Twitter, i.e., the process of reposting messages of other users, increases traffic considerably [5]. [3] found that, on average, messages get 1000 retweets, even when the number of followers of the original message poster was low.

The limited length of messages causes people to adopt different strategies, from retweets to links to full articles, websites, pictures, and videos, to share much more information than 140 characters could alone. People frequently include links in tweets to share links to sites, articles, videos, and images [6]. [7] found that 25% – 29% of tweets posted daily contain links. URL sharing patterns may be used to identify the links shared in a certain geographical region and which may be of interest to other users in that region. Additionally, link analysis may provide potential applications in recommendation, advertising and detecting a spam [7].

Twitter posts includes several fields not shown to the user that contain information about the poster, including user location, time zone, and geographic coordinates, which can be used to identify the name of the location of the user by geotagging the tweet location and/or using the user stated location. Employing such geographic features for developing location-aware news services provides two possible advantages. First, allowing users to view news-related tweets by location, and second, reducing the amounts of tweets streaming to the user, thus providing a filtering mechanism.

The abundance of user generated contents provided by Twitter poses a challenge for users to find quality news and information on news events. [8] indicated that interactive querying and analytics could be developed that would enhance the reading experience of Twitter posts. That is, given the scale and the immediacy of Twitter in news reporting, it has become necessary to investigate different ways for viewing and tracking tweets related to major world events.

In this work we seek to understand the relationship between geographic location of users and the behavioural characteristics of their participation, related to creating original tweets or retweeting, and the inclusion of links using Twitter for news events. We examined these characteristics on three different types of news, finance, disaster and world politics, and across countries of origin for six stories from 2015.

This paper is organized as follows, background, methodology, results, analysis and discussion. Lastly, contributions, conclusion and future work are outlined.

## 2 Background

Twitter research covers a broad spectrum of topics, ranging from popular areas such as event discovery and network analysis to novel ones such as mental depression detection [9]. In the following subsections we explore related work in news sharing, social network analysis, and modeling user behavior.

## 2.1 News Sharing in Twitter

News is among the most common content shared in Twitter. Twitter is ranked second in news discussions after Facebook [10]. Three news genres stand out: financial news, global disaster news, and political news. According to [11] news sharing is among the most important reasons for using Twitter, and due to the widely-adopted retweet feature, news spreads in Twitter in a very fast rate. While the types of news posted in Twitter vary from serious news events to movie and celebrities' news, in this research we focus on finance, disaster and politics news. Researchers have long investigated the relationship between social media activity and market prices to improve investment strategies [12]. [13] used Twitter data to monitor stock price change and trade volume. They created a trust network of Twitter users, using the followers of the Financial Times Twitter account. The study found that trade volume was more correlated than the daily price change with tweets, and the trust network approach improved predicting trade volume. Twitter has long been associated with reporting and commenting on global disasters. [14] presented a Twitter parser to capture and analyze tweets to support the work of the Global Disaster Alert and Coordination System (GDACS). The study analyzed Twitter data containing the word "earthquake" from the end of October 2011 until the middle of January 2012. They found three peaks of heavy activity coinciding with three earthquakes that happened in that period. Finally, engagement through social media on politically motivated events is a global reality. [15] analyzed Spanish tweets related to the Boston terror attack gathered during one week, in April of 2013. The study found that during this event the retweet frequency was much higher, 44%, than replies, 5%. From these results, they concluded that users did not use Twitter as a communication interaction means, but rather to broadcast information and breaking news.

## 2.2 Information Propagation

Information propagation or information diffusion examines how information spreads from one user to another in a social network. Much work has been done in this area. [3] investigated the topological characteristics and information diffusion in Twitter and found that the follower-following topology did not follow a power-law distribution and had low reciprocity. [16] analyzed the propagation patterns in tweets during breaking news using 58,000,000 tweets about Michel Jackson's death. They found that 37% of the messages propagated more than three hops away from the first tweeter, 75% of replies happened within 16 minutes, and 75% of message flows were short lived, lasting less than an hour. [17] studied message propagation on Twitter by analyzing message sharing related to the Iranian election in 2009 and found that the retweet probability depended heavily on tweet's contents. [18] studied the propagation patterns of news from twelve news sources in Twitter, including the BBC and the New York Times. The study found that BBC had the maximum spread and longest lifespan of the twelve news sources.

