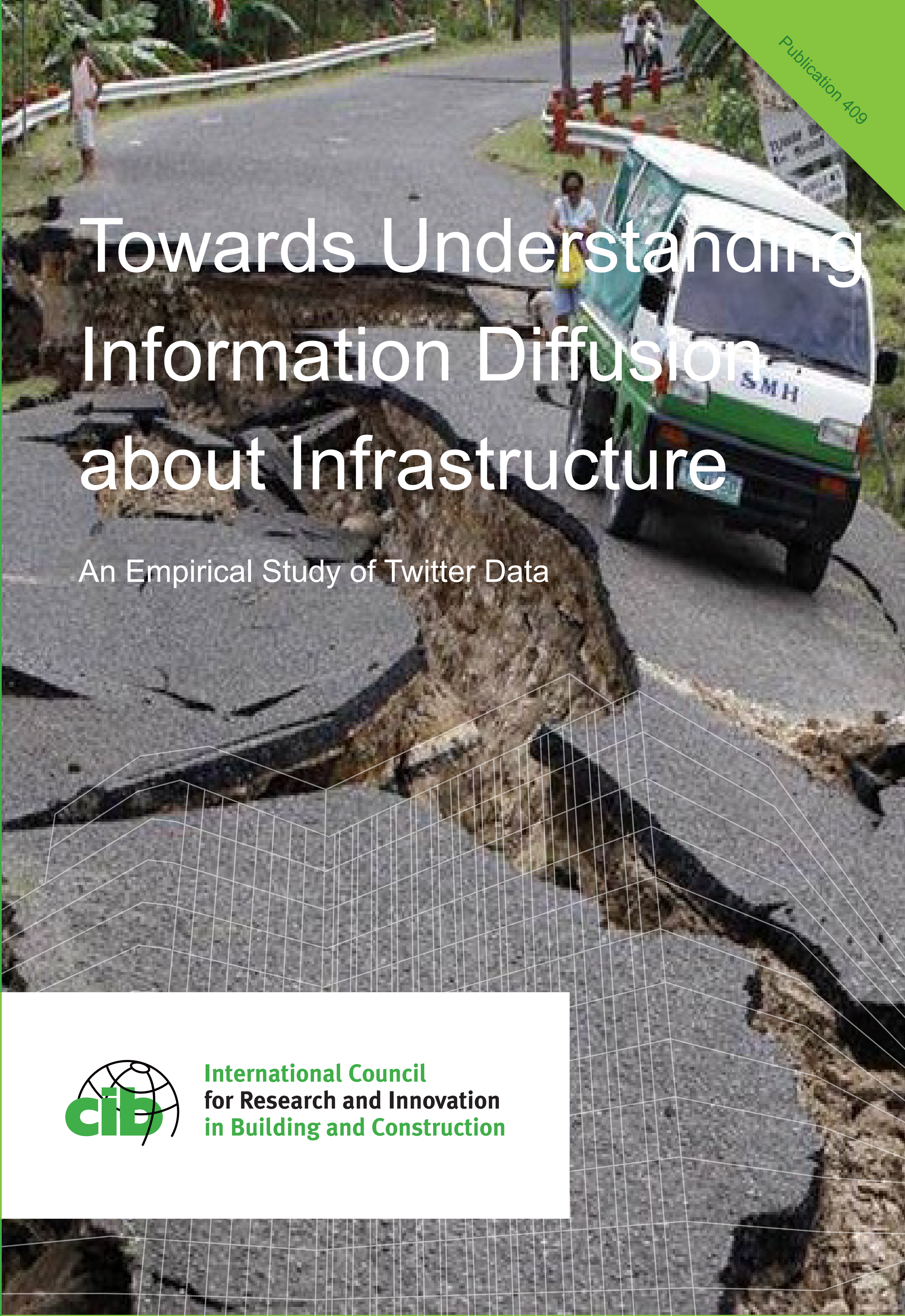
Publication 409



Towards Understanding Information Diffusion about Infrastructure

An Empirical Study of Twitter Data

International Council
for Research and Innovation
in Building and Construction

Title	Toward Understanding Information Diffusion about Infrastructure
Subtitle	An Empirical Study of Twitter Data
Serial title	409
Year	2016
Authors	Qi Wang
Contributors	Khalil Benali, Edwin Gonzalez-Montalvo, Jose Guevara-Maldonado, Yasaman Shahtaheri
Language	English
Pages	34

ISBN 978 9063 63 09 42

Publisher CIB General Secretariat
PO Box 516, 2600 AM Delft
The Netherlands
E-mail secretariat@cibworld.nl
www.cibworld.nl

Foreword

The purpose of the Sebestyén Future Leaders Award is to expose future leaders of the built environment to emerging developments in building and construction research. CIB Student Chapters compete for this award and are encouraged to collaborate with CIB Working Commissions to develop and implement their proposals.

The CIB Student Chapter of Virginia Tech in the United States was the award winner in 2014. The Chapter completed this project in collaboration with Working Commission 120 – Disasters and the Built Environment. The project examines the diffusion of infrastructure information through social media and contemplates applications of such platforms during natural hazard events to enhance infrastructure management and resilience. This report presents the results of the Chapter's efforts.

The Virginia Tech Chapter gratefully acknowledges the support of Working Commission 120, specifically Dr. Lee Bosher of Loughborough University and Dr. Jason Von Meding of the University of Newcastle, throughout this project. In addition, Dr. John Taylor of Virginia Tech was instrumental in the project's success.

Dr. Michael Garvin
Co-Advisor
Virginia Tech CIB Student Chapter

June 2016

EXECUTIVE SUMMARY

Cities around the world are facing the challenge of improving the resilience of their infrastructure systems. Resilience is the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. If the United States is taken as an example, the American Society of Civil Engineers grades the nation's infrastructure system as D+. Even more pressing, human societies are facing more and more natural hazards. Infrastructure failures during natural hazards can lead to economic loss, inconvenience to urban dwellers, and even injuries and deaths.

To develop a resilient infrastructure system, all stakeholders need to collaborate. Stakeholders need accurate and prompt information. To understand information diffusion, this project focuses on information sharing on Twitter. We created a data collection system to stream 'tweets' with geographical coordinates embedded. Using the system we retrieved all the geotagged tweets in the U.S. related to infrastructure in the year of 2014. The tweets were cleaned and processed using natural language processing. We created a word cloud related to infrastructure based on the key words from the tweets. The word cloud provides an indication of the prominence of infrastructure systems and related issues among Twitter users.

We also used hashtags as links and studied the communication network regarding U.S. infrastructure. This analysis found that the degrees of the individuals in the network still roughly follow the scaling law. This demonstrates that infrastructure has prompted discussions among individuals and organizations, which can be used to inform policy makers. We also studied the basic components of the communication network. While the network has a complex structure, it

can be broken down into different components. Other than one giant component, the rest have relatively simple structures.

