Online social activity reflects economic status

Jin-Hu Liu [a,b], Jun Wang [a,b], Junming Shao [b,c], Tao Zhou [a,c,*]

[a] CompleX Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] Data Mining Lab, Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 611731, China
[c] Big Data Research Center, University of Electronic Science and Technology of China, Chengdu 611731, China

HIGHLIGHTS

- The online social activity shows a strong linear correlation with economic development.
- The macro economic structure can be well reflected by social activity.
- Online social activity allows analyzing the economy status instantaneously and thus supports in-time decisions from individuals to countries.

ARTICLE INFO

ABSTRACT

To characterize economic development and diagnose the economic health condition, several popular indices such as gross domestic product (GDP), industrial structure and income growth are widely applied. However, computing these indices based on traditional economic census is usually costly and resources consuming, and more importantly, following a long time delay. In this paper, we analyzed nearly 200 million users' activities for four consecutive years in the largest social network (Sina Microblog) in China, aiming at exploring latent relationships between the online social activities and local economic status. Results indicate that online social activity has a strong correlation with local economