Title
Revisiting the death of geography in the era of Big Data: the friction of distance in cyberspace and real space

Journal
International Journal of Digital Earth, 11(5)

ISSN
1753-8947

Authors
Han, Su Yeon
Tsou, Ming-Hsiang
Clarke, Keith C

Publication Date
2018-05-04

DOI
10.1080/17538947.2017.1330366

Peer reviewed

Taylor & Francis
Taylor & Francis Group

International Journal of Digital Earth

ISSN: 1753-8947 (Print) 1753-8955 (Online) Journal homepage: http://www.tandfonline.com/loi/tjde20

Revisiting the death of geography in the era of Big Data: the friction of distance in cyberspace and real space

Su Yeon Han, Ming-Hsiang Tsou & Keith C. Clarke

INTERNATIONAL JOURNAL OF DIGITAL EARTH, 2017
https://doi.org/10.1080/17538947.2017.1330366

Revisiting the death of geography in the era of Big Data: the friction of distance in cyberspace and real space

Su Yeon Han[a,b], Ming-Hsiang Tsou [a] and Keith C. Clarke [b]

[a]Department of Geography, San Diego State University, San Diego, CA, USA; [b]Department of Geography, University of California, Santa Barbara, CA, USA

ABSTRACT

Many scholars have argued that the importance of geographic proximity in human interactions has been diminished by the use of the Internet, while others disagree with this argument. Studies have noted the distance decay effect in both cyberspace and real space, showing that interactions occur with an inverse relationship between the number of interactions and the distance between the locations of the interactors. However, these studies rarely provide strong evidence to show the influence of distance on interactions in cyberspace, nor do they quantify the differences in the amount of friction of distance between cyberspace and real space. To fill this gap, this study used massive amounts of social media data (Twitter) to compare the influence of distance decay on human interactions between cyberspace and real space in a quantitative manner. To estimate the distance decay effect in both cyberspace and real space, the distance decay function of interactions in each space was modeled. Estimating the distance decay in cyberspace in this study can help predict the degree of information flow across space through social media. Measuring how far ideas can be diffused through social media is useful for users of location-based services, policy advocates, public health officials, and political campaigners.

ARTICLE HISTORY
Received 24 October 2016
Accepted 10 May 2017

KEYWORDS
First law of geography; distance decay; Big Data; Twitter; social media; information diffusion

1. Introduction

Long a focus of geographical spatial analytic theory, the effect of distance on spatial interaction has been examined by many scholars over the past decades (for a definition of spatial interaction, see the Background). Tobler (1970) suggested the concept of distance decay in proposing his first law of geography: 'everything is related to everything else, but near things are more related than distant things.' Distance decay describes how spatial interaction decreases with increasing distance between two places because of the penalties in travel time and cost associated with longer distances. This effect has been termed the 'friction of distance.'

However, the influence of distance on spatial interaction has been controversial since the Internet became a means of mass communication in the late 1990s. Since then, many scholars have devalued the effect of distance decay on spatial interaction, calling this trend the 'death of geography' (Bates 1996), 'end of geography' (O'brien 1992), or the 'death of distance' (Kolko 2000; Cairncross 2001). It is argued that physical proximity in spatial interaction does not matter in cyberspace since people living in faraway towns are as accessible as your immediate neighbors through the Internet. The

CONTACT Ming-Hsiang Tsou mtsou@mail.sdsu.edu Department of Geography, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182-4493, USA
Supplemental data for this article can be accessed at https://doi.org/10.1080/17538947.2017.1330366.

death of geography because of the Internet has been the topic of many studies using different terms such as the emergence of a 'borderless world (Green and Ruhleder 1995; Ohmae 1995)', 'global village' (Fiore and McLuhan 1967; McLuhan and Powers 1989), and 'telecottages' (Toffler 1980).

On the other hand, others have argued that distance is still very much alive (Morgan 2004; Polèse and Shearmur 2004). Some scholars have noted that there are more spatial interactions among individuals living in nearby places than with those in faraway places in cyberspace (Takhteyev, Gruzd, and Wellman 2012; Han, Tsou, and Clarke 2015). As opposed to the death of geography, Goldenberg and Levy (2009) counter-intuitively argued that the friction of distance on spatial interaction has increased with the Internet revolution. Despite considerable debate about the role of distance in spatial interaction, studies rarely demonstrate quantitative evidence to show the difference in the role of distance between cyberspace and real space, nor do they investigate how much difference exists in the influence of distance between the two spaces.

The first goal of this study is to reveal the difference in the distance decay effect on spatial interaction between cyberspace and real space in a quantitative manner. To empirically quantify the rate of distance decay in cyberspace and real space, we used social media data (Twitter). The distance decay curve describes by how much the frequency of interaction decreases with an increase in distance, usually in the form of a power law. We expected to find different shapes of the distance decay function between cyberspace and real space. Among Twitter users living in different cities, interactions on Twitter do not require travel cost and time, while face-to-face interaction does. Therefore, our hypothesis was that the distance decay curve in cyberspace will show a more gentle decline than the distance decay curve in real space, since it is less subject to the friction of distance. The second part of this study includes spatial analyses and visualizations of the effect of distance decay in both cyberspace and real space.

This research is intended to address the following questions:

- Is the distance decay effect measurable for human interactions in cyberspace as well as real space? If so, how is the distance decay function different?
- What is the quantitative difference in the influence of distance on spatial interactions between cyberspace and real space?
- What are the differences in the geographic patterns of related spatial interactions between cyberspace and real space?

2. Background

2.1. Term definitions and assumptions

Many scholars have defined spatial interaction in different contexts. A simple definition of spatial interaction is 'the flow of products, people, services, or information among places, in response to localized supply' (Swanson and Tayman 2012). MacLachlan (2011) defined spatial interaction in more detail:

> Spatial interaction is a dynamic flow process that articulates one location with another. It is a general concept that may refer to the movement of human beings such as intra-urban commuters or intercontinental migrants but may also refer to traffic in goods such as raw materials or to flows of intangibles such as information.

Fotheringham and O'Kelly (1989) demonstrated some specific examples of spatial interactions such as 'migration, shopping trips, commuting, trips for recreational purposes, trips for educational purposes, freight flows, the spatial pattern of telephone calls, emails and world-wide web connections, and even the use of healthcare facilities'. As the examples show, he defined spatial interaction as any connection between the two places. We also define spatial interaction in a broad context: spatial interaction can be either actual or potential flow among places, with any type of connection among places. In this study, we assume that spatial interaction does not necessarily involve communication or the movement of tangible objects among places. Just looking at messages of people living in other places on the Internet is a spatial interaction. Also, recognizing place names is a spatial interaction.

Based on our definition of the spatial interaction in this broad contexts, in this study, we assume that there are spatial interactions in the following cases. First, when Twitter users travel to other city, we assume that there is spatial interaction between the city where the users live and the city to which they travel. Second, when Twitter users mention city names in their messages, we assume that there is spatial interaction between the city where the users live and the city names that they mention. When someone hears a news item relating to a city, they may be simply mentioned the city name to talk about the news. In this case, there is no communication between people living in the different cities. However, as we defined above, we defined the spatial interaction as any type of connection between places. Third, when Twitter users are followed by other users, we assume that there are spatial interactions between the users and their followers. Connected users on the Twitter website are referred to as 'following' or 'follower.' For example, if you are 'following' John, you can see John's updates such as messages, and the postings that John is sharing on your timeline. 'Followers' of John are people who can see John's updates. The users that John follows or who are followers of John are often his friends, acquaintances, or a person or organization that John is interested in. Therefore, we assume that those connected users in the Twitter website can be an indicator of spatial interactions between users living in different physical places.

The term cyberspace as a metaphor for the Internet and World Wide Web was coined by William Gibson (Gibson 1984) and became popular with many scholars, especially geographers adopting the term (e.g. Zook and Graham 2007; Couclelis 2009; Kellerman 2010; Devriendt et al. 2011; Tsou et al. 2013). However, the use of the term 'cyberspace' as a metaphor for the Internet has been criticized (Graham 2013; Leszczynski 2015). In this study, cyberspace refers to the notional environment where humans interact or communicate over the computer networks or via the Internet.