This study shed light on the new phenomenon of information diffusion in online social networks and discovered some important findings. The new insights enable us to understand the basic characteristics of information diffusion in online social networks regarding infrastructure and paves the way for future research regarding how such platforms might be used to enhance infrastructure management, performance and resilience.

Contents

List of Figures

List of Tables

1. INTRODUCTION

1.1 Urban Population, Infrastructure Systems, and Natural Hazards

Human society has never been so urbanized. The urban population, increased from 1.01 billion in 1960 to 3.69 billion in 2012, has surpassed the rural population in 2009 for the first time in human history (United Nations 2009; World Bank 2013). The growing population has made urban areas more active and complex which requires more infrastructures to stomach the urban dwellers and their activities. Urban residents and infrastructure systems have formed an interdependent sociotechnical system.

The interdependence calls for more resilient infrastructure. According to the National Infrastructure Advisory Council (NIAC), infrastructure resilience can be defined as “the ability to reduce the magnitude and/or duration of disruptive events” (Moteff 2012). Also, more recently, the Presidential Policy Directive (PPD) defined resilience as “…the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions” (Obama 2013). Consequently, resilience looks to resist or mitigate the magnitude of the impact after a disruptive event, thus providing enhanced safety to the people and saving the national economy in the long run.

However, infrastructure resilience is a global challenge. Take the U.S. as an example. Based on the report from ASCE in 2013[1], the nation’s infrastructure system only received a score of D+. An estimate of 3.6 trillion dollars is needed by 2020 to fix the deteriorating system. Our lives depend on infrastructure system so much that we hardly pay any attention to them. However, any infrastructure failures can have broad impact such as power outage, traffic jams. Some of the

[1] The report can be found on the ASCE website: http://www.infrastructurereportcard.org/.

failures can result catastrophic consequences such as dam and bridge collapse, water population etc.

The challenge is made more pressing due to the increasing impact from natural hazards. From 2000 to 2012, natural hazards caused 1.7 trillion dollar of economic loss, killed 1.2 million people and affected half of the entire population on earth (UNISDR 2013). Looking into the future, both the magnitudes and frequencies of natural hazards are expected to increase. Such increase is owing to the global environmental change. Human activities and fossil fuel consumption is causing global warming. Analysis shows that the global annual mean surface air temperature has increased during the last 65 years (Stocker et al. 2013). Almost all simulation models project that such trend will continue for the next 20 years. The trend of global warming has caused steady retreat of glacier (Oerlemans 1994) and impacted the sea levels (Meehl et al. 2005). Global warming also caused anomalous tropical sea surface temperatures and global aridity. Dry periods lasting for decades have occurred multiple times during the last millennium across the world (Dai 2013). While aridity will increasingly influence most of regions in the world, places like the United States which have avoided prolonged droughts during the last 50 years due to natural climate variations may see persistent droughts in this century. Global warming is influencing tropical cyclones as well. Records show that the frequency of Atlantic tropical cyclones has increased for the last 100 years, and the increasing trend has accelerated since the 1980s (Landsea et al. 2010). By 2100, the intensity of tropical cyclones is predicted to increase by 2-11% (Knutson et al. 2010). It is predicted to be a long-term, continuous battle between humans and natural hazards.

Recent natural hazard events have shown how they can impact urban infrastructure. First, natural hazards can severely impact transportation networks in cities. As Pan et al. (2007) pointed out, overcrowding and crushing during emergency situations can cause incidents and thus injuries and the unnecessary loss of lives. The snow storm that happened in Atlanta, Georgia is an excellent example. Alarmed by the approach of a severe snow storm in December 2013, the City of Atlanta, Georgia, issued a snow storm warning and advised people to leave school and work early and return home. The unfortunate consequence of this warning was that residents all crammed onto the city's roads and highways at the same time, causing a city-wide traffic jam. Many were still stuck on the roads when the storm hit, forcing them to abandon their vehicles and seek shelter (Beasley 2014). Second, natural hazards can force people for relocation. During Hurricane Sandy, the New York City government issued a mandatory evacuation order. Similar situations happened during the attack of Typhoon Haiyan. While people were ordered to seek shelter in the City of Tacloban, many people moved to concrete buildings instead of seeking higher ground. Many of these buildings were flooded and/or collapsed and claimed many people's lives (Teves and Bodeen 2013).

Besides the threat to people, there is also the threat to our economy. Following each hazardous event, the economy of a state or nation is impacted. Indeed, natural disasters can lead to degraded infrastructure which is strongly related to the economy of countries. Furthermore, impacts on infrastructure can negatively affect aspects like business productivity, international competitiveness, employment, and the gross domestic product (GDP), among others. Thus, ensuring that the infrastructure systems are sustainable and resilient is increasingly becoming one of the top priorities of policy makers. It is becoming a challenge to predict and mitigate the consequences of these natural hazards (i.e. storms, hurricanes, etc.). Such challenges call for innovative research and technologies to improve disaster evacuation, timely responses and relief plans – thereby enhancing infrastructure resilience.

1.2 Information Diffusion during the Occurrences of Natural Hazards

Information delivery is critical during the occurrence of natural hazards. Take Hurricane Sandy as an example. Before Hurricane Sandy made land fall, the government of New York City issued a mandatory evacuation order. However, a survey showed that only 71 percent of the residents living in the evacuation zones were aware of the mandatory order. Not everyone took the order seriously enough; roughly 50 percent of those aware of the order decided to not move. The actual evacuation rate was only about 35%. Unfortunately, most of the fatalities occurred in the mandatory evacuation zones (CDC 2013). More alarmingly, even those who evacuated were not entirely safe. In a comparison of the flooding areas reported by FEMA with the mandatory evacuation areas, the flooding area was found to be 15 percent larger. Consequently, people could have possibly moved out of the evacuation areas and think they were safe; but in reality they possibly faced a severe threat (Wang and Taylor 2014). This case from New York City demonstrates how important it is to promptly deliver accurate information to affected individuals.

Word-of-Mouth (WOM) is one of the most effective ways to deliver key information related to infrastructure. The WOM communication channel is especially effective in large-scale social networks. People are members of social groups rather than isolated individuals, and they are more influenced by local communities and trusted social connections (Stern et al. 1986). WOM is a strong tool that can potentially disperse information to people through local networks. Therefore, information sharing should rely on WOM communication through existing social networks and community organizations (Stern 1992). Researchers have studied the properties of WOM communication networks, including network position and centrality (Arndt 1967; Vilpponen et al. 2006), tie strength (Brown and Reingen 1987; Goldenberg et al. 2001; Reingen and Kernan 1986; Vilpponen et al. 2006; Wang and Taylor 2013), communities and cliques

(Vilpponen et al. 2006; Webster and Morrison 2004), density (Vilpponen et al. 2006), social dimensions (Allsop et al. 2007), etc. The advantage of social network analysis is its ability to integrate micro-level analysis to macro-level system patterns (Brown and Reingen 1987) and quantitatively examine the interrelation between social network structures and communicators' positions and influences in WOM communications. Although WOM is deemed an important way to share information, limited research has quantitatively studied information diffusion (File et al. 1994; Hennig-Thurau et al. 2004; Sweeney et al. 2008), especially the data related to infrastructure systems.