The geographic properties of users in a social network have been found to be important in the discussion of a news events [19]. The Internet and social networks allow a user to connect to other users regardless of the physical distance between them. However, distance, language, country and the number of airplane flights between the different locations have been shown to have an influence on the formation of Twitter ties [1, 20].

## 2.3 Modeling User Behavior

Twitter has a unique network structure, unlike traditional social networks where links are reciprocal in most of the cases [3]. Twitter includes a large number of unidirectional ties. For instance, news media

outlets and celebrities accounts are followed by thousands, but seldom follow their followers. Thus, understanding the influence of users in such a network cannot rely solely on network structure, and the behavior of users must be considered and analyzed [21]. User behavior modeling has been used in areas such as discovering influential users, providing recommendation [22], building users profiles [23], and predicting new information spread [21]. Combining geography and news type with user behavior modeling is an area that has not been adequately addressed in the literature.

Retweet behavior has been shown to be the dominant message spreading mechanism for users. Retweets have been widely adopted in disaster situations. For instance, communication between the government of Japan and people during the of nuclear radiation leakage disaster that took place after the Fukushima earthquake in 2011, was conducted largely through Twitter by means of retweets [24]. Results showed that the most retweeted messages were in the intent of alarm. The study also found that retweeted messages differed dramatically in content from what newspapers and television were broadcasting, which contained more reassuring messages, while Twitter messages contained more sense of alarm. [25] also used retweets as a filtering mechanism for real-time tweet retrieval in disaster events. Results showed that the retweet-based methods were well suited for quick tweet retrieval in disaster situations.

Sharing links is among the most common behaviors of Twitter users. Link related research investigates link usage properties [26] as well as spam and security issues [7]. [6] studied the characteristics of links shared in Twitter to improve website navigation for large sport events. The study found that 25-47% of all of the tweets analyzed contained links, and the number of links shared was affected by the type of the event and the characteristics of the official website.

### 3 Methodology

This research employed a quantitative analytical approach. Data collection and preparation process were done by applying ETL (Extract-Transform-Load) to the data [27]. The data were collected from Twitter Streaming API in JSON format (JavaScript Object Notation) and parsed into fields of interest then transformed to a relational database for analysis.

371,379 tweets were collected from the streaming API resource of the DNLP, Dalhousie Natural Language Processing research group which consisted of a collection of news tweets (1% sampling) collected during 2014 and 2015, and organized by days and categories, such as finance, auto and retail.

While news channels broadcast stories of different news types daily, not all news stories motivate social media users to post and comment. During some events the tweeting and retweeting activity between social media users increase significantly [28]. Tweeting may be initiated for many reasons, both endogenous or exogenous. Endogenous trends occur when popular ideas spread widely by viral contagion or information cascade. Exogenous trends are typically associated with real-world events and are spread widely through direct user interaction. For the purpose of this research we chose events that triggered exogenous attention and chose news events significant enough to trigger social media activity and generate sufficient tweets for analysis. Three types of news were chosen for this research, finance, disaster and politics, based on previous research in these areas. We extracted tweets related to six events, two per news genre with more than 10 thousand tweets each, from the 2015 DNPL data set. The two stories chosen for political news were the shooting in the Charlie Hebdo magazine office in Paris, and the story of the teenage boy who was arrested in a Texas high school for inventing a clock that his teacher

mistakenly thought was a bomb. For financial news, the stories chosen were the Chinese stock market fall causing disturbance in shares market around the world, and the story of the automaker Volkswagen, which was accused for manipulating pollution control systems. The stories selected for disaster were the Germanwings airplane crash, an airbus crashed in French Alps killing 150 people, and the Nepal earthquake of 7.8 magnitude that killed over 9,000 people.

For each news event, tweets were collected for a 7-day period, starting at the day the event took place. We chose 1-week period, since tweeting activity for most trending topics lasts for a week or less [3]. Keyword searches were applied to find those tweets related to each of the six-news story. The total count of tweets for processing was 371,379 tweets.