2.2. Distance decay of interactions in the pre- and post-Internet eras

Distance decay is inherent in spatial gravity models, but the slope and range of the distance decay vary depending on the type of human interaction (Fellmann, Getis, and Getis 2007). For example, Mok and Wellman (2007) found that both the frequency of face-to-face contacts and phone calls decrease with increasing distance – they both follow the distance decay function – but the slope of the distance decay function of phone contact is less steep than that for face-to-face contact. On the other hand, Mok, Wellman, and Carrasco (2010) examined the frequency of emails, phone calls, and face-to-face contacts between the 1970s and the 2000s, and compared frequencies of human interaction by different means of communication between the pre- and post-Internet eras. They found that the effect of distance decay in human interaction by each means of communication was similar for the pre- and post-Internet eras – that is, the importance of distance remained the same between the 1970s and the 2000s. Kellerman (2016) envisages that probably this trend may have continued into the era of social networking – that is, even though people are well connected through social networking services, the importance of distance in human communications will possibly remain the same as the pre-Internet era. Mok et al.'s previous studies focused on one-to-one communications to examine the effect of distance decay of interactions via emails, phone calls, and face-to-face contacts. However, our study focused on one-to-many communications on the social networking platform (Twitter) to examine the effect of distance in cyberspace. The way communications work on social networking sites is described in Section 2.4.

Scholars have argued that human interactions in cyberspace and real space are not entirely separate, but are interrelated. Warf (2013) argued that the distance decay effect for human interactions in real space cannot help creating a distance decay effect on human interactions in cyberspace. Rainie and Wellman (2012) noted 'the more Internet contact, the more in-person and phone contact' among friends and family members. Kellerman (2016) also noted that communications online are usually accompanied by face-to-face communications with the same person, and 'online networking may facilitate the development of stronger ties among individuals, as long as there exist already some social ties among these individuals in real space'. In fact, Stephens and Poorthuis (2015) showed that 'Twitter networks primarily replicate existing social and spatial patterns, such as flights and telecommunications patterns'.

2.3. Social media data as a proxy for human mobility observations

The feasibility of using large-scale Twitter data as a proxy of human mobility has recently been explored (Hawelka et al. 2014; Jurdak et al. 2015; Liu et al. 2015). The coordinates of users' locations are recorded when they post social media messages, and can be used to locate the users. Some Twitter messages include the coordinates of each user's location and are called geo-tagged tweets. Scholars have argued that Twitter can be a proxy for global mobility patterns and for city-scale human mobility (Hawelka et al. 2014; Wang and Taylor 2015). Jurdak et al. (2015) revealed that geo-tagged tweets can capture the diverse trajectories of individuals and movements within and between cities. Liu et al. (2015) showed that there is a correlation between the distribution of Twitter users and the population distribution reported in the census. As assumed in this study, this research demonstrated that geo-tagged tweets can reveal patterns of human movement. We assume that when people travel they have interactions with someone living in the places they visit.

Prior research also found that human travel patterns captured by geo-tagged messages can be approximated by a power law (Hawelka et al. 2014; Jurdak et al. 2015) and loosely follow the gravity model (Liu et al. 2015). On the other hand, according to Noulas et al. (2012), 'a rank-based movement model accurately captures real human movements in different cities'. These studies used social media data to reveal human travel patterns and find universal laws representing the effect of distance decay on human movements. However, these studies failed to investigate the influence of distance on communications in social networking sites. By examining the human network for Twitter, the patterns of interactions in cyberspace with increasing distance can be examined.

2.4. The nature of social media

Social media are revolutionary forms of communication and have unique characteristics that are differentiated from other conventional forms of communication such as televisions, radio, telephone calls, emails, and video chatting.

First, unlike the traditional forms of communication, social media provide multimedia-style platforms and asynchronous communications. Social media provide user-friendly and interactive interfaces that help their users to share news, advertisements, and their daily lives easily and promptly by effectively posting hyperlinks, images, and layouts of shared messages. When the users open up their own page on social networking sites, they often can see all the shared postings from their friends on the first page. When they share their postings or leave comments on the postings, they do not need to worry about other people's availability at the time when they share items or respond. Once they post their messages, their friends can see the posting anytime and voluntarily reply to their postings whenever they like.

Second, social media are self-chosen and self-filtered news channels via personal networks. Broadcast media such as radio and television usually deal with public concerns such as riots in Baltimore, earthquakes in Nepal, or terror attacks in Bangkok. Unlike broadcast media, in social networking sites, users can spontaneously pick the topic they would like to talk about. For this reason, an individual user's trivial interest can go viral in social networking sites. One example of a viral diffusion of the social media user's trivial interest is the dispute about the color of a dress, on 26 February 2015. The dispute was started from a washed-out photograph of a dress posted by one user to social networking sites. Within a few days, the photo went viral and drew more than 10 million Twitter users' attention to dispute over whether the dress pictured was blue and black or white and gold. The dispute implies that even trivial issues and socially crucial issues have the potential for spreading extensively and attracting many peoples' attention via social networking.

Third, through social media, millions of people can communicate with one another about the same issue spontaneously and simultaneously. Information that is contained in the social media messages can be copied and cross-posted by massive numbers of social media users who are interconnected to one another by world-wide networks, so the messages can increase exponentially and diffuse extensively. Therefore, the content of messages spreading through social networks has the greatest social, political, and cultural ripple effect compared to other types of media such as televisions, radios, and emails.

Due to these characteristics of social media, the contents of social media messages embody dynamically changing records of human activities, individual's ideas and perceptions. Therefore, the statistical exploration and data mining of social media messages allow us to understand human dynamics.

2.5. The use of social media and Big Data to understand human dynamics

Several studies show that studying the relationship between cyberspace and real space using vast quantities of social media data (Big Data) can help to produce insightful knowledge to understand human activities. Some scholars (Laney 2001; McAfee et al. 2012; White 2012; Kwon and Sim 2013) have characterized Big Data using the term, 'the 3Vs': 'volume' to highlight the large volume of data such as terabytes or petabytes of data; 'velocity' which means a high speed of data generation, data being created in or almost in real time; and 'variety' to represent that data are created from various sources and are diversely structured. In terms of studying the complexities of human dynamics, Tsou (2015) redefined Big Data as 'a large dynamic dataset created by or derived from human activities, communications, movements, and behaviors … . The term Big Data refers to big ideas, big impacts, and big changes for our society in addition to big volume of datasets'. Social media data (Big Data) have been used for tracking flu (Signorini et al. 2011; Nagel et al. 2013; Aslam et al. 2014), anticipating election results (Tsou et al. 2013), understanding social networks (Leskovec and Horvitz 2008; Lerman and Ghosh 2010), understanding human emotions (Kamvar and Harris 2011; Shook et al. 2012), and for disaster management (Sakaki, Okazaki, and Matsuo 2010; MacEachren et al. 2011). In the present study, we seek to examine the nature of the friction of distance in both real and cyberspace using social media data with a large number of records. Social media interactions are assumed to closely approximate actual human interactions, and Twitter data are assumed to be a representative sample of social media exchanges and interactions in the context of a sample of four US cities.

3. Study area

We select the four largest US cities – New York, Los Angeles, Chicago, and Houston – to examine the distance decay effect for spatial interaction among Twitter users. We compared Twitter users living in these four cities with those living outside the four cities. The reason why we selected the four cities is that they are large enough to produce Big Data for tweets. In addition, each city is located in a geographically different region. Figure 1 shows the census boundary of the selected four cities and other cities in the US. This boundary, the metropolitan statistical area, was downloaded from the website of the United States Census Bureau. The estimated population in 2014 is available from LandScan (Laboratory 2016). The map indicates that this boundary covers the whole region of each individual city where most of the US population lives. In fact, 99.7% of all US tweets were created in these metropolitan areas. The tweets that were created within those cities were used for analysis.

4. Research design and methodology

To understand the patterns of spatial interaction in cyberspace and in real space, the methodological framework of this study was based on inductive reasoning, a 'bottom up' logic where a conclusion is a

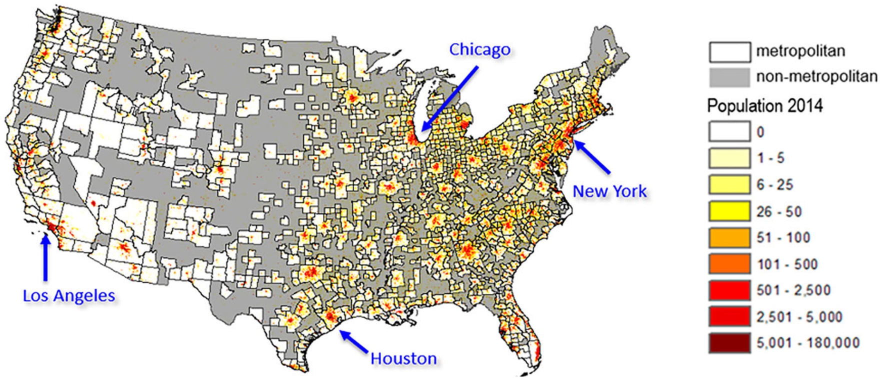

Figure 1. The boundaries of the census metropolitan areas in the US. The city boundaries were used to identify the home cities of Twitter users and the four selected cities: New York, Chicago, Los Angeles, and Houston.

broad generalization from many observations of human activities, as revealed by massive amounts of Twitter data. The use of inductive reasoning for discovering new knowledge contained in the tweets in this study is based on the fourth paradigm of data-driven science and has to do with 'new forms of empiricism that declare 'the end of theory', the creation of data-driven rather than knowledge-driven science'(Kitchin 2014).