One of the difficulties that has prohibited the study of information diffusion is the lack of sufficient and complete communication data. Traditionally, it is a complicated task to track information sharing among individuals. This issue is less a constraint in online social networks. New social media, like Facebook, Twitter, and Google+, have amassed an unprecedented user population to connect and communicate. In these online social networks, any individual can build and maintain a large number of social connections with hardly any social capital cost (Kwak et al. 2010). Billions of individuals are active on these platforms, sharing their status, opinions, and activities. They provide a tremendous amount of data with rich content and complete history for research. Since such data can have different types of noise and characteristics, it is also important to understand its complexity.

Despite its potential importance as a research area, very limited research on information diffusion in online social networks exists, especially information related to infrastructure systems. This study not only serves to fill this gap, but it also contributes as a first step towards analyzing the interaction between urban stakeholders at micro (individuals and residents), meso (organizations and communities), and macro (institutions) levels (Bosher et al. 2016).[2] We introduce an innovative data collection system to assemble geo-social networking data from a popular social media, i.e. 'tweets' from Twitter (Section 2). We also introduce how we processed and analyzed the data to understand what topics have stimulated interest on social media in Section 2. In Section 3, we construct a communication network based on the hashtags from each tweet and conducted multiple network analyses to reveal the structure and property of the network. We discuss our findings and suggest future directions of this research in Section 4.

[2] Consequently, this work aligns well with the recently developed research roadmap of CIB's W120 – Disasters and the Built Environment.

2. DATA COLLECTION AND PROCESSING

To understand information diffusion about infrastructure systems on social networks, we collected data from one of the largest social media, Twitter. The reasons why Twitter was selected were twofold. First, Twitter, as an open, public platform, attracts a broad user population. The platform amasses over 400 million active users, making it an ideal platform to collect a large quantity of data and study communication and information diffusion. Second, Twitter implements both broadcasting and word-of-mouth (WOM) for information sharing. Research has found that WOM is one of the most effective means to disseminate information. Twitter also allows people to post text messages limited to 140 characters, called "tweets." Users can also choose to let Twitter add location information, called a geotag, to each tweet they post. Each geotag contains the exact geographical coordinate of the tweet. The geolocation helps to define the scale of our study.

2.1 Representativeness of Twitter Users

We first examined demographic information of Twitter users to confirm that Twitter users are representative of the general population. Unfortunately, only limited demographic information of Twitter users was available, of which most is concentrated on U.S. users. Therefore, we constrained our study only to tweets from the U.S. We examined the demographic information of U.S. Twitter users (Bennett 2013) and compared it with demographic information about the general U.S. population (Howden and Meyer 2010; USCB 2012; USCB 2012; USCB 2013; USCB 2014). The comparison and data sources can be found in Table 1, which illustrates that Twitter users are representative of the general population by gender, race, age, and income. A noticeable difference is that while people with higher education tend to user Twitter more, less of the general population has completed higher education.

Table 1. Demographics Comparison

Categories	Demographics	Twitter Users[1] (%)	U.S. Population (%)
Gender	**Men**	50.8	49.2[2]
	Women	49.2	50.8[2]
Age	**18-29**	18.8*	22.0*[4]
	30-49	33.3*	36.1*[4]
	50-64	28.3*	25.0*[4]
	65+	19.6*	16.9*[4]
Race	**White, Non-Hispanic**	70.9	68.1[3]
	Black, Non-Hispanic	9.6	13.1[3]
	Hispanic	11.7	16.9[3]
Education	**High school grad or less**	26.6	42.1[5]
	Some college	30.0	29.0[5]
	College+	42.8	28.9[5]
Income	**Less than $30,000/yr**	22.6	29.9[6]
	$30,000-$49,999	17.9	20.2[6]
	$50,000-$74,999	12.9	18.1[6]
	$75,000+	33.6	31.6[6]

Table 1 is adapted from Wang and Taylor (2015). 1: (Bennett 2013); 2: (Howden and Meyer 2010); 3: (USCB 2014); 4: (USCB 2012); 5: (USCB 2013); 6: (USCB 2012)
*Since data from Twitter users only included those whose ages were above 18, we excluded people under 18 in the general population. Then, we adjusted the values for the rest of the age groups such that the sum equals to 100 percent.

2.2 Process Map for Twitter Data Collection

After confirming the representativeness of Twitter users, we have developed a data collection system to gather Twitter data. The system relies entirely on the Twitter public application program interface (API) (Wang and Taylor 2015). The system generally contains three steps, as shown in Figure 1. The steps are described in the following sections.

2.2.1 Step 1: Collect Geotagged Tweets

The first step was to collect tweets that contained geotags. We created a *tweet streaming module*, which established a continuous connection between a computer in the research lab and the Twitter server. The module streamed every tweet containing a geotag in real-time. *Tweepy*, a Python package for implementing the Twitter API, was used to develop the module. Each streamed tweet contains not only the exact geographical coordinate, but also its ID, place name, name and ID of the user who posted it, and a time stamp. The collected tweets were stored in a database called **Tweets**. A reconnecting mechanism is embedded in the module to ensure that if the continuous connection was lost for 30 seconds, a restart message was displayed and a new connection established.

Following this process, we collected geo-tagged tweets from January 1, 2014 to December 31, 2014. In total, we collected about 1 billion tweets around the world.

