Use of social media for the detection and analysis of infectious diseases in China

Article

Use of Social Media for the Detection and Analysis of Infectious Diseases in China

Xinyue Ye [1], Shengwen Li [2], Xining Yang [3] and Chenglin Qin [4,*]

[1]   Department of Geography, Kent State University, Kent, OH 44240, USA; xye5@kent.edu
[2]   School of Information Engineering, China University of Geosciences, Wuhan 430074, China; swli@cug.edu.cn
[3]   Department of Geography and Geology, Eastern Michigan University, Ypsilanti, MI 48197, USA; westoxgeorge@gmail.com
[4]   School of Economics, Jinan University, Guangzhou 510632, China
[*]   Correspondence: qinchlin@jnu.edu.cn; Tel./Fax: +86-27-6877-8969

Abstract: Social media activity has become an important component of daily life for many people. Messages from Twitter (US) and Weibo (China) have shown their potential as important data sources for detecting and analyzing infectious diseases. Such emerging and dynamic new data sources allow us to predict how infectious diseases develop and evolve both spatially and temporally. We report the dynamics of dengue fever in China using messages from Weibo. We first extract and construct a list of keywords related to dengue fever in order to analyze how frequently these words appear in Weibo messages based on the Latent Dirichlet Allocation (LDA). Spatial analysis is then applied to detect how dengue fever cases cluster spatially and spread over time.

Keywords: social media; infectious disease; space; time; China

1. Introduction

Since 2013, 2.35 million cases of dengue fever have been reported in the Americas, and 37,687 of these dengue fever cases were considered to be severe [1]. Identifying the geographical ranges helps the public understand the risk posed by infectious disease outbreaks [2]. Early detection of disease activity, followed by a rapid response, can largely reduce the impact of both seasonal and pandemic influenza [3]. Social media analytics enable the possibility of infectious disease surveillance at a fine scale and in a timely manner [2]. Progress in the areas of geospatial analytics has led to increased intelligence in investigating the outbreak, transmission, and treatment of diseases at both local and global scales. Furthermore, geospatial services and sensor apps affect people's daily behavior and lifestyle [4].

The geographical context of health research has shifted from a data-scarce to a data-rich environment [5]. Social media websites such as Twitter (US) and Weibo (China) serve as platforms for both sharing and communicating information. The widespread application of mobile smart devices has facilitated social media usage because most micro-blog users update information in their social media using mobile devices. Social media messages contain new ideas and report events in real time [6]. The large spatiotemporal data in Weibo represents a goldmine to understand and model various social phenomena across spaces. For example, a web-enabled Geo-Twitter analytics system can be used to conduct crisis management based on place, time, and concept [7]. Based on the location of the tweets, obesity patterns have been mapped in a GIS environment [8]. Social media can also be used to assess the effects of natural disasters, such as post-earthquake microblogging [9] and real-time earthquake detection [10].