This study consisted of following four procedures (Figure 2). Each subsection describes the tasks in each step.

4.1. Step 1: Data collection by streaming APIs, cleaning, and preprocessing

The first step was to collect Twitter data. Twitter is an Internet-based social networking and micro-blogging service that allows registered users to post messages up to 140 characters in length. According to the third quarter of 2015 statistics, Twitter has an average of 307 million monthly active users.

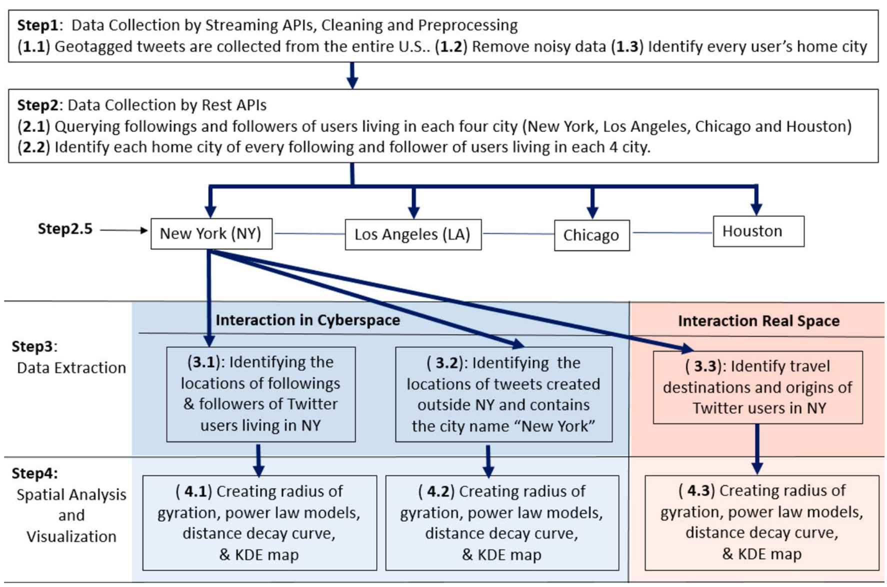

Figure 2. The four steps of this study.

Twitter Streaming API allows us to collect geo-tagged tweets created within a bounding box that the user specifies. In this study, a bounding box that covers the whole US was set. When users turn on location service functions in their smartphones, the smartphones allow Twitter to exactly locate the users' locations based on the global position system, Wi-Fi, or triangulated locations from cellphone towers. Geo-tagged tweets contain the coordinate locations of where users were when they posted their messages. It is known that only about 5% of tweets are geo-tagged. To discover knowledge based on the analysis of the tweets, the sample size of tweets should be large enough to be representative of the entire city population. Therefore, to collect sufficient data in terms of volume and scale, we collected tweets for the whole US for about three months from 16 November 2015 to 17 January 2016.

Once the collection of geo-tagged tweets across the US was completed, the noisy data were removed. It is well known that tweets are created not only by humans but also by robots. Tweets created by non-human users for advertising, web gaming, and job posting can produce biased analyses. This step aims to remove those noisy tweets in order to have tweets created only by humans based on two criteria. The first criterion was used in a previous study (Hawelka et al. 2014); all tweets from the authors who moved faster than 1000 km/h were removed because the maximum speed of an airplane is 1000 km/h. The second criterion to remove noisy data assumes that humans cannot tweet more than 12 times within 1 minute. Thus, all tweets created by authors who tweeted more than 12 times within 1 minute were removed. Based on the two criteria, about 11% of the tweets were considered noise. Once the noisy data were removed, the home cities of Twitter users were identified. The statistical result shows that from 16 November 2015 to 17 January 2016, Twitter users created on average 90% of their tweets only in one city with the rest in the other cities. In addition, during the same period, Twitter users who had tweeted at least once in New York and once in cities other than New York created 76.8% of their tweets within one of the US cities and the rest in other cities. Based on this, the city where the users tweet most over several months is most likely to be their home city. Therefore, the city where each user created the most tweets during the data collection period was considered his/her home city. In rare cases, the home city was not able to be identified because users created an equal number of tweets in each different city, but tweets created by those authors were less than 1% of all tweets and were not used in this study. Figure 1 shows each city boundary that was used to identify the home cities of each Twitter user.

4.2. Step 2: Data collection by Rest APIs

Twitter's Rest APIs were used to identify followings and followers of Twitter users. How the users are connected by following someone on Twitter cannot be identified from the dataset collected by Streaming API at the first step. However, Rest API allows us to query the user ids of followings and followers of a particular user. By using the Rest APIs, we queried followings and followers of users living in each of the four cities. Since every user's home city has been identified in Step 1.3, the home city of each following and follower has also been identified. Step 2.5 is to select one of the four biggest US cities, New York, Los Angeles, Chicago, and Houston, to examine the spatial interactions among Twitter users living in these cities and those living outside them. Steps 3 and 4 are the analyses for New York-based tweets. The same analyses were then performed for the remainder of the three cities, Los Angeles, Chicago, and Houston.

4.3. Step 3: Data extraction, spatial analysis, and visualization

Step 3 was to retrieve particular tweets from those processed in the previous steps. To examine the effect of distance decay on spatial interactions in cyberspace, we used two different datasets. The first dataset was tweets containing the locations of Twitter users that are connected with the users living in the four biggest cities who were following and being followed by someone (Step 3.1 in Figure 2).

Another dataset that we used to examine the effect of distance decay in cyberspace was geo-tagged tweets containing the four biggest US city names (i.e. New York, Los Angeles, Chicago, Houston). Therefore, Step 3.2 in Figure 2 focuses on filtering out tweets that contain those city names. The act of mentioning city names in their messages can be an indicator of spatial interactions between the city where the user lives and the city that the user mentions. For example, if users living in Boston created a large volume of messages mentioning the city name New York, this implies that there is potentially active interaction between Boston and New York in cyberspace. Furthermore, it implies that there is potentially active movement of people, ideas, goods, service, or labor between Boston and New York in real space. This example shows that the act of mentioning a city name in cyberspace can imply spatial interaction between the two cities.

To examine the effect of distance decay in real space in Step 3.3 of Figure 2, trajectories of Twitter users who traveled among greater than or equal to two cities were examined. Based on the trajectory of each user, we identified the tweeting locations of users during their travel outside of each of the four cities where they are living, and tweeting locations of Twitter users who had ever been to each of the four cities. The users' home cities were already identified in Step 2. Since the city in which each user is living had been identified, it was also possible to identify the city to which each user traveled. For example, if a user whose home city is Boston tweeted in New York, the tweet created by this user in New York means that this person actually traveled to New York. The method we use to identify home city and the city where users traveled is the same method that was used in a previous study for the same purpose (Hawelka et al. 2014).