To identify the geographic location of the user, the user location and time zone fields of the tweet, found in over 70% of the tweets, were extracted. Using these two fields, MapQuest geocoding API was applied to identify the country name. If the user location was not found after geocoding, the time zone field was used to approximate a geographic region. Any tweet without a user location or a time zone was eliminated from the dataset. Manual screening was applied to 1000 randomly selected geocoded user locations to test validity and for 72% of the sample the users' locations had been correctly geocoded. The number of tweets remained after this step was 280,436 tweets.

### **3.1 Analysis**

The aim of the analysis is to examine the relationships between user participation, defined by occurrences of original tweets and retweets and links in the 280,436 tweets, as well as geographic locations, and news type. In the first phase, a general descriptive analysis examined the characteristics of the whole dataset, i.e., the 6 stories combined. In the second phase, a statistical logistic regression analysis was applied to model the relationships between country, news type and user behavior.

#### **3.1.1 General Descriptive Statistics for All 6 Stories**

This analysis provides an overall description of all the tweets used in the study.

#### **3.1.2 Characteristics by country**

Using the whole dataset (all 6 stories), the top 10 countries were identified, and analyzed using the numbers of originals and retweets, tweets with links, the original/retweet distribution for tweets with links were obtained for each of the top 10 countries. Only three countries appeared in each individual story, USA, UK, and India, the tweets for these countries were then analyzed in the same manner.

#### **3.1.3 Characteristics by news type**

News stories of each news type were then collapsed into individual datasets, resulting in three datasets by news type: finance, disaster and politics. The three datasets were analyzed by the same characteristics used for the whole database including an analysis of the Top 10 countries and the common 3 countries for each of the 3 news types.

#### **3.1.4 Characteristics by news story**

Each of the pairs of news stories were examined on the behavioral and geographic characteristics overall as well as for the top 10 countries and the common 3 countries.

### 3.1.5 Statistical analysis

Binary logistic regression was then applied to the data to provide two basic outcomes. First, to find the significance of the co-relations of the variables country and news type. Second, to model the relationship between the variables country, news type and behavior. From this model, odds ratios were obtained to facilitate predictive comparisons among countries and among news types in relation to user behavior.

## 3.2 Assumptions

User locations considered for this research were obtained from the user profile data. For this study, we assumed that for a certain user, the location obtained was the location where the user lived, but not necessarily where the tweet came from. We also assumed that the user stayed in the original location during the period of data collection, so if the user appeared twice in the dataset only the first location was fetched and saved in the database.

## 4 Results

The results of applying the analyses described earlier are presented in this section. First, the results describing the whole dataset are presented, the results by country and news type are reported next. Finally, we explain the results of the statistical tests applied for analyzing the relationships between original/retweet, links, and both country and news type.

### 4.1 General Descriptive Statistics for All 6 Stories

The general characteristics for the whole dataset provide an overall description of all the tweets used in the study. These results are presented in Table 1.

**Table 1 General characteristics of tweets in the database of the 6 stories combined**

Characteristics	Count	%Total	
		Tweets	/Retweet
Total tweets	280,436		
Unique users	203,363		
Average tweets per	1.38		
Unique countries	235		
Originals	114,695	41%	
Avg. originals per user	0.57		
Retweets	165,741	59%	
Avg. retweets per user	0.81		
Tweets with Links	214,697	77%	
Originals & Links	89,166	32%	78%
Retweets & Links	125,531	45%	76%

The original to retweet ratio was found to be 41% to 59%, indicating that overall most of the tweets were retweets. 77% of tweets contained links, and 76% of retweets contained links. This link-use ratio in this data set for news was considerably higher than that reported by [7] which were between 25% – 29%. This could be due to the different nature of the tweets used in our research as opposed to a dataset of tweets with no specific event.

## 4.2 Characteristics by Country

73% of tweets originated from 10 of the 235 countries, and only three countries were in the top 10 producers of tweets in each of the 6 stories. In all cases the number of tweets from individual country followed a power law, Zipf-like distribution.

Two subsets of the main database were created for further analysis: the first with 10 countries and the second with 3 countries. The two subsets, as shown in **Error! Reference source not found.**, exhibited similar patterns to the whole database in terms of percentages of originals, retweets and tweets with links to total tweets in each one of them.

