Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

A spatial econometric modeling of online social interactions using microblogs

Zheye Wang[a], Xinyue Ye[a], Jay Lee[a], Xiaomeng Chang[b,c,*], Haimeng Liu[d], Qingquan Li[b,c]

[a] Department of Geography, Kent State University, Kent, OH, USA
[b] Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China
[c] Key Laboratory for Geo-environment Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and Geoinformation, Shenzhen University, Shenzhen 518060, China
[d] Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Keywords:
Social media
Death of distance
Spatial and social network
Spatial econometrics
Gravity model

ABSTRACT

With the advent of Information and Communication technology (ICT) in modern age, the statement of "death of distance" has received numerous discussions. This article contributes a new empirical study to the debate of "death of distance" by considering the effect of spatial autocorrelation in the estimation of distance decay effect with the incorporation of network autocorrelation in spatial econometric modeling. This work is based on a city-level dataset from China's largest social networking site called Weibo. The findings are shown as following. First, the coefficient value of network autocorrelation term (0.007, significant at 0.01 level) suggests that the city-level online social links are spatially dependent. In other words, these social connections are not randomly distributed across space but tend to form spatial clusters where neighboring links are more similar. Second, controlling spatial autocorrelation in the data, a distance decay effect on the formation of online social links is unveiled with a much smaller scaling exponent of the distances (i.e., 0.276) as compared to those (e.g., 2.0, 1.8, 1.45, 1.06, 1.03, 0.4, and 0.5) in existing studies. This research provides a useful modeling framework to analyze the real-world driving forces that characterize the patterns of social interactions in virtual space and thus advance our understanding in the connection of virtual and real spaces.

1. Introduction

The technological innovations in transportation and communication have brought about dramatic changes in human interactions. With automobiles, high-speed trains, and airplanes, people can travel much longer distances than ever before. Likewise, telecommunications enable users to carry out long-distance conversations in a virtually instantaneous manner (Shaw, Tsou, & Ye, 2016). In this context, it has been asserted that the frictional effects of geography on human mobility and social interactions are diminishing. In terms of social interactions, some claimed the "death of distance" thesis, stating that geographical distance becomes less constraining on the formation of social connections with the advent of Information and Communication technology (ICT) in modern age. Nonetheless, studies have showed evidence against the "death of distance" thesis and have criticized it as an exaggeration of some technological advances. More specifically, just because some technologies can let people expand their social connections everywhere it does not necessarily mean people will do that. Furthermore, some researchers such as Goldenberg and Levy (2009) point out that the role of geographical distance actually became stronger since people stick to their already existing social links, regardless of the influence of new technologies. The debate is still ongoing.

In recent years, empirical efforts attempting to contribute to this debate have utilized emerging data sources derived from location-based services (Andris, 2016; Han, Tsou, & Clarke, 2017). Online social networks (Backstrom, Sun, & Marlow, 2010; Crandall et al., 2010; Goldenberg & Levy, 2009; Grabowicz, Ramasco, Gonçalves, & Eguíluz, 2014; Kaltenbrunner et al., 2012; Lengyel, Varga, Ságvári, Jakobi, & Kertész, 2015; Liben-Nowell, Novak, Kumar, Raghavan, & Tomkins, 2005; Quercia, Capra, & Crowcroft, 2012; Scellato, Mascolo, Musolesi, & Latora, 2010; Takhteyev, Gruzd, & Wellman, 2012; Volkovich, Scellato, Laniado, Mascolo, & Kaltenbrunner, 2012) and mobile communication networks (Calabrese, Smoreda, Blondel, & Ratti, 2011; Expert, Evans, Blondel, & Lambiotte, 2011; Gao, Liu, Wang, & Ma, 2013; Kang, Zhang, Ma, & Liu, 2013; Krings, Calabrese, Ratti, &

* Corresponding author at: Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China.
E-mail addresses: zwang31@kent.edu (Z. Wang), xye5@kent.edu (X. Ye), jlee@kent.edu (J. Lee), changx@szu.edu.cn (X. Chang), haimengliu@163.com (H. Liu), liqq@szu.edu.cn (Q. Li).

https://doi.org/10.1016/j.compenvurbsys.2018.02.001
Received 5 September 2017; Received in revised form 15 January 2018; Accepted 3 February 2018

Blondel, 2009; Lambiotte et al., 2008; Onnela, Arbesman, González, Barabási, & Christakis, 2011; Shi, Wu, Chi, & Liu, 2016) are drawing particular attention from researchers. With these datasets, quantitative analyses such as measurement and modeling have been conducted to explore what the datasets imply.

From a measurement perspective, various metrics have been proposed to gauge the geographical length of social links (Scellato et al., 2010), geographical span of social communities (Onnela et al., 2011), geographical span of egonetworks (Quercia et al., 2012), distance disparity of triad edges (Grabowicz et al., 2014), distance dependence probability (Lambiotte et al., 2008; Lengyel et al., 2015), space-time co-occurrences of friends (Crandall et al., 2010), among others. From a modeling perspective, spatial interaction modeling has been widely used to test the role of geographical distance in empirical studies (Backstrom et al., 2010; Expert et al., 2011; Gao et al., 2013; Goldenberg & Levy, 2009; Kang et al., 2013; Krings et al., 2009; Liben-Nowell et al., 2005).