4.4. Step 4: Spatial analysis and visualization

Step 4 focuses on creating the distance decay function based on each dataset prepared in Steps 3.1, 3.2, and 3.3 in Figure 2, and conducted the spatial analysis. To examine how each dataset spatially spread or concentrated, the radius of gyration was calculated based on the equation given below. The radius of gyration was adopted from a previous study and modified (Hawelka et al. 2014).

$$\text{Radius of gyration} = \frac{1}{n}\sum_{i=0}^{n} Di \times Xi/p, \tag{1}$$

where Di is the distance from one of the four cities (i.e. NY, LA, Chicago, and Houston) to tweeting locations outside of each city. Table 1 takes New York as an example. Column D in Table 1 shows what the distance (Di) refers to depending on the different spatial interactions in both cyberspace and real space. Great-circle distances were used to calculate the distance. Xi refers to the frequency of tweets in each city (i.e. each metropolitan area in Figure 1). Column E in Table 1 represents what the frequency represents for each spatial interaction. p is the total number of users or the total number of tweets in each metropolitan area. Column F of Table 1 shows p for each different spatial interaction. Dividing Xi by p provides a fair comparison between the areas with high population and those with low population. Without the normalization, Xi is always high in the big cities just because of their population size. The normalization is needed to remove the effect of an uneven distribution of population. In terms of normalization at Step 4.1 (Table 1 and Figure 2), each number of followings and followers in each region is divided by the total number of Twitter authors and multiplied by an adjustment factor (the first row of Column F). The adjustment factor is the average number of common followings or followers of users living in New York. For example, Mary living in LA is a friend of both John and Tom in NY. In this case, users living in NY (John and Tom) have the same friend (Mary) in LA. In this case, Mary should be counted as two people. In other words, the adjustment factor is for adjusting the total number of users for the purpose of unbiased normalization.

The power law was used to model each distance decay curve representing the spatial interaction in cyberspace (Steps 4.1 and 4.2 in Figure 2) and in real space (Step 4.3 in Figure 2). The power law is

Table 1. Variables needed to model the spatial interaction between New York (NY) and other US cities in real space and cyberspace.

(A) Space	(B) Interaction name	(C) Sub category	(D) Distance $P(\Delta\Gamma)$	(E) Frequency	(F) Normalized by
Analysis of Cyberspace	Follow (Step 4.1) :	followings	The distance between NY and the locations of users who are followed by users in NY	The frequency of users who are living in the cities outside NY and are followed by users living in NY	Total Twitter authors × adjustment factor
		followers	The distance between NY and the locations of users who follow users in NY	The frequency of users who are living in the cities outside NY and follow users in NY	
		friends	The distance between NY and the locations of friends of users in NY	The frequency of users who are living in the cities outside NY and have friends in NY	
	Awareness (Step 4.2):		The distance between NY and tweets that are created outside NY and contain the city name, 'New York'	The frequency of tweets that are created outside of NY and contain the city name, 'New York.'	Total tweets
Analysis of Real Space	Travel (Step 4.3):	residents' travel	The distance between NY and a destination city other than NY for users who traveled from NY	Frequency of travel between NY and cities other than NY by residents of NY	Total number of users
		visitors' travel	The distance between NY and tweeting location outside of NY for users who have ever been to NY	Frequency of travel between NY and cities other than NY by visitors to NY	

Note: Each analysis was conducted for Los Angeles, Chicago, and Houston in the same way. The frequency of follow (Step 4.1) in column B is the average frequency of followings and followers (see columns C and E). The frequency of travel (Step 4.3) in column B is the average frequency of residents' travel and visitor's travel (see columns C and E).

given as follows:

$$P(\Delta\Gamma) = \alpha\Delta\Gamma^{-\beta}, \tag{2}$$

where $\Delta\Gamma$is the distance between two places, α is a constant, and β is the power law exponent, which decides the shape of the distance decay curve. The bigger β, the steeper the slope of the distance decay curve. The goodness of fit for the power law was measured by R^2. For example, in terms of the use of a power law to model spatial interaction between New York (NY) and cities other than NY at Step 4, Table 1 shows values of $\Delta\Gamma$ and $P(\Delta\Gamma)$ at each step. $\Delta\Gamma$ is the distance in column D. $P(\Delta\Gamma)$ is the normalized frequency; each value in column E is divided by the value in column F.

In addition, the distributional difference in the effect of distance decay on spatial interaction between cyberspace and real space was examined through kernel density estimation.

Maps of kernel density estimation were created to visualize each of the datasets identified in Steps 3.1, 3.2, and 3.3. Through this process, the distributional patterns of spatial interactions between cyberspace and real space were compared, and the distributional differences between them were revealed.

5. Findings

5.1. Results of data collection

The tweets that were collected using the Twitter Streaming API are all geo-tagged. It is known that geo-tagged tweets are about 5% of the entire body of tweets. However, the size of the dataset that was collected was big enough to show general trends. A total of 51,019,087 tweets were collected in the US from 16 November 2015 to 17 January 2016. These tweets were created by 1,764,293 users in the US.

Table 2. The number of tweets or users during the data collection period.

	Type	New York	Los Angeles	Chicago	Houston
Interaction in cyberspace	(A) Followers	1,917,849	1,445,176	668,277	392,937
	(B) Followings	2,402,599	1,670,319	1,045,119	590,841
	(C) Friends	940,733	675,827	390,379	248,481
	(D) Awareness	158,239	27,708	71,391	61,848
Interaction in real space	(E) Residents' travel	24,912	18,100	9,837	1,933
	(F) Visitors' travel	41,669	39,027	21,961	12,324
N/A	(G) Total users	144,982	104,277	57,724	35,604
N/A	(H) Total tweets	3,700,607	2,957,339	1,565,104	1,292,300

Note: Each row from (A) to (F) corresponds to each row of column C of Table1. See column E of Table 1 for the specific description. In this study, followings and followers who created only geo-tagged tweets were used. The number of followings and followers including non-geo-tagged tweets is on average 19 times greater in each city than those in this table.

From these sampled data, we identified the number of Twitter users living in each city and the number of tweets created in each city in Rows G and H respectively in Table 2. Rows A–F in Table 2 show the frequencies that were used for representing spatial interactions in cyberspace and real space during the data collection period. Row C in Table 2 represents the frequency of a both-side relationship between users – that is, user A follows user B, and user B also follows user A. Unlike friends, followers and followings do not guarantee the both-side relationship; user A follows user B, but user B does not necessarily follow user A.

5.2. Measuring the level of spread of different types of spatial interactions

In terms of the level of spread, the distribution of data representing spatial interaction in cyberspace was more spatially dispersed than the distribution of data representing spatial interaction in real space. Figure 3 shows the result of computing the radius of gyration as a percentage. The higher the number, the more spread in each dataset. There is not much difference in the level of spread between followers and followings of users living in Chicago and Houston. However, there is a big difference in the level of spread between followers and followings of users living in New York and Los Angeles. The reason for this big difference is possibly the celebrities living in those two cities. A huge number of users are followers of celebrities, while celebrities are not usually following

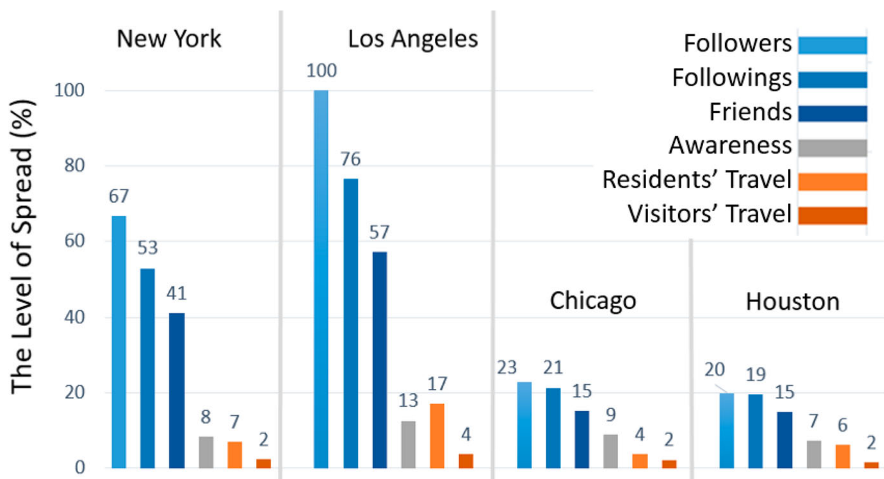

Figure 3. Radius of gyration computed based on Equation (1). The description of each variable is in columns C, D, and E of Table 1. Followers, followings, friends, and awareness represent spatial interaction in cyberspace. Residents' and visitors' travel represent spatial interaction in real space. Awareness of New York means that a user mentioned New York in the tweets.

their fans. In fact, according to Schwartz (2014), the city that has the largest number of celebrities is Los Angeles, and the city that has the second largest celebrities is New York. Two users are friends only if they are following each other. Friends of users are always more spatially concentrated around each of the four cities than followings or followers of users in each city. The reason is that many users tend to follow Twitter accounts of famous people and big organizations, but those accounts do not follow most of their followers. The distribution of cities visited by residents in each of the four cities is more likely to be spread out than the distribution of cities visited by visitors to the same cities. In terms of awareness, tweets containing each city name tend to be more spatially concentrated than followings, followers, and friends, and more spread out than residents' and visitors' travel except for Los Angeles (LA). This exception is because in cities that are far away from LA, users mentioned Los Angeles many times. The exceptional pattern for LA was also observed in Figures 5(B), 7, and 8 (E).