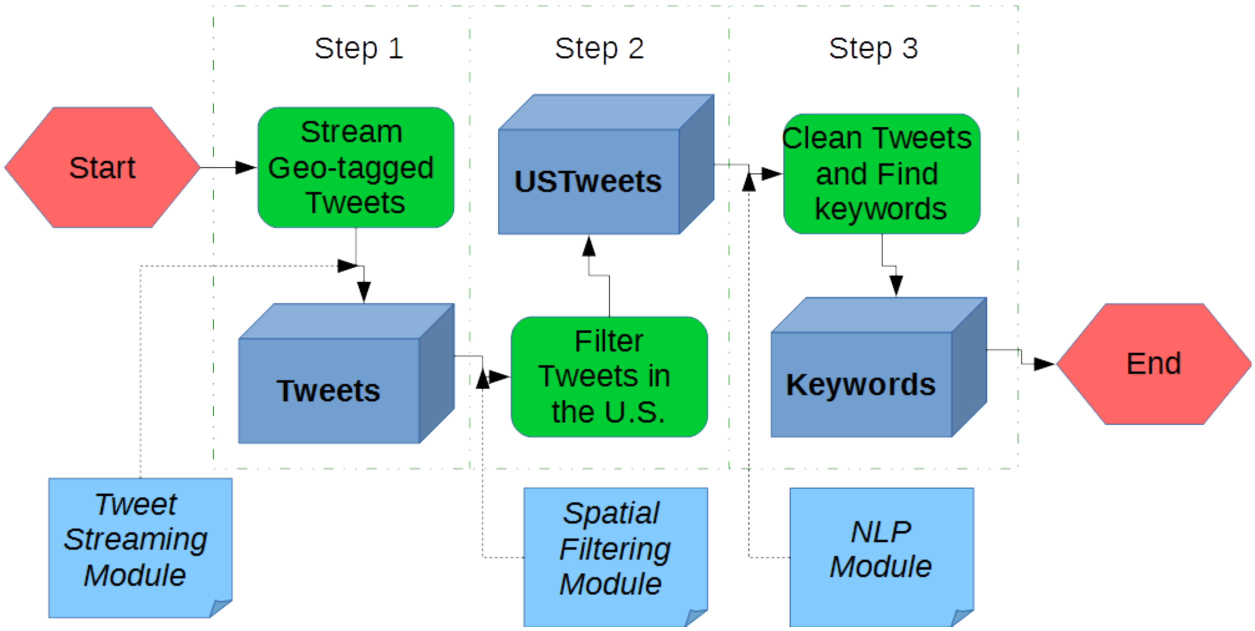


Figure 1. Process Map for Twitter Data Collection

2.2.2. Step 2: Filter Tweets in Specific Geographical Boundaries

This step filters data and retrieves only the ones in certain areas. We conducted preliminary data processing using two filters. First, we defined the geographical boundary of the research area. In this study, we focused our research in the U.S. A shapefile of the U.S. was imported to ArcGIS, and then, we used *ArcPy* to filter all the data and kept only the ones residing within the boundary of the shapefile. We developed a *spatial filtering module* to perform this task. The module retrieved each tweet in the **Tweets** database and determined its geographical coordinate. We filtered the data and retrieved the ones that were located in the United States. If the geographical coordinate of a tweet resides in the U.S., the tweet was examined by the second filter. The second filter was enforced by examining the content of the tweet. If the tweet contained the word "infrastructure", then the tweet was exported into a new database called **USTweets**. After the filtering, there were 6,495 tweets in the **USTweets** database.

2.2.3 Step 3: Clean Tweets and Remove Stop Words

In this step, we cleaned up the tweets for further analysis. First, we retrieved all the content of the tweets and converted any upper case words into lower cases. Second, we converted all the internet addresses, i.e. “www.*” or “https?://*”, to “URL”. Third, we converted any mentions in the tweets, i.e. “@username” to AT_USER. Fourth, we removed all the extra spaces and all the newlines in a tweet. Last, we removed all the hashtag notation, i.e. “#”. Last, we looked for 2 or more repetitions of character and replaced it with the character itself. An example of the cleaning can be found in Table 2.

After processing the tweets, we arrive at the essential word list. We broke down each tweet into different keywords, and then removed all the stopwords. Stopwords were filtered out because they contribute no meaning to a tweet. The list of stopwords is shown in Table 3. Most search engines ignored these words since they are commonly used that improve no precision or recall. An example can be found in Table 2. The processing gave 50,210 keywords. The keywords that appeared over 50 times other than “infrastructure” are listed in Table 4. These 102 words represent about 20 percent of all the keywords used by Twitter users.

Table 2**.** An Example of a Processed Tweet

Before	Winding #conduit and #tunnel entrance to the #PATH #nyc #underground #subway #infrastructure #urban @… http://t.co/Yucfa8RrRc
After	winding conduit and tunnel entrance to the path nyc underground subway infrastructure urban AT_USER URL
Keywords	winding, conduit, tunnel, entrance, path, nyc, underground, subway, infrastructure, urban

Table 3. The List of Stopwords

'a','about','above','across','after','again','against','all','almost','alone','along','already','also','although'
,'always','among','an','and','another','any','anybody','anyone','anything','anywhere','are','area','areas'
,'around','as','ask','asked','asking','asks','at','away','b','back','backed','backing','backs','be','became','
because','become','becomes','been','before','began','behind','being','beings','best','better','between','
big','both','but','by','c','came','can','cannot','case','cases','certain','certainly','clear','clearly','come','co
uld','d','did','differ','different','differently','do','does','done','down','downed','downing','downs','duri
ng','e','each','early','either','end','ended','ending','ends','enough','even','evenly','ever','every','everyb
ody','everyone','everything','everywhere','f','face','faces','fact','facts','far','felt','few','find','finds','firs
t','for','four','from','full','fully','further','furthered','furthering','furthers','g','gave','general','generally'
,'get','gets','give','given','gives','go','going','good','goods','got','great','greater','greatest','group','grou
ped','grouping','groups','h','had','has','have','having','he','her','here','herself','high','higher','highest','
him','himself','his','how','however','i','if','important','in','interest','interested','interesting','interests','
into','is','it','its','itself','j','just','k','keep','keeps','kind','knew','know','known','knows','l','large','largel
y','last','later','latest','least','less','let','lets','like','likely','long','longer','longest','m','made','make','ma
king','man','many','may','me','member','members','men','might','more','most','mostly','mr','mrs','mu
ch','must','my','myself','n','necessary','need','needed','needing','needs','never','new','newer','newest',
'next','no','nobody','non','noone','not','nothing','now','nowhere','number','numbers','o','of','off','often
','old','older','oldest','on','once','one','only','open','opened','opening','opens','or','order','ordered','ord
ering','orders','other','others','our','out','over','p','part','parted','parting','parts','per','perhaps','place','p
laces','point','pointed','pointing','points','possible','present','presented','presenting','presents','proble
m','problems','put','puts','q','quite','r','rather','really','right','room','rooms','s','said','same','saw','say','

says','second','seconds','see','seem','seemed','seeming','seems','sees','several','shall','she','should','show','showed','showing','shows','side','sides','since','small','smaller','smallest','so','some','somebody','someone','something','somewhere','state','states','still','such','sure','t','take','taken','than','that','the','their','them','then','there','therefore','these','they','thing','things','think','thinks','this','those','though','thought','thoughts','three','through','thus','to','today','together','too','took','toward','turn','turned','turning','turns','two','u','under','until','up','upon','us','use','used','uses','v','very','w','want','wanted','wanting','wants','was','way','ways','we','well','wells','went','were','what','when','where','whether','which','while','who','whole','whose','why','will','with','within','without','work','worked','working','works','would','x','y','year','years','yet','you','young','younger','youngest','your','yours','z','AT_USER','URL'