**Table 1 General characteristics of tweets in the database of the 6 stories combined.**

Dataset	Total Tweets	Originals	Retweets	Tweets with links
<b>All countries</b>	280,436	114,695	165,741	214,697
<b>Top 10</b>	205,296	83,977	121,319	156,272
<b>Common 3</b>	132,240	59,409	72,831	103,729
<b>% All countries</b>		41%	59%	77%
<b>% Top 10</b>		41%	59%	76%
<b>% Common 3</b>		45%	55%	78%

### 4.2.1 Top 10 countries

The top 10 countries produced 73% of the dataset, these countries include USA, France, UK, India, Netherland, Canada, Spain, Greece, Indonesia and Germany. These countries along with their ranks are presented in Table 3.

**Table 2 Top 10 countries in the whole dataset**

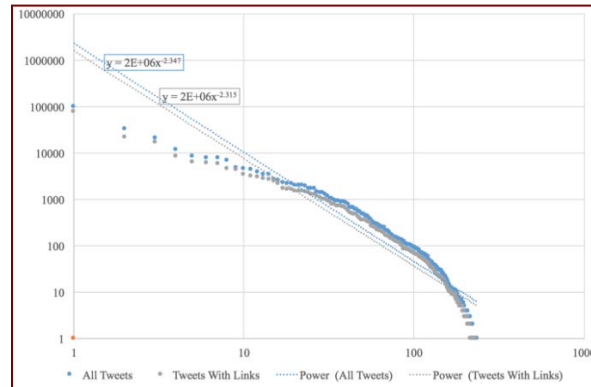
Country	Tweets	% Originals	% Retweets	% Tweets with Links
<b>USA</b>	99,295	44%	56%	79%
<b>France</b>	32,463	25%	75%	67%
<b>UK</b>	21,316	46%	54%	79%
<b>India</b>	11,629	49%	51%	73%
<b>Netherlands</b>	8,604	33%	67%	81%
<b>Canada</b>	7,854	45%	55%	70%
<b>Spain</b>	7,813	31%	69%	74%
<b>Greece</b>	6,872	29%	71%	67%
<b>Indonesia</b>	4,841	76%	24%	90%
<b>Germany</b>	4,609	44%	56%	76%
<b>Total</b>	205,296	41%	59%	76%

The common three countries USA, UK and India were the first, second and fourth ranks respectively. France occupied the second place due to the large number of tweets generated by Charlie Hebdo story, which represented a large portion of the dataset. The number of retweets was more than originals in all countries except for Indonesia. All the top 10 countries had high percentages of tweets with links, the lowest in Greece (67%) and the highest in Indonesia (90%).



#### 4.2.2 Common 3 countries

The three countries found in all the six stories were used for more intensive analysis in this research study, USA, UK and India representing 47% of all the tweets in the database. The log log scale of the distributions of raw counts of all tweets from all countries in the database and tweets with links followed a power law distribution as presented in **Figure 1**.



**Figure 1** Log log scale of the distribution of raw counts of tweets and tweets with links in all countries

### 4.3 Characteristics by News Type

To examine characteristics of tweets by news type, and to develop a sense of what differentiated the news types, news stories of the same type were collapsed into one dataset, resulting in three datasets by news type: finance, disaster and politics. In the following subsection, the general descriptive statistics are presented for the three datasets.

#### 4.3.1 General characteristics by news type

The same descriptive analysis was done for each of the three combined datasets as was done for whole dataset presented above. **Error! Reference source not found.** shows the general characteristics of the combined datasets for each news type, the total count of tweets in all types was 280,436.

Table 4 shows that the highest proportion of original tweets was in finance stories (64%) and 92% of these original tweets contained links. The lowest proportion of original tweets was in political stories (27%) and (39%) in disaster stories. The retweet ratio was highest in political stories (73%).

**Table 3** General characteristics of the combined datasets of each news type

	Finance		Disaster		Politics	
	Tweets Count	% Total	Tweets Count	% Total	Tweets Count	% Total
Total tweets	93,148		38,147		149,141	
Originals	59,179	64%	14,861	39%	40,655	27%
Retweets	33,969	36%	23,286	61%	108,486	73%
Tweets with Links	84,439	91%	26,693	70%	103,565	69%
Originals & Links	54,234	58%	11,248	29%	23,684	16%
Retweets & Links	30,205	32%	15,445	40%	79,881	54%



The patterns of behavior in the different news type were different from the patterns of the whole dataset. The top 10 countries for each news type were identified and analyzed in terms of the use of original, retweets and links. These results are presented in the next subsections.