To see the change of frequency of spatial interaction with increasing distance from each of the biggest four cities, the mainland of the US was divided into buffer zones in 50 km increments from the boundaries of the four cities. Maps representing every 50 km buffer zone from each of four cities are shown in Figure S1, computed using great-circle distances and shown projected. To examine the distance decay effect, the frequencies of each interaction (i.e. column E of Table 1) within each buffer were counted and normalized by each value in column F of Table 1.

The normalization reveals the effect of distance decay. For example, Figure 4(A) shows the frequency of followings of users in Los Angeles (LA) with the increase of distance. Figure 4(B) shows the frequency that was normalized according to column F in Table 1. In each graph, the spot at 0 km is LA, the spot at approximately 450 km is San Francisco, and the spot at approximately 3850 km is New York. In Figure 4(A), the highest peak is New York and the second highest peak is San Francisco. In Figure 4(A), the frequency of followings of users in LA is high in the area where the

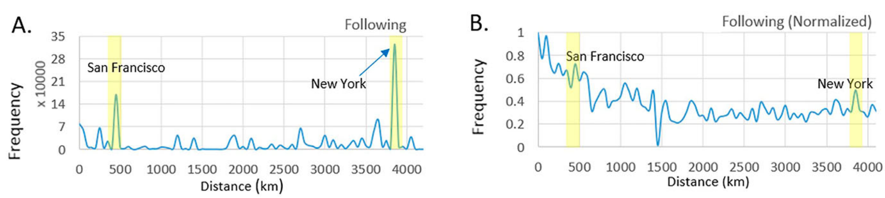

Figure 4. Graph A shows the frequency of Twitter users living in the cities outside Los Angeles (LA) who are followed by users in LA. Graph B is after normalization: frequency/(total Twitter authors × adjustment factor). See column F of Table 1.

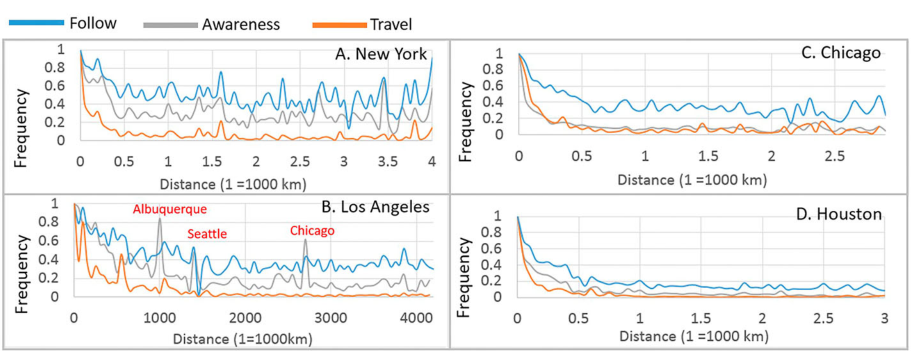

Figure 5. The frequency of spatial interactions in cyberspace and in real space. The blue line represents the distance decay curve of spatial interactions by following. The gray line represents the distance decay curve of spatial interactions by containing each of the four city names. The orange line represents the distance decay curve of spatial interactions by travel (Color online).

population is high. In Figure 4(B), New York has a relatively high frequency of followings among eastern cities, but it is not the biggest peak. Similarly, San Francisco in Figure 4(B) is relatively high among surrounding areas, but it is not as high as in Figure 4(A). Another characteristic of Figure 4(B) is that it shows the effect of distance decay, which was not observed in Figure 4(A). Figures S2–S5 show a change of interaction frequency in cyberspace with the increase of distance from each of the four cities. Figures S6–S9 show a change of interaction frequency in real space with the increase of distance from each of the four cities.

The slope of the distance decay curve is generally steeper in real space than in cyberspace in each of the four cities. In Figure 5, follow, awareness, and travel are described in column B of Table 1. Each frequency of follow, awareness, and travel was normalized according to column F in Table 1. In addition, for a fair comparison of the three lines in Figure 5, each of the frequencies was normalized by dividing by the maximum value, which makes the maximum value of each line become 1. Blue and gray lines represent the distance decay curve of spatial interaction in cyberspace. The orange line graph shows the distance decay curve of spatial interactions in real space. The distance decay effect of the spatial interaction in real space is stronger compared to the other spatial interactions in cyberspace. In terms of the spatial interaction in real space (orange lines in Figure 5), the effect of distance decay can be observed in all four cities; a dramatic decrease of the spatial interaction can be observed until 500 km, which is about 5–6 hours driving time. Users rarely have interactions beyond 500 km in physical space. In terms of the interactions by followings and followers of someone (blue lines in Figure 5) in cyber space, the mild decrease of the interaction is observed until around 750–1000 km with the increase of distance for all four cities; however, beyond 1000 km, distance does not exert an impact on the interaction.

Interactions by mentioning city names (gray line in Figure 5) have a steeper distance decay curve than the interactions by following and followers of someone, but have a less steep distance decay curve than the interactions by traveling in real space. However, the gray line in Figure 5(B) shows an exceptional pattern which has unusually high peaks at 1000, 1400, and 2700 km where the major cities, Albuquerque in New Mexico, Seattle in Washington state, and Chicago in Illinois, are respectively located. In other words, these three unusual high peaks mean that Twitter users living in each city mentioned ‘Los Angeles’ much more than others at a similar distance. To identify the reasons for the peaks, the contents of the Twitter messages in each city were examined. The reason of the high peak in Albuquerque was that 74% of tweets were created by only two users. By examining the contents of the tweets, we identified that one user formerly lived in Los Angeles and the other user’s hometown is Los Angeles. In terms of the high peaks in Seattle and Chicago, each word cloud in Figure 6 describes the reasons. The words such as ‘Seattle’, ‘WA’, ‘Galaxy’, ‘CenturyLink’, ‘Field’, ‘VS’, ‘Sounders’, and ‘FC’ are outstanding among the words consisting of Twitter messages created by users living in Seattle (the left image of Figure 6). The reason is that most of the tweets were about a soccer game, Los Angeles Galaxy vs. Seattle Sounders FC at CenturyLink Field in

Figure 6. Word Cloud. The left image was created with 60 Twitter messages from users in Seattle. The right image was created with 174 Twitter messages from users in Chicago. The bigger the words, the more people mentioned the word. D3.js, a JavaScript library, was used to create these word clouds.

Seattle, WA. Among the words consisting of Twitter messages created by users living in Chicago, the words such as 'DODGERS', 'Acquired', 'Infielder', 'outfielder', 'Contract'. 'Signed', Pitcher, 'MLB', 'NHL', 'Books', and 'Review' were outstanding (the right image of Figure 6). The reason is that most of the topics related to Los Angeles in Twitter messages created in Chicago were related to sports. In addition, many users in Chicago also shared multiple articles from the *Los Angeles Review of Books* (a literary review journal).

5.3. The power law model of the distance decay curve

Each frequency for follow, awareness, and travel (column B of Table 1) was modeled based on the power law. Previous studies have found that distributions of human movements follow a power law (Hawelka et al. 2014; Jurdak et al. 2015). The power law is described in Equation (2). Each follow, awareness, and travel in Figure 7 is the result of fitting a power law to each of the lines in Figure 5. The results of power law model fitting such as R^2, β, and residual sum-of-squares (RSS) help to understand the characteristics of each distance decay curve in Figure 5. The goodness fit of a power law was measured based on R^2 in Figure 7. The higher the R^2, the better the distribution is approximated by the power law. R^2 closer to one means a better power law fit. In the case of interactions by traveling, R^2 ranges between 0.64 and 0.88. In the case of interactions by following on Twitter, R^2 ranges between 0.44 and 0.78. In addition, in each of four cities, interaction by traveling always shows higher β than the interaction by following on Twitter. Therefore, the distribution of interactions by traveling generally had a better power law fit than the distribution of interactions by following in the four cities. In the case of Houston, R^2 of follow was 0.78, so it follows the power law distribution. However, in the case of New York, R^2 of follow was 0.43, so it is hard to