Table 4. The Frequency of Keywords

Keyword	Frequency	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
jobs	249	bill	90	fund	68	bridges	56
people	175	obama	89	development	68	talk	56
build	165	spending	89	spend	68	own	56
building	163	cloud	89	system	67	government	55
water	163	business	88	technology	66	key	55
city	159	internet	86	cost	66	billion	55
money	159	data	85	lack	66	improve	54
investment	156	care	85	tech	65	community	53
public	155	green	84	future	65	lot	53
support	138	pay	83	traffic	65	stop	53
world	136	economy	81	taxes	65	ready	52
bike	136	war	80	talking	65	road	52
invest	136	funding	80	etc	64	services	52
transportation	135	power	79	issues	64	gop	52
education	135	cities	79	economic	63	repair	51
time	131	growth	79	love	62	start	51
america	127	built	75	change	60	network	51
health	112	country	74	look	60	american	51
help	110	investing	74	social	60	president	51
crumbling	110	job	74	schools	59	day	51
tax	109	poor	71	bridge	59	major	50
energy	100	create	71	transit	57	rebuild	50
projects	99	real	70	focus	57	issue	50
fix	94	bad	69	congress	56	yes	50
roads	94	local	69	via	56	gas	50
critical	94			plan	56		

Using all the keywords and their frequencies, we generated a word cloud as shown in Figure 2. The word cloud was created using the python package "word_cloud". From the frequency and word cloud, we can identify which topics have sparked interests among Twitter users. While infrastructure systems are clearly important aspects of civil engineering and construction, these systems impact social and economic sy. The word cloud clearly shows how people have related "infrastructure" to "people", "jobs", "money", "investment", etc. Such knowledge can be critical when policymakers seek support from different stakeholders.

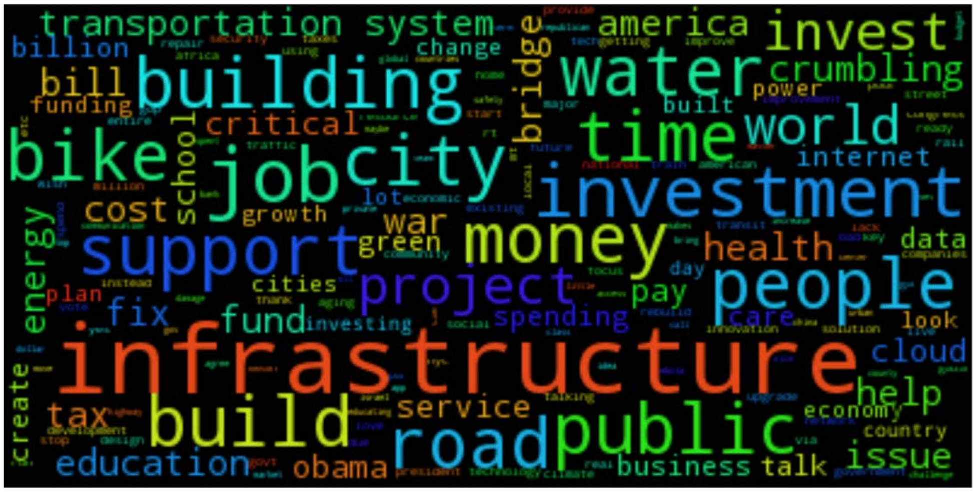


Figure 2. Word Cloud of Keywords Related to Infrastructure

3. COMMUNICATION NETWORK ANALYSIS

Beyond understanding the topics addressed, recognizing how stakeholders have shared these topics with one another is equally significant. Twitter is used to broadcast and share information and opinions. Therefore, we studied how individuals have shared tweets regarding infrastructure.

3.1. Constructing the Hashtag Network

We reexamined the tweets and all the hashtags inside the tweets. We retrieved all the hashtags in the content of each tweet. A hashtag is a word with a "#" at the front of it. A hashtag is used to mark keywords or topics in a tweet which is created organically by Twitter users to categorize messages. After processing the data, we found 4,499 hashtags, with 2,746 unique ones.

Then, we created a hashtag network. If two different tweets from two different users share the same hashtag, then we added a link between them. Altogether, 56,468 links were created. Using all the links, we developed the network (Figure 3). Each node represents a user. Each edge represents the total number of hashtags shared by a dyad of two users, which is also known as *weight*. The size of the nodes represents the weighted degrees, i.e. the total weights of edges that connect to a node. The opaqueness of an edge is determined by the weight of it. The layout of the network is generated using the Fruchterman-Reingold algorithm.