#### 4.3.2 Finance stories

The top 10 countries for finance stories are presented in **Error! Reference source not found.** The second column displays the number of tweets in each country, and the next three columns display the percentages of originals, retweets and tweets with links to total tweets in each country.

**Table 4 Top 10 countries in finance stories**

Country	Tweets	% Originals	% Retweets	% Tweets with Links
USA	42,226	64%	36%	91%
UK	10,061	65%	35%	89%
India	4,976	69%	31%	86%
Canada	4,081	60%	40%	89%
Nigeria	1,897	70%	30%	98%
Indonesia	1,845	93%	7%	97%
Australia	1,748	57%	43%	90%
Netherlands	1,598	64%	36%	90%
Germany	1,465	61%	39%	92%
France	1,318	53%	47%	87%
Total	71,215	65%	35%	90%

Most countries had similar percentages of originals and retweets except for Indonesia, which had higher percentage of originals (93%). 90% of the tweets of the top 10 countries contained links.

#### 4.3.3 Disaster stories

The top 10 countries in disaster stories included USA, India, UK, Indonesia and Mexico among other countries. The top 10 countries followed the pattern of original/retweet of the overall disaster dataset (39/61) except for Indonesia (73/27) and Mexico (61/39). More than 60% of disaster tweets contained links in all the top 10 countries

#### 4.3.4 Politics stories

USA, France, UK, Netherlands and Spain were among the top 10 countries in politics stories. Most of politics tweets were retweets, 73% retweets versus 27% originals. The top 10 countries followed the same pattern as the whole politics dataset. Links were in 69%-84% of politics tweets, similar to the overall percentage of links in the politics dataset.

In summary, the characteristics of the tweets viewed by news type revealed differences in the ratio of originals vs retweets and in the use of links across the different news types. Finance had the highest originals ratio (64%), while politics had the highest retweets ratio (73%). Finance tweets had the most links (91%).

The top 10 countries followed the overall pattern of original/retweet ratios in tweets links in each dataset. This is likely because these countries represented such a high proportion of each dataset. Indonesia, however, showed different behavior in many scenarios, but since the total counts of Indonesia's tweets

is relatively small, the different pattern Indonesia exhibited did not affect the overall results. This finding agrees with the finding of [29]. They studied the top 10 countries they found active in Twitter and reported that Indonesia ranked first in tweets per user, and had the fewest retweets, compared to the other countries.

#### 4.4 Statistical Analysis

To understand the relationships between tweeting activity and news type and country, a binary logistic regression analysis using SPSS was applied on the test dataset that combined the tweets for all six stories from the common three countries USA, UK and India. The reasons for using this subset rather than the entire dataset was the difficulty of applying regression on all 235 countries and because of the consistency found in the common 3 countries with the whole database with all the 6 stories in terms of the percentages of originals, retweets and tweets with links as presented in Table 2.

The predictive variables used in this study were the categorical variables: country, news type and the interaction between country and news type. Each of the predictive variables (country and news type) have three levels: country includes USA, UK and India, and news type includes finance, disaster, and politics. The outcome variables were original tweets and use of links. The interaction included all combinations between country and news type. Thus, logistic regression was applied two times, i.e., creating two prediction models, one for each of the outcome variables (originals and links) using the two predictive variables (country and news type) and the interaction terms.

**Error! Reference source not found.** presents the result of the regression analysis. Model prediction is 67.7%, 72.4% and 78.4% in the logistic regression interaction models of analyzing the relationships between county news type and the behaviors of generating originals and links respectively. The standard error (SE) in all of the results was relatively small, indicating that the coefficients estimates were precise. The Nagelkerke R square value is 16% in originals model and 10% in links model.

**Table 5 Regression analysis result summary**

	Originals	Links
<b>Model Prediction</b>	67.7%	78.4%
<b>Chi Square</b>	16,620	129,063
<b>Significance</b>	.000	.000
<b>Nagelkerke R Square</b>	16%	10%

The odd ratios for comparing users' behaviors among countries, shown in **Error! Reference source not found.**, indicate that USA and UK had similar odds ratios, however India was 1.24 and 1.19 times more likely to generate originals in politics than USA and UK respectively.