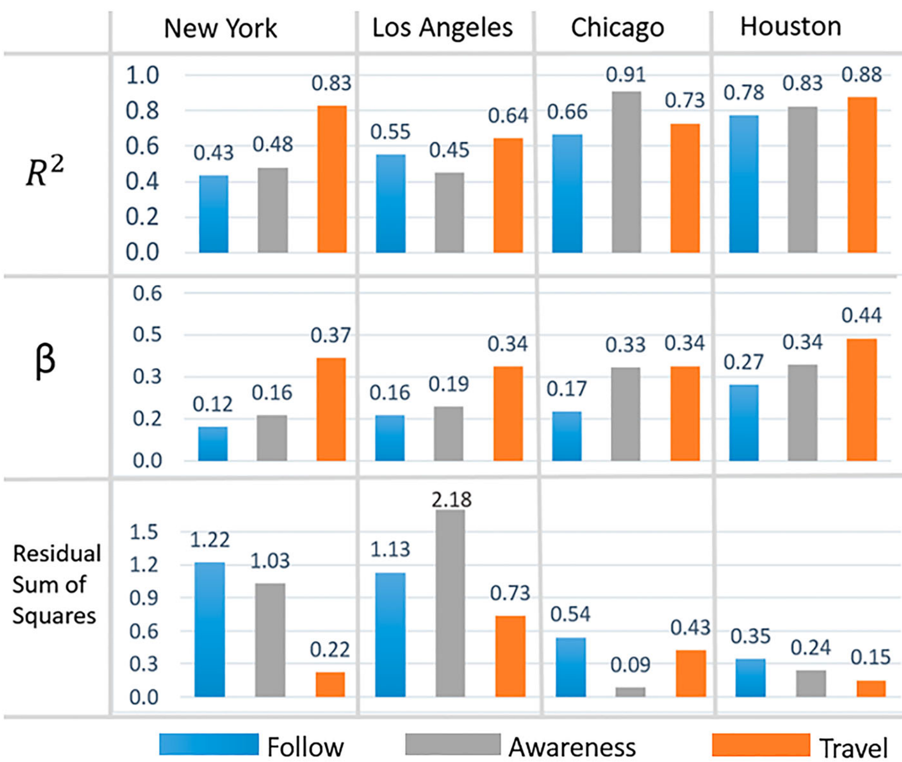

Figure 7. The result of fitting a power law. Follow and awareness represent spatial interaction in cyberspace. Travel represents spatial interaction in real space.

tell that it actually follows the power law distribution. On the other hand, in the case of Houston, all three different types of interactions were fit well by the power law – that is, R^2 of each follow, awareness, and travel was 0.78, 0.83, and 0.88, respectively.

The power law exponent, β reflects the shape of the distance decay curve – that is, the bigger the β, the steeper the slope of the distance decay curve. In the case of interactions by traveling, β ranges between 0.34 and 0.44. In the case of interaction by following on Twitter, β ranges between 0.12 and 0.27. In addition, in each of the four cities, the interaction by traveling always showed a higher β than the interaction by following on Twitter. These findings show that distribution of interactions by traveling have a steeper slope (higher friction of distance) than the distribution of interactions by following in each of the four cities.

RSS measures the overall difference between the actual data and the values predicted by a power law model. The closer RSS is to zero, the better the model fits the data. In other words, the bigger RSS, the more the distance decay curve differs from the power law. In each of four cities, the interaction by traveling always shows lower RSS than the interaction by following on Twitter (Figure 7) – that is, the distance decay curve of interactions by followings differs more than the distance decay curve of interactions by traveling. The unusually high RSS of awareness of Los Angeles mirrors the peaks at 1000, 1400, and 2700 km in Figure 5(B).

5.4. Difference in the geographic pattern of spatial interactions between cyberspace and real space

The above findings are also supported by the visualization of the effect of distance decay on spatial interaction in cyberspace and in real space (Figure 8). Kernel density estimation was used to visualize the distribution of each interaction in each of the four cities (Figure 8). Maps of follow, awareness, and travel visualize each frequency (of column B of Table 1) normalized by column F of Table 1 in each city. Maps of follow in Figure 8(A,D,G,J) show the spatial distribution of frequency of users by following. Maps of awareness in Figure 8(B,E,H,K) show the spatial distribution of tweets mentioning each city name. Maps of travel in Figure 8(C,F,I,L) show the spatial distribution of travel frequency. The darker the color, the higher the frequency. The different extent of the dark-colored region represents the difference in the effect of distance decay between cyberspace and real space. For example, in terms of a comparison of Figure 8(A) with Figure 8(C), in cyberspace there is a considerable number of users who had connections in New York and also reside in western US cities such as Seattle, San Francisco, and Los Angeles (Figure 8(A)). However, in real space there are few users who had interactions with New York in real space and also reside in western US cities (Figure 8(C)). Most interactions by travel to or from New York (Figure 8(C)) are clustered in the northeastern US where New York City is located.

The right four maps of Figure 8(C,F,I,L) represent the distribution of spatial interactions in real space and the left four maps of Figure 8(A,D,G,J) represent the distribution of spatial interactions in cyberspace. These maps reveal an obvious contrast in spatial interactions in real space versus in cyberspace. The distributions of spatial interactions in real space in each of the four cities (Figure 8(C,F,I,L)) show that human interactions are clustered around each of the four cities. In contrast, Figure 8(A,D,G,J) shows similar distributions of spatial interactions in cyberspace; these four similar maps imply that people are well connected in cyberspace through the Twitter network. Even though these four maps look similar (Figure 8(A,D,G,J)) in terms of the spread of spatial interactions, each of maps shows that the frequency of spatial interactions is highest around each of the four cities. In summary, the evidence in Figure 8(A,D,G,J) indicates that there is a distance decay effect in cyberspace, but the effect is weak. The weak distance decay effect that is visualized on each of map of cyberspace (Figure 8(A,D,G,J)) corresponds to the distance decay curve of each city in Figure 5 (the blue line in each graph). There is a mild decrease of the frequency of spatial interaction until around 750–1000 km and beyond 1000 km, there is no more distance decay.

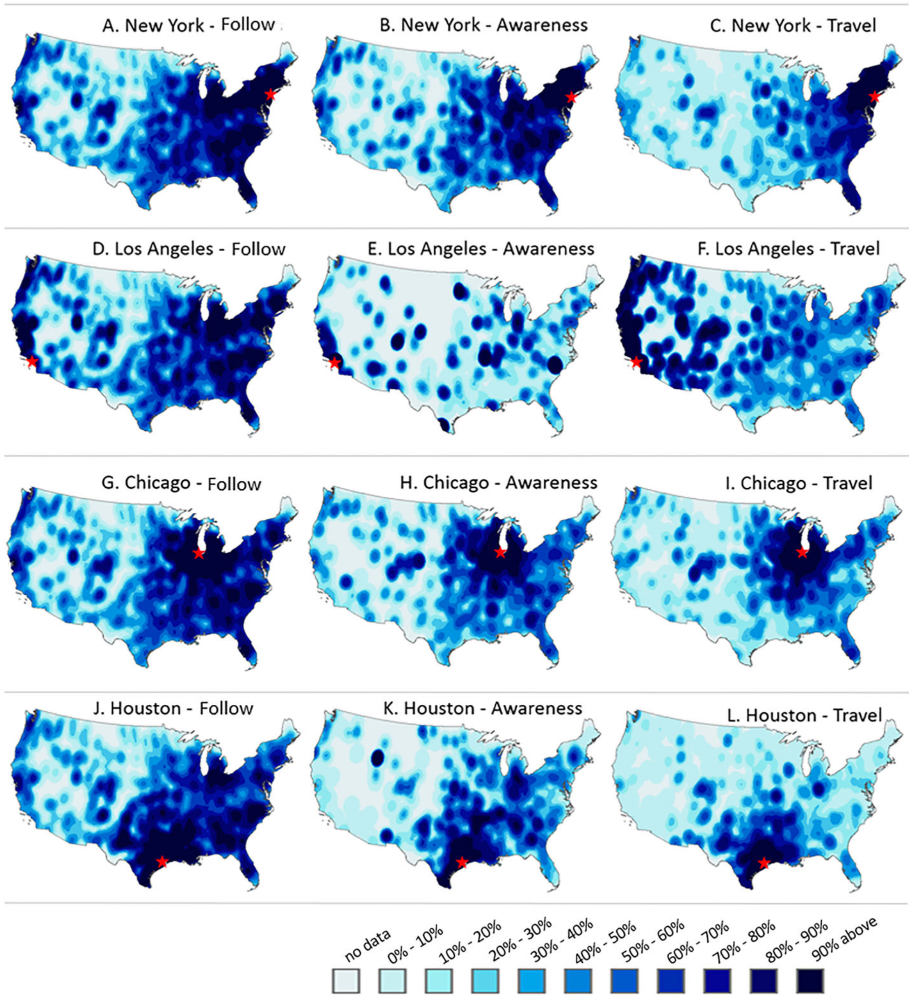

Figure 8. Maps of the effect of distance in spatial interaction in cyberspace and real space. The darker the color, the higher the frequency of spatial interaction. The quantile classification method was used to draw the maps. In the case of maps A–C, the area with 90% above (the darkest color) represents that the frequency of interactions with New York belongs in the top 10% among the entire frequency. The same rule applied to the rest of the classes.