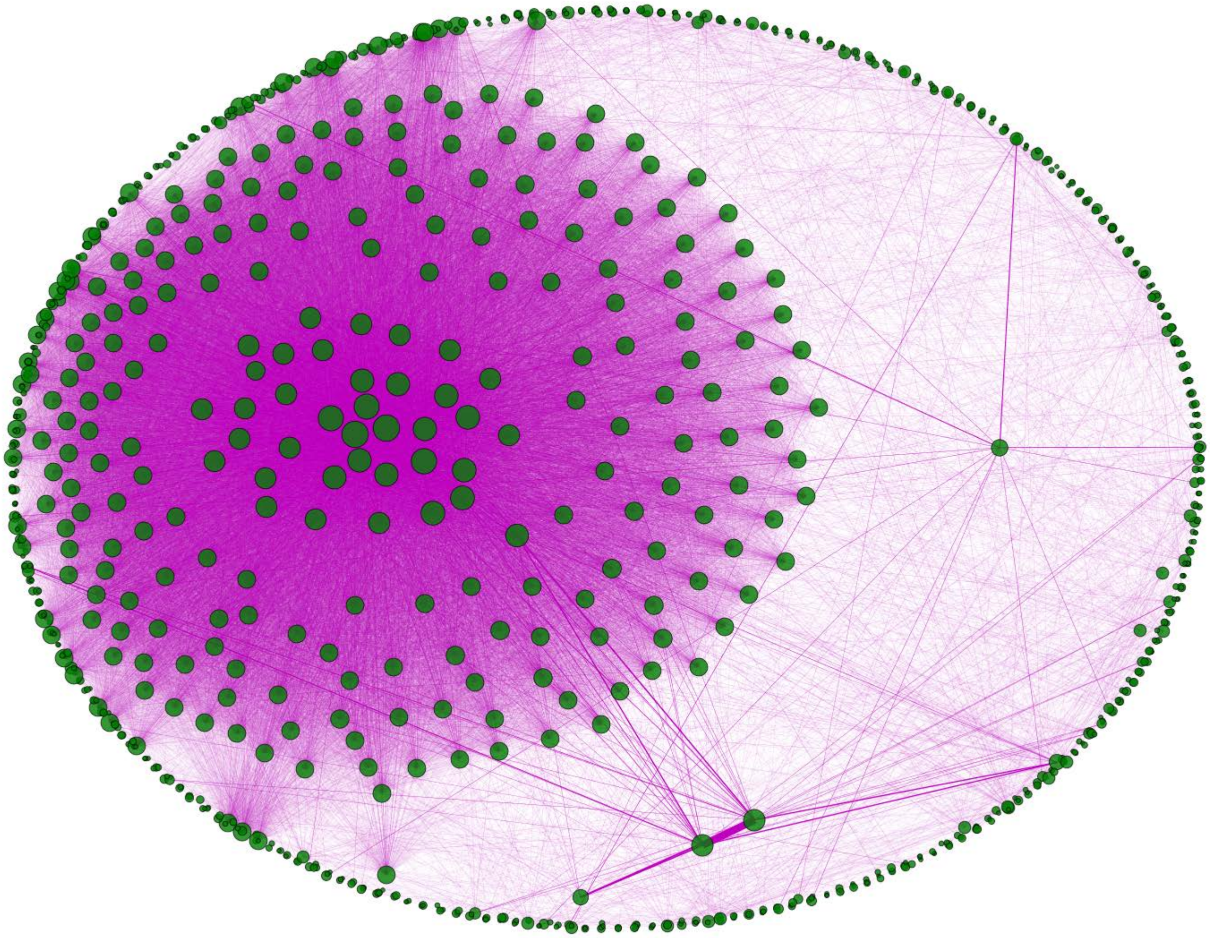


Figure 3. Hashtag Network

3.2 Network Degree

We analyze the degree of the network to understand how communications are connected. There are a total of 988 nodes, i.e. individual Twitter users, in the network. We calculated the weighted degree of each node. The results show that while 181 individuals out of the 988 (about 18.3%) only have a degree of 1, an individual has a weighted degree of 1,570. The distribution of degrees can be found in Figure 4. The chart demonstrates that the general distribution follows a scaling law that has a heavy tail. The results demonstrate that while a large portion of individuals are only casually sharing information related to infrastructure, some people are extremely active.

They hold the central positions of information diffusion in the network and should potentially be treated as advocates for infrastructure, and infrastructure issues such as resilience.

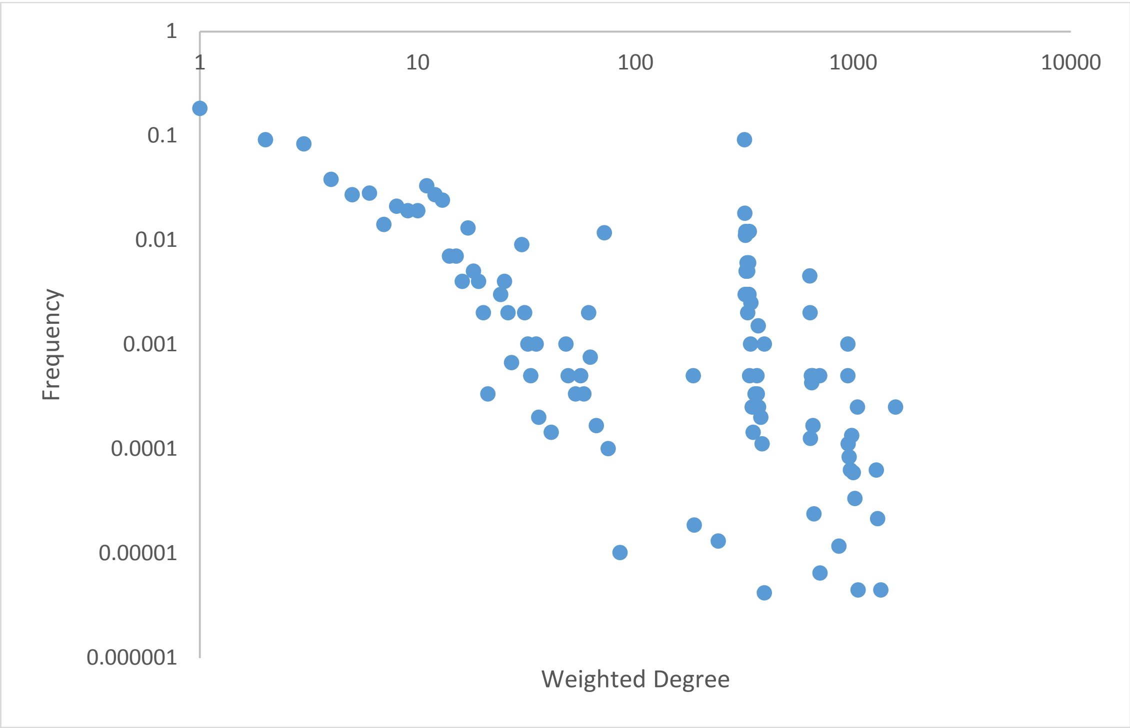


Figure 4. Distribution of Weighted Degrees

3.3. Components

The hashtag network has multiple components. Using *networkx*, a python package, we conducted multiple analyses to study the components of the network. Based on our analysis, there are 88 components. They include one giant component which contains 777 nodes out of the 988, and 87 small components. Table 5 summarizes the components in the network: 2 components contain 6 nodes, 1 component contains 5 nodes, 5 components contain 4 nodes, and so on.

Table 5 Summary of Components

Number of Nodes Included	Number of Components
2	63
3	16
4	5
5	1
6	2
777	1

The break-down of components demonstrates that the sharing of infrastructure information is a collective behavior. The giant component consists of most of the individuals, almost 80 percent of all of the users studied. Information diffusion of infrastructure tends to reach a broad audience.

Then, we analyzed the configurations of the components. We studied each of the components and find only 11 different configurations. Except the complex configuration from the giant component, the other ten configurations showed a quite simple structure, as depicted in Figure 5. The results reveal that information diffusion takes very simple forms. It provides a trajectory to break down the complexity of information diffusion in online social networks and to uncover the foundation of the motivations, decisions, and paths of information sharing among different stakeholders.