Gravity model has been recognized as an important tool in the investigation of various spatial interactions such as migrations, commodity flows, and so forth. As an aggregate spatial interaction model, the specification of a conventional gravity model assumes that the amount of spatial interactions is proportional to the size of the independent populations and inversely proportional to the distance between the locations (Fotheringham, 1981). The application of gravity model has revealed the effect of distance decay with different magnitudes on social flows in Cyberspace (e.g., Expert et al., 2011; Gao et al., 2013; Kang et al., 2013; Krings et al., 2009). Krings et al. (2009) shows that inter-city mobile phone communication in Belgium is subject to distance decay effect with a scale exponent close to 2. Kang et al. (2013) finds that mobile phone calls and their time durations are constrained by geographical distances with scaling exponents of 0.5 and 0.4, respectively. However, it remains unclear whether the distance decay estimation is free from model misspecifications. Are there any advances in spatial interaction modeling that we can adopt to mitigate potential biases in results?

Since 1970s, researchers have examined the impact of spatial structure on spatial interaction modeling (Tiefelsdorf, 2003). One manifestation of this spatial structure effect is the existence of spatial autocorrelation contained in origin and destination geographic distributions (Griffith, 2007). More specifically, as stated by Griffith and Jones (1980), flows from an origin are "enhanced or diminished in accordance with attributes displayed by neighboring origin locations", and similarly, flows received by a destination are "enhanced or diminished in accordance with attributes displayed by neighboring destination locations". As evidenced by existing studies (Griffith & Jones, 1980, Griffith, 2007, Chun, 2008, Chun, Kim, & Kim, 2012, LeSage & Pace, 2008), conventional gravity models that do not take into account the spatial structure/autocorrelation could result in biased estimations of distance decay parameters. To address this problem, Chun et al. (2012) suggest the use of network autocorrelation. Network autocorrelation measures the degree to which spatial dependency exists among flows. Such dependence suggests that flows may be dependent on the proximity of origin regions and destination regions (Chun & Griffith, 2011). LeSage and Pace (2008) identify three types of spatial dependence for interregional flows: (1) flows with the same destination region and neighboring origin regions tend to be similar, suggesting an origin-based dependence; (2) flows with the same origin region and neighboring destination regions tend to be similar, denoting a destination-based dependence; (3) flows with both neighboring origin regions and neighboring destination regions tend to be similar, signifying an origin-destination dependence. Spatial econometric modeling has been integrated with traditional gravity models to embed the network autocorrelation (Chun, 2008; Chun et al., 2012; Chun & Griffith, 2011; Fischer & Griffith, 2008; LeSage & Pace, 2008). To the best of our knowledge, network autocorrelation has not been given sufficient attention by researchers when investigating the impact of geographical distances on social interactions in Cyberspace. To fill this gap, our research attempts to contribute to the ongoing debate of "death of distance" by using network autocorrelation with spatial econometric method to obtain more reliable distance parameter estimation.

The remainder of the article is organized as follows. Section 2 describes the data and methods that utilized in this research. Section 3 provides the analytical results. We offer the observations and findings in Section 4 as concluding remarks.

2. Data and methodology

Social links are retrieved and established from a Chinese social networking site called Weibo (http://www.weibo.com). Weibo is a Chinese version of Twitter. It was launched in August 2009 and has become the largest social media outlet in China (Wang, Wang, Ye, Zhu, & Lee, 2015). In the first quarter of 2017, monthly active users on Weibo reached 340 million (https://www.chinainternetwatch.com/20636/weibo-q1-2017/). Weibo data have been applied in various fields including environmental issues (Wang et al., 2017), human mobility (Wu, Wang, & Dai, 2016), housing prices (Li, Ye, Lee, Gong, & Qin, 2016), natural disaster (Wang et al., 2015), public health (Ye, Li, Yang, & Qin, 2016), urban expansion (Long, Zhai, Shen, & Ye, 2017), etc. Here, we extend the methodological specification discussed in LeSage and Pace (2008) and Chun et al. (2012) to Cyberspace by incorporating the concept of network autocorrelation in spatial interaction models with spatial econometric methods.

2.1. Data

In this study, 7,286,310 user profiles were collected during the period from September 12th 2011 to June 25th 2012 using the Open API released by Sina Weibo. Data collection was based on a chain-referral technique called snowball sampling. With this algorithm, a set of users were first randomly selected as seeds and their profiles were retrieved; then, the profiles belonging to the seeds' friends were solicited; finally, the profiles of friends' friends were collected. Please note that all of the duplicated user information was removed prior to the analysis.

After the data were collected, a filtering process was implemented to obtain active Weibo users by removing inactive users and fake followers. A user is defined as an inactive user if he/she has no activities in the past 6 months. Fake followers are generally registered by 'Internet bots' to follow people and send out spams. Internet bots are automated programs that could perform simple and repetitive tasks over the Internet. Therefore, these fake followers are rarely being followed, but they follow a large number of users. That is, these "users" usually have few followers but many friends. Based on this observation, we removed those users who have more than 500 friends but with the followers/friends ratio less than 0.1. Additionally, since we focus on the social links in mainland China, the users outside of mainland China were removed from our dataset. A geocoding process was conducted to geolocate Weibo users. Following Kryvasheyeu, Chen, Moro, Van Hentenryck, and Cebrian (2015), this process was based on provinces and cities reported in user profiles. Finally, we obtained the profiles from 5,845,329 geo-located active users that accounted for 80.2% of the original dataset. Among all the users, 55.524% were male, while 44.476% were female. In the collected dataset, each Weibo user has an average of 292.4 friends.

We set up the social links among users based on the follower-friend relationship. More specifically, if user A followed user B, i.e., A is B's follower and B is A's friend, a social link would go from user A to user B, represented by A → B. In this sense, the social links between users were directional. In total, we retrieved 367,044,740 directional social connections. Based upon users' profile locations, all social links were then spatially aggregated to cities. Notably, to maintain the continuity of analytical units, cities in the collected dataset consisted of centrally administrated municipalities, cities designated in the state plan, sub-