Maps of awareness in Figure 8(B,H,K) show the medium extent between maps of follow (Figure 8 (A,G,J)) and maps of travel (Figure 8(C,I,L)). Figure 8(E) has an exceptional pattern compared to the maps of awareness for other cities (Figure 8(B,H,K)).

6. Discussion

A contribution of this study is to show that quantifying the extent of areas where people are frequently interacting can help to measure how far interactions can spread. The interactions that can spread over space include not only tangible things such as people and goods, but also intangible things such as information and diseases. Each map of Figure 9(A,B) is another visualization of the same data that are used to visualize the respective maps in Figure 8(A,C). Figure 9(A,B) only visualizes the areas above 70% to emphasize the contrast between the two maps.

The reddish areas with 70% and above mean that people living in the areas are 70% or more likely to interact people living in New York. The more vivid the red color, the higher the probability that

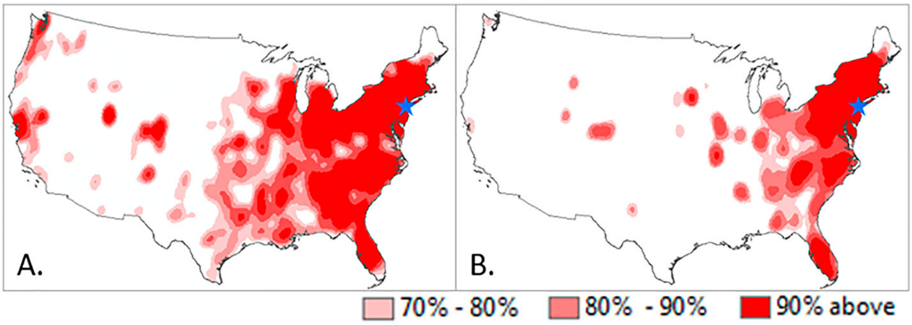

Figure 9. The areas where people are 70% likely to interact with people in New York through cyberspace A and real space B.

people interact. Figure 9(A) represents the extent of the areas where information can be diffused online. On the other hand, Figure 9(B) is the extent of the areas is where interactions spread through face-to-face contact. When the two maps, Figure 9(A,B), are compared to each other, the reddish colored area is much wider in Figure 9(A) than in Figure 9(B). These two different sizes and shapes of areas can be used to improve the well-being of individuals in society by many private and public organizations. For example, if an epidemic such as influenza started in New York, the reddish area in Figure 9(B) is the area where the epidemic is highly likely to spread through travel to and from New York. The extent of the area in Figure 9(B) could be used by public health officials to predict the extent of areas where diseases will spread. On the other hand, if public health officials in New York post information about the prevention of epidemics on a social networking website, the reddish region in Figure 9(A) is the areas where the information can be easily diffused through users that are connected with users in New York through the social networking website. In addition, the boundary in Figure 9(A) can be used in political campaigns, if a user posted some information to support a particular candidate during a national election campaign, the information will be easily diffused through human interactions from New York to the reddish regions in Figure 9(A). In these example cases, the areas with 70% and above in Figure 9(A,B) were selected from the maps in Figure 8(A,C). However, different percentages should be chosen depending on the purpose of the maps.

7. Conclusion

The use of Twitter data enabled the quantification of the interactions between users in cyberspace and those in real space. In terms of the level of spread, the distribution of interactions in cyberspace is more spatially spread out than the distribution of interactions in real space. The distance decay effect on spatial interaction was found to apply in both cyberspace and real space. However, the slope of the distance decay curve representing interactions is steeper in real space than in cyberspace. The distance decay curve of cyberspace fluctuates more than the distance decay curve of real space. Kernel density maps revealed that users are much more interconnected in cyberspace with users living in every part of US than in real space.

This study provides empirical evidence to inform the debate over the 'death of distance' by quantifying the influence of distance on interactions between cyberspace and real space. Based on the empirical evidence found in this study, we suggest a new corollary to Tobler's first law of geography; 'In both real space and cyberspace, everything is related to everything else, but near things are more related in real space than in cyberspace.' Measuring the rate of distance decay in cyberspace in this study can help predict the degree of information flow across space through social media. Measuring how far ideas can be diffused through social media is useful for users of location-based services, policy advocates, public health officials, and political campaigners.

8. Limitations

When the spatial interactions in cyberspace and real space were examined, this study could not examine every spatial interaction happening in the two spaces. To understand the patterns of spatial interaction in cyberspace and in real space, the methodological framework of this study is based on inductive reasoning. Thus, after the observations of spatial patterns of human interactions of millions of Twitter users in each cyberspace and real space, we have generalized the distance decay curve in cyberspace and real space. However, the use of inductive reasoning to generalize the patterns of distance decay curve in cyberspace and real space has limitations. In real space, counting the number of Twitter users' travels using the geo-tagged tweets represents only a small proportion of human travel in real space. Previous studies have demonstrated that Twitter can be a proxy for human mobility patterns. Thus, we assume that Twitter data can reveal the patterns of movement. However, there is a possibility of overgeneralization of human travel patterns. Also, in cyberspace, Twitter is only one of many channels where people interact. Spatial interactions counted based on relationships like following do not represent the entire set of spatial interactions happening in cyberspace. In fact, interactions in cyberspace happen not only on Twitter, but also at other social networking sites such as Facebook and Instagram, by online chatting, and by email. There are numerous other channels where people can interact in cyberspace. Even though this study provides strong evidence for distance decay in cyberspace based on the observations of interactions in the Twitter network, it is possible that interactions by following in Twitter might not equate to the entire set of interactions happening in real space.

Acknowledgements

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This material is based upon work supported by the National Science Foundation [grant number 1416509], IBSS project titled 'Spatiotemporal Modeling of Human Dynamics Across Social Media and Social Networks' and National Science Foundation [grant number 1634641], IMEE project titled 'Integrated Stage-Based Evacuation with Social Perception Analysis and Dynamic Population Estimation'.

ORCID

Ming-Hsiang Tsou http://orcid.org/0000-0003-3421-486X
Keith C. Clarke http://orcid.org/0000-0001-5805-6056

References

Aslam, Anoshé A., Ming-Hsiang Tsou, Brian H. Spitzberg, Li An, J. Mark Gawron, Dipak K. Gupta, K. Michael Peddecord, Anna C. Nagel, Christopher Allen, and Jiue-An Yang. 2014. "The Reliability of Tweets as a Supplementary Method of Seasonal Influenza Surveillance." *Journal of Medical Internet Research* 16 (11): e250.

Bates, Stephen. 1996. "The Death of Geography, the Rise of Anonymity, and the Internet." *American Enterprises* 7 (2): 50–52.

Cairncross, Frances. 2001. *The Death of Distance: How the Communications Revolution Is Changing Our Lives*. Boston: Harvard Business Press.

Couclelis, Helen. 2009. "Rethinking Time Geography in the Information Age." *Environment and Planning A* 41 (7): 1556–1575.