**Table 6 Odd ratios for comparing countries**

	Countries	Finance	Disaster	Politics
<b>Originals</b>	UK vs USA	1.042	0.935	0.935
	India vs USA	1.240	0.907	1.240
	India vs UK	1.190	0.969	1.326
<b>Links</b>	UK vs USA	0.827	0.881	1.087
	India vs USA	0.654	0.641	0.785
	India vs UK	0.791	0.728	0.723

The odd ratios for comparing users' behaviors within news types are shown in **Error! Reference source not found.** The logistic models for interaction between <Country, NewsType> and originals and links, were statistically significant.

**Table 7 Odd ratios for comparing news types**

	News Types	USA	UK	India
<b>Originals</b>	Finance vs Disaster	3.102	3.456	4.242
	Finance vs Politics	4.632	5.160	4.632
	Disaster vs Politics	1.493	1.092	1.092
<b>Links</b>	Finance vs Disaster	3.963	3.721	4.047
	Finance vs Politics	4.154	3.161	3.463
	Disaster vs Politics	1.048	0.856	0.856

However, the variation among news types in general was greater than the variation among countries. For instance, for the financial news examples, all countries were 3 to 5 times more likely to generate originals and links than the politics or disaster stories. These results showed that news type was more influential than country on user behavior. These results are consistent with the descriptive analysis results presented earlier. Finance events were related to user behaviors that were different from the other two news types in terms of generating originals, retweets and tweets with links; finance had larger original/retweet ratio, more tweets with links.

## 5 Discussion

In this section, we explain how these results might provide the basis for developing methods that could improve the user experience in similar contexts. The descriptive analysis showed that overall user behavior for all countries, top 10 and common 3 was different by news type and by country. The logistic regression analysis of the common 3 countries verified those relationships.

The descriptive statistics by news type showed that the profile of user behavior was related to the type of news. Finance had the highest originals ratio at 64% and 91% of tweets had links; politics had the highest retweet ratio at 73%; and Disaster had only 40% original tweets and links in 70%.

The logistic regression models showed that news type was more significant in the relationship between <Country, NewsType> than was country. Beside answering the research questions, the analyses conducted revealed some additional findings. The distributions of tweets by country followed a power-law distribution, similar to a Zipf curve, and that distribution held for retweets, original and tweets with links. The Top 10 countries generated nearly three quarter of the total tweets in the whole database, and the common 3 generated nearly 50% of all tweets. The top 10 and the common 3 countries had very similar patterns in the use of originals, retweets and tweets with links to total tweets in each of them. The common 3 countries, however, created a higher ratio of new content over retweet. This is an area of further investigation. When we examined the profile of use by news type it was clear that finance news tweets had distinct user behavior in that there were more original tweets and more links in those tweets.

Retweets were used more frequently in disaster and breaking political news, e.g. Nepal earthquake, Charlie Hebdo and clock boy stories. This is consistent with the social reaction to these types of events since retweeting is an effective response for spreading links quickly for widespread information spread. This is consistent with the finding of Wang & Zheng (2014) who noted that the single-spike pattern as result of an event had a larger retweet ratio.

The utilization of links was also interesting. In each case the distribution also followed the power-law distribution, i.e., a few links were shared many times while a large number of links were shared few times. This result is consistent with Nizam, et al. (2014) who showed that social media users share more links during big events. BBC appeared in the top 3 most shared links in 5 of the 6 stories, not surprisingly as the stories used in the study had global interest and the tweets were in English. The results found by Bhattacharya and Ram (2012) found that BBC had the larger article spread and longer lifespan than the other news sources they investigated.

## **6 Conclusion and Future Work**

In this research we found that user behavior is related to country, news type and the interaction between country and news type. Users from different geographic locations reacted differently to the different news events. The common countries contributed more in creating original contents than the other 223 countries in the dataset. The investigation also revealed that the behavior of a country may be explained partially by factors such as, the proximity to the origin of the event, e.g., India's contribution to the Nepal earthquake story, and perhaps whether there are economic or diplomatic relations between the country and the country of the event origin, e.g., Nigeria and China, and certainly whether the country is involved in the event, e.g., Spain's engagement with the Germanwings plane crash. Much of the user behavior could not, however, be explained easily by reference to the country of user and further inquiry of countries' social, economic, political situations might be needed to obtain knowledgeable explanations of their social media behaviors.