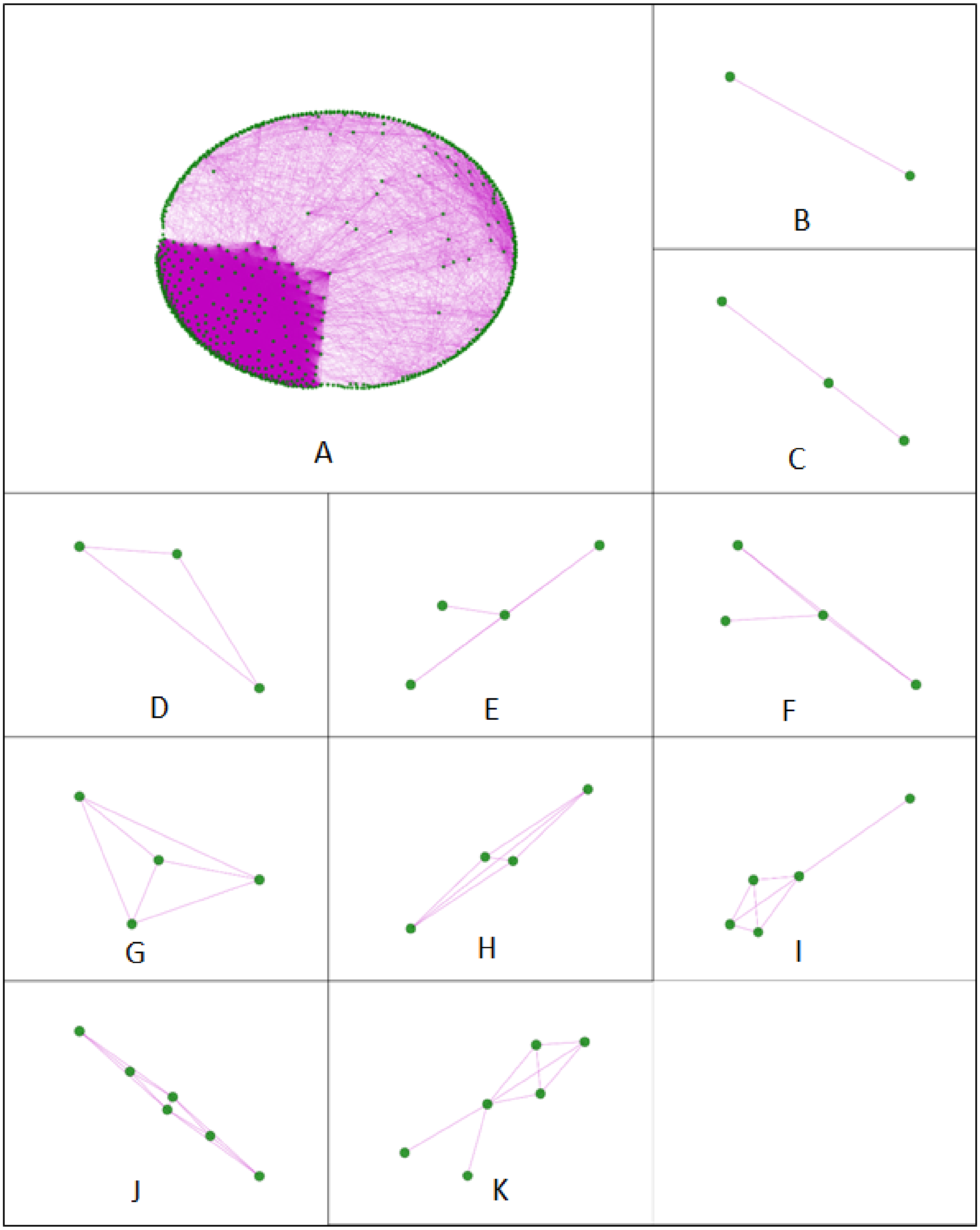


Figure 5. Configurations of Components.
Although it is a complex communication network, the hashtag network can be decomposed into 11 components. **A** shows the giant component which includes almost 80 percent of all the individuals. **B – K** is the rest of the components which all have relatively simple structures.

4. DISCUSSION, CONCLUSIONS AND FUTURE STUDY

4.1 Discussion and Conclusions

Infrastructure resilience is a critical topic for all cities globally. The U.S. only scored D+ on its infrastructure system based on a report from ASCE in 2013. Approximately 3.6 trillion dollar is needed by 2020 to fix the deteriorating system. Even more pressing, natural hazards, which are exacerbating due to the influence of climate change, can exert increasing impact on urban infrastructure. In the event of natural hazards, infrastructure failures can lead to inconveniences, such as traffic jams and public transportation delays, to more catastrophic consequences, such as human injuries and fatalities. We need to motivate all stakeholders to take actions to improve infrastructure system performance, particularly during hazardous natural events when accurate information about safe locations, system outages, etc. is crucial.

Effective actions require accurate, sufficient, and prompt information. Therefore, this project focused on how Twitter, a popular online social networking platform, has been used for information sharing. We created a data collection system to stream tweets with geographical coordinates embedded in the tweet. In this way, we were able to retrieve all the data from the United States. Our analysis focused on the geo-tagged tweets for the year of 2014. This innovative data collection system not only contributes sufficient data for this report, but it also can be used to study other areas of urban infrastructure.

We analyzed the words shared by Twitter users. Using natural language processing, we cleaned each tweet, removed stop words and analyzed the frequencies of the word. The analysis shows that different stakeholders care about different aspects of infrastructure. The word cloud developed indicates relationships between infrastructure and a variety of issues. Improving U.S.

infrastructure is a complicated problem, so understanding the concerns from different parties is important; the Twitter data collected helps to demonstrate such concerns and how it might be used in public engagement efforts.

We used Twitter hashtags as links and studied a communication network regarding U.S. infrastructure. The effort generated a massive hashtag network. While the degrees of the individuals in the network still roughly follows a scaling law, the network has a surprisingly high connectivity compared to other communication networks found on Twitter. This confirms that infrastructure provokes interest and discussions among individuals and organizations. This finding demonstrates that policy makers should seriously consider the utilization of online social media to provoke discussions and generate feedback.

We also identified the basic components of the hashtag network. While the network has a complex structure, it can be broken down into different components. Other than one giant component, most of the others have simple structures. This finding provides an important insight to understand the elements of the infrastructure communication network within a social media platform. It might contribute to developing simulations like agent-based models.

4.2 Future Study

Though an innovative effort, the study can be extended to several directions. First, the study can be extended to non-geotagged tweets. We only examined geotagged tweets. Our database generated 6,495 tweets, a sufficient amount to conduct this research. However, most tweets are not geotagged, so it is possible that many tweets related to infrastructure are not included. Examining them may provide more insights about information diffusion on Twitter.

Second, it will be beneficial to study the characteristics of the stakeholders who are involved in the information diffusion. In line with Bosher (2016), stakeholders can possibly be divided into three different levels (micro, meso, and macro) and can include different categories, such as government, non-profit organizations, private companies, individuals, etc. Analyzing their opinions and their roles in information sharing can help us understand whether there are gaps in communication and how we can eliminate them. This can also shed light on evaluating the different levels of engagement between individuals, organizations, and institutions.

Finally, geotagged Twitter data could be analyzed during known events such as tornadoes, winter storms or power outages to determine what and how Twitter users communicate during such events. These types of analysis would allow greater comprehension of the potential of social media for pre, per and post event communication, which is critical to hazardous event preparedness, response and recovery. Indeed, social media will likely play an important role in such events and in future infrastructure performance and resilience.

5. REFERENCES

Allsop, D. T., Bassett, B. R., and Hoskins, J. A. (2007). "Word-of-mouth research: Principles and applications." *Journal of Advertising Research*, 47(4), 398-411.

Arndt, J. (1967). "Role of Product-Related Conversations in the Diffusion of a New Product." *Journal of Marketing Research*, 4(3), 291-295.