Devriendt, Lomme, Andrew Boulton, Stanley Brunn, Ben Derudder, and Frank Witlox. 2011. "Searching for Cyberspace: The Position of Major Cities in the Information Age." *Journal of Urban Technology* 18 (1): 73–92.
Fellmann, Jerome D., Arthur Getis and Judith Getis. 2007. *Human Geography: Landscapes of Human Activities*. 10th ed. New York: McGraw-Hill Higher Education.
Fiore, Quentin, and Marshall McLuhan. 1967. *The Medium Is the Massage*. New York: Random House.
Fotheringham, A Stewart, and Morton E. O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications*. Boston: Kluwer Academic.
Gibson, W. 1984. *Neuromancer*. New York: Berkley Pub.
Goldenberg, Jacob, and Moshe Levy. 2009. "Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era." arXiv preprint arXiv:0906.3202.
Graham, Mark. 2013. "Geography/Internet: Ethereal Alternate Dimensions of Cyberspace or Grounded Augmented Realities?" *The Geographical Journal* 179 (2): 177–182.
Green, Carolyn, and Karen Ruhleder. 1995. "Globalization, Borderless Worlds, and the Tower of Babel: Metaphors Gone Awry." *Journal of Organizational Change Management* 8 (4): 55–68.
Han, Su Yeon, Ming-Hsiang Tsou, and Keith C. Clarke. 2015. "Do Global Cities Enable Global Views? Using Twitter to Quantify the Level of Geographical Awareness of US Cities." *PloS One* 10 (7): e0132464.
Hawelka, Bartosz, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. "Geo-located Twitter as Proxy for Global Mobility Patterns." *Cartography and Geographic Information Science* 41 (3): 260–271.
Jurdak, Raja, Kun Zhao, Jiajun Liu, Maurice AbouJaoude, Mark Cameron, David Newth, and Ye Wu. 2015. "Understanding Human Mobility from Twitter." *PloS One* 10 (7): e0131469.
Kamvar, Sepandar D., and Jonathan Harris. 2011. "We Feel Fine and Searching the Emotional Web." Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, Hong Kong, Feb 9–12.
Kellerman, Aharon. 2010. "Mobile Broadband Services and the Availability of Instant Access to Cyberspace." *Environment and Planning A* 42 (12): 2990–3005.
Kellerman, Aharon. 2016. *Geographic Interpretations of the Internet*. Switzerland: Springer.
Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts." *Big Data & Society* 1 (1): 2053951714528481.
Kolko, Jed. 2000. "The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage." In *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*, edited by Ingo Vogelsang and Benjamin M. Compaine, 73–98. Cambridge, MA: The MIT Press.
Kwon, Ohbyung, and Jae Mun Sim. 2013. "Effects of Data Set Features on the Performances of Classification Algorithms." *Expert Systems with Applications* 40 (5): 1847–1857.
Laboratory, Oak Ridge National. 2016. "Geographic Information Science and Technology." Retrieved February 24, 2016, from http://web.ornl.gov/sci/landscan/.
Laney, D. 2001. "3-D Data Management: Controlling Data Volume, Variety and Velocity." *META Group File* 949: 1–4.
Lerman, Kristina, and Rumi Ghosh. 2010. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." *ICWSM* 10: 90–97.
Leskovec, Jure, and Eric Horvitz. 2008. "Planetary-scale Views on a Large Instant-Messaging Network." Proceedings of the 17th International Conference on World Wide Web, Beijing, China, April 15–28.
Leszczynski, Agnieszka. 2015. "Spatial Media/Tion." *Progress in Human Geography* 39 (6): 729–751.
Liu, Jiajun, Kun Zhao, Saeed Khan, Mark Cameron, and Raja Jurdak. 2015. "Multi-scale Population and Mobility Estimation with Geo-tagged Tweets." Data Engineering Workshops (ICDEW), 2015, 31st IEEE International Conference, Seoul, South Korea, April 13–17.
MacEachren, Alan M., Anthony C. Robinson, Anuj Jaiswal, Scott Pezanowski, Alexander Savelyev, Justine Blanford, and Prasenjit Mitra. 2011. "Geo-Twitter Analytics: Applications in Crisis Management." 25th International Cartographic Conference, France, July 3–8.
MacLachlan, Ian. 2011. *Quantitative Models for Geographical Analysis Additional Required Reading for Geography 3235*. Fourth Revision. Retrieved June 25, 2016, from http://classes.uleth.ca/201101/geog3235a/Quantitative_Models_coursepack_March2011.pdf.
McAfee, Andrew, Erik Brynjolfsson, Thomas H. Davenport, D. J. Patil, and Dominic Barton. 2012. "Big Data. The Management Revolution." *Harvard Business Review* 90 (10): 61–67.
McLuhan, Marshall, and Bruce R. Powers. 1989. *The Global Village: der Weg der Mediengesellschaft in das 21. Jahrhundert*. Vol. 21. Paderborn: Junfermann.
Mok, Diana, and Barry Wellman. 2007. "Did Distance Matter Before the Internet? Interpersonal Contact and Support in the 1970s." *Social Networks* 29 (3): 430–461.
Mok, Diana, Barry Wellman, and Juan Carrasco. 2010. "Does Distance Matter in the Age of the Internet?" *Urban Studies* 47 (13): 2747–2783.
Morgan, Kevin. 2004. "The Exaggerated Death of Geography: Learning, Proximity and Territorial Innovation Systems." *Journal of Economic Geography* 4 (1): 3–21.
Nagel, Anna C., Ming-Hsiang Tsou, Brian H. Spitzberg, Li An, J. Mark Gawron, Dipak K. Gupta, Jiue-An Yang, Su Han, K. Michael Peddecord, and Suzanne Lindsay. 2013. "The Complex Relationship of Realspace Events and

Messages in Cyberspace: Case Study of Influenza and Pertussis Using Tweets." *Journal of Medical Internet Research* 15 (10): e237.
Noulas, Anastasios, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2012. "A Tale of Many Cities: Universal Patterns in Human Urban Mobility." *PloS One* 7 (5): e37027.
O'brien, Richard. 1992. *Global Financial Integration: The End of Geography*. New York: Council on Foreign Relations Press.
Ohmae, Kenichi. 1995. *The Borderless World: Power and Strategy in an Interdependent Economy*. New York: Harper Business.
Polèse, Mario, and Richard Shearmur. 2004. "Is Distance Really Dead? Comparing Industrial Location Patterns Over Time in Canada." *International Regional Science Review* 27 (4): 431–457.
Rainie, Lee, and Barry Wellman. 2012. *Networked: The New Social Operating System*. Cambridge, MA: MIT Press.
Sakaki, Takeshi, Makoto Okazaki, and Yutaka Matsuo. 2010. "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors." Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, April 26–30.
Schwartz, Harriette. 2014. "10 Cities for Spotting Celebrities." THERICHEST (accessed February 27).
Shook, Eric, Kalev Leetaru, Guofeng Cao, Anand Padmanabhan, and Shaowen Wang. 2012. "Happy or Not: Generating Topic-based Emotional Heatmaps for Culturomics Using CyberGIS." E-Science (e-Science), 2012 IEEE 8th International Conference, Chicago, IL, October 8–12.
Signorini, Alessio, Alberto Maria Segre, Philip M. Polgreen, and Alison P. Galvani. 2011. "The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. During the Influenza A H1N1 Pandemic." *PloS One* 6 (5): e19467.
Stephens, Monica, and Ate Poorthuis. 2015. "Follow Thy Neighbor: Connecting the Social and the Spatial Networks on Twitter." *Computers, Environment and Urban Systems* 53: 87–95.
Swanson, David A., and Jeff Tayman. 2012. *Subnational Population Estimates*. Vol. 31. Dordrecht: Springer Science & Business Media.
Takhteyev, Yuri, Anatoliy Gruzd, and Barry Wellman. 2012. "Geography of Twitter Networks." *Social Networks* 34 (1): 73–81.
Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46: 234–240.
Toffler, Alvin. 1980. *The Third Way*. New York: Bantam.
Tsou, Ming-Hsiang. 2015. "Research Challenges and Opportunities in Mapping Social Media and Big Data." *Cartography and Geographic Information Science* 42 (sup1): 70–74.
Tsou, Ming-Hsiang, Jiue-An Yang, Daniel Lusher, Su Han, Brian Spitzberg, Jean Mark Gawron, Dipak Gupta, and Li An. 2013. "Mapping Social Activities and Concepts with Social Media (Twitter) and web Search Engines (Yahoo and Bing): A Case Study in 2012 US Presidential Election." *Cartography and Geographic Information Science* 40 (4): 337–348.
Wang, Qi, and John E. Taylor. 2015. "Process Map for Urban-human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms." *Journal of Computing in Civil Engineering* 30 (2): 04015004.
Warf, Barney. 2013. *Global Geographies of the Internet*. Dordecht: Springer.
White, Tom. 2012. *Hadoop: The Definitive Guide*. Sebastopol, CA: O'Reilly Media.
Zook, Matthew A., and Mark Graham. 2007. "Mapping DigiPlace: Geocoded Internet Data and the Representation of Place." *Environment and Planning B: Planning and Design* 34 (3): 466–482.