The descriptive analysis for the three news types datasets showed that users' behavior varied considerably by the type of news. Finance had the highest proportion of originals (64%), politics had the highest proportion of retweets (73%) and finance tweets had the most links (91%). Odds ratios for comparing countries showed variations ranging from 4.5% to 33%, whereas odd ratios for comparing news types ranged from -89% to 5 times. Variations related to news type were found to be more significant than those related to the country of the user.

Twitter usage is an active research area, nevertheless research involving user behaviors by region and purpose are limited. This research study provided some insight on user tweeting behavior related to a sample of global political, financial and disaster events.

The popularity of smart phones and location aware devices, in addition to the availability of geocoding and mapping services, encourages us to continue this line of research to provide real time prediction of news type. The use of hashtags, not reported in this paper, was a widely-used feature in the dataset. Using social network analysis would allow further exploration of hashtags related to the type of news and geographic location of the user and potentially to the geographic location of the event.

## **7 Research Contributions and Limitations**

The main research contributions of this research study include the following points: there are relationships between the country of the user, the type of news and user tweeting behavior for news related tweets; News type is strongly correlated to users' behavior and was found to have a stronger relationship to users' behavior than geographic location.

This research had limitations in scale and scope. Importantly, the data set of tweets used were sampled from the tweet stream; only 3 news genres were used; only 2 sample stories per genre were chosen; and all events were restricted to a six-month period in 2015; and only English tweets were used. Further investigation is needed before more generalized conclusions can be made.

Social media applications are advancing at a rapid rate; however, the human factor remains instrumental in shaping the way information is shared. In this study, the behavior of social media users on Twitter was analyzed considering two dimensions, geographic location and news type. The results indicated that user behavior is related to the type of news and to the location of the user, although geographic location less so. The findings of this type of study informs researches and developers in areas concerning modeling user behavior, detecting anomalies, filtering, ranking and recommendation systems for future social media systems.

## REFERENCES

- [1] Takhteyev, Y., Gruzd, A., Wellman, B., *Geography of Twitter networks. Social Networks. Social Networks*, 2012. 34(1), 73–81. doi: 10.1016/j.socnet.2011.05.006
- [2] Oliveira, J., Gloor, P., *The Citizen IS the Journalist: Automatically Extracting News from the Swarm*. Springer International Publishing Switzerland, 2016. p. 141-150
- [3] Kwak, H., Lee, C., Park, H., Mon, S., *What is Twitter, a social network or a news media?* In Proc. WWW '10, ACM 591–600, 2010.
- [4] Choi, D., Kim, P., *Sentiment analysis for tracking breaking events: A case study on Twitter*. In A. Selamat et al. (Eds.) *Intelligent Information and Database Systems*, 2013. 7803, p. 285-294
- [5] Wu, B., Shen, H., *Corrigendum to Analyzing and predicting news popularity on Twitter*. *International Journal of Information Management*, 2015. 35 (6): p. 702–711
- [6] Nizam, N. Watters, C., Gruzd, A., *Link sharing on Twitter during popular events: Implications for social navigation on websites*. In the 47th Hawaii International Conference on System Science, 2013. HI.
- [7] Cao, C., Caverlee, J., *Behavioral Detection of Spam URL Sharing: Posting Patterns versus Click Patterns*. 2014.
- [8] Yang, X., Ghoting, A., Ruan, Y., Parthasarathy, S., *A Framework for summarizing and analyzing Twitter feeds*. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2012*. doi:10.1145/2339530.2339591
- [9] Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., Bao, Z., *A depression detection model based on sentiment analysis in micro-blog social network*. *Trends and Applications in Knowledge Discovery and Data Mining*, 2013. 201–213. doi:10.1007/978-3-642-40319-4\_18
- [10] Gabelkov, M., Ramachandran, A., Chaintreau, A., Legout, A., *Social Clicks: What and Who Gets Read on Twitter?* *Proceedings of SIGMETRICS 2016*