Beasley, D. (2014). "Deadly ice storm turns Atlanta into parking lot, strands thousands." <http://www.reuters.com/article/2014/01/29/us-usa-weather-idUSBREA0Q1DK20140129>. (December 10, 2014).

Bennett, S. (2013). "Facebook, Twitter, Instagram, Pinterest, LinedIn: 2013 Social Media Demographics." <http://www.mediabistro.com/alltwitter/social-demographics-2013_b53515>.

Bosher, L., Von Medling, J., Johnson, C., Arefian, F.,Chmutina, K., and Chang-Richards, Y. (2016). "Disasters and the Built Environment: Research Roadmap Report for Consultation." International Council for Research and Innovation in Building and Construction, Delft, Netherlands.

Brown, J. J., and Reingen, P. H. (1987). "Social Ties and Word-of-Mouth Referral Behavior." *Journal of Consumer Research*, 14(3), 350-362.

CDC (2013). "Deaths Associated with Hurricane Sandy — October–November 2012." C. f. D. C. a. Prevention, ed.Atlanta, GA.

Dai, A. (2013). "Increasing drought under global warming in observations and models." *Nature Climate Change*, 3(1), 52-58.

File, K. M., Cermak, D. S. P., and Prince, R. A. (1994). "Word-of-mouth effects in professional services buyer behaviour." *Service Industries Journal*, 14(3), 301-314.

Goldenberg, J., Libai, B., and Muller, E. (2001). "Talk of the network: A complex systems look at the underlying process of word-of-mouth." *Marketing letters*, 12(3), 211-223.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., and Gremler, D. D. (2004). "Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?" *Journal of interactive marketing*, 18(1), 38-52.

Howden, L. M., and Meyer, J. A. (2010). "Age and Sex Composition: 2010. 2010 Census Briefs.", <http://www.census.gov/prod/cen2010/briefs/c2010br-03.pdf.>.

Knutson, T. R., McBride, J. L., Chan, J., Emanuel, K., Holland, G., Landsea, C., Held, I., Kossin, J. P., Srivastava, A., and Sugi, M. (2010). "Tropical cyclones and climate change." *Nature Geoscience*, 3(3), 157-163.

Kwak, H., Lee, C., Park, H., and Moon, S. "What is Twitter, a social network or a news media?" *Proc., Proceedings of the 19th international conference on World wide web*, ACM, 591-600.

Landsea, C. W., Vecchi, G. A., Bengtsson, L., and Knutson, T. R. (2010). "Impact of Duration Thresholds on Atlantic Tropical Cyclone Counts." *Journal of Climate*, 23(10), 2508-2519.

Meehl, G. A., Washington, W. M., Collins, W. D., Arblaster, J. M., Hu, A., Buja, L. E., Strand, W. G., and Teng, H. (2005). "How much more global warming and sea level rise?" *Science*, 307(5716), 1769-1772.

Moteff, J. D. (2012). *Critical Infrastructure Resilience: The Evolution of Policy and Programs and Issues for Congress*, Congressional Research Service.

Obama, B. (2013). "Improving critical infrastructure cybersecurity." *Executive Order, Office of the Press Secretary*, 12.

Oerlemans, J. (1994). "Quantifying global warming from the retreat of glaciers." *Science*, 264(5156), 243-245.

Pan, X., Han, C. S., Dauber, K., and Law, K. H. (2007). "A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations." *Ai & Society*, 22(2), 113-132.

Reingen, P. H., and Kernan, J. B. (1986). "Analysis of Referral Networks in Marketing: Methods and Illustration." *Journal of Marketing Research*, 23(4), 370-378.

Stern, P. C. (1992). "What psychology knows about energy conservation." *American Psychologist; American Psychologist*, 47(10), 1224.

Stern, P. C., Aronson, E., Darley, J. M., Hill, D. H., Hirst, E., Kempton, W., and Wilbanks, T. J. (1986). "The effectiveness of incentives for residential energy conservation." *Evaluation Review*, 10(2), 147-176.

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M. (2013). "Climate change 2013: The physical science basis." *Intergovernmental Panel on Climate Change, Working Group I Contribution to the IPCC Fifth Assessment Report (AR5)(Cambridge Univ Press, New York)*.

Sweeney, J. C., Soutar, G. N., and Mazzarol, T. (2008). "Factors influencing word of mouth effectiveness: receiver perspectives." *European Journal of Marketing*, 42(3/4), 344-364.

Teves, O., and Bodeen, C. (2013). "Haiyan Storm Surges Caught Philippines by Surprise." <http://bigstory.ap.org/article/haiyan-storm-surges-caught-philippines-surprise>. (November 18, 2014).

UNISDR (2013). "Disasters Impacts 2000-2012."

United Nations (2009). "Urban and Rural Areas 2009." United Nations, Department of Economic and Social Affairs, Population Division, United Nations, New York.

USCB (2012). "Age and Sex Composition in the United States: 2012. Table 1. Population by Age and Sex: 2012.", <http://www.census.gov/population/age/data/files/2012/2012gender_table1.xls>.

USCB (2012). "Income, Expenditures, Poverty, & Wealth: Household Income.", <http://www.census.gov/compendia/statab/cats/income_expenditures_poverty_wealth/household_income.html>.

USCB (2013). "Educational Attainment in the United States: 2013 – Detailed Tables. ." <https://www.census.gov/hhes/socdemo/education/data/cps/2013/tables.html>.

USCB (2014). "People QuickFacts." <http://quickfacts.census.gov/qfd/states/00000.html.>.

Vilpponen, A., Winter, S., and Sundqvist, S. (2006). "Electronic word-of-mouth in online environments: Exploring referral network structure and adoption behavior." *Journal of Interactive Advertising*, 6(2), 71-86.

Wang, Q., and Taylor, J. (2015). "Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms." *Journal of Computing in Civil Engineering*, 0(0), 04015004.

Wang, Q., and Taylor, J. (2015). "Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms." *Journal of Computing in Civil Engineering*, 04015004.

Wang, Q., and Taylor, J. E. "Energy saving information cascades in online social networks: An agent-based simulation study." *Proc., Winter Simulation Conference (WSC), 2013*, 3042-3050.

Wang, Q., and Taylor, J. E. (2014). "Quantifying Human Mobility Perturbation and Resilience in Hurricane Sandy." *PLoS ONE*, 9(11), e112608.

Webster, C. M., and Morrison, P. D. (2004). "Network analysis in marketing." *Australasian Marketing Journal (AMJ)*, 12(2), 8-18.

World Bank (2013). "World Development Indicator: Urban Population." The World Bank Group, New York City, New York.

CIB General Secretariat
post box 516
2600 AM Delft
The Netherlands
E-mail: secretariat@cibworld.nl
www.cibworld.nl

CIB Publication 409 / ISBN 978-9-063630-9-42