- [11] Subašić, I., Berendt, B., *Peddling or creating? Investigating the role of Twitter in news reporting*. Advances in Information Retrieval, 2011. 207–213. doi:10.1007/978-3-64220161-5\_21
- [12] Yang, S. Y., Mo, S. Y. K., *Social media and news sentiment analysis for advanced investment strategies*. Sentiment Analysis and Ontology Engineering, 2016. doi:10.1007/978-3-319-30319-2\_11
- [13] Ruan, Y., Alfantoukh, L., Duresi, A., *Exploring stock market using Twitter trust network*. 2015 IEEE 29th International Conference on Advanced Information Networking and Applications, 2015. doi:10.1109/aina.2015.217
- [14] Stollberg, B., de Groeve, T., *The use of social media within the Global Disaster Alert and Coordination System (GDACS)*. Proceedings of the 21st International Conference Companion on World Wide Web - WWW, 2012. doi:10.1145/2187980.2188185
- [15] Cuesta, Á., Barrero, D. F., R-Moreno, M. D., *A descriptive analysis of Twitter activity in Spanish around Boston terror attacks*. Computational Collective Intelligence. Technologies and Applications, 2013. doi:10.1007/978-3-642-40495-5\_63
- [16] Ye, S., Wu, F., *Measuring message propagation and social influence on Twitter.com*. Lecture Notes in Computer Science, 2010. 216–231. doi:10.1007/978-3-642-16567-2\_16
- [17] Zhou, Z., Bandari, R., Kong, J., Qian, H., Roychowdhury, V., *Information resonance on Twitter: Watching Iran*. Proceedings of the First Workshop on Social Media Analytics – SOMA, 2010. doi:10.1145/1964858.1964875
- [18] Bhattacharya, D., Ram, S., *Sharing news articles using 140 characters: A diffusion analysis on Twitter*. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012. doi:10.1109/asonam.2012.170
- [19] Agrawal, D., Budak, C., El Abbadi, A. E. A., Georgiou, T., Yan, X., *Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks*. In Databases in Networked Information Systems, 2014. p. 1–16. Springer International Publishing.
- [20] Lima, A., Musolesi, M., *Spatial dissemination metrics for location-based social networks*. Proceedings from the 2012 ACM Conference on Ubiquitous Computing, 2012.
- [21] Bogdanov, P., Busch, M., Moehlis, J., Singh, A. K., Szymanski, B. K., *Modeling individual topic-specific behavior and influence backbone networks in social media*. Social Networks Analysis and Mining, 2014. doi:10.1007/s13278-014-0204-6
- [22] Yeung, C. A. Y., Iwata, T., *Modelling User Behavior and Interactions: Augmented Cognition on the Social Web*. In Foundation of Augmented Cognition, 2011. p. 277–287
- [23] Macskassy, S., Michelson, M., *Why Do People Retweet? Anti-Homophily Wins the Day!*. Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011. p. 209-216
- [24] Li, J., Vishwanath, A., Rao, H. R., *Retweeting the Fukushima nuclear radiation disaster*. Communications of the ACM, 2014. 57(1), 78–85. doi:10.1145/2500881

- [25] Toriumi, F., Baba, S., *Real-time Tweet Classification in Disaster Situation*. In WWW, 2016. doi:10.1145/2872518.2889365
- [26] Antoniadis, D., Polakis, I., Kontaxis, G., Athanasopoulos, E., Ioannidis, S., Markatos, E. P., Karagiannis, T., *we.b: the web of short urls*. Proceedings of the 20th international conference on World wide web – WWW, 2011. doi:10.1145/1963405.1963505
- [27] Ji, X., Chun, S. A., Wei, Z., Geller, J., *Twitter sentiment classification for measuring public health concerns*. Social Network Analysis and Mining, 2015. 5(1). doi:10.1007/s13278-015-0253-5
- [28] Imran, M., Castillo, C., Diaz, F., Vieweg, S. *Processing social media messages in mass emergency*. ACM Computing Surveys, 2015. 47(4), 1–38. doi:10.1145/2771588
- [29] Poblete, B., Garcia, R., Mendoza, M., Jaimes, A., *Do All Birds Tweet the Same? Characterizing Twitter Around the World*. CIKM 2011, October 24–28, 2011. Glasgow, Scotland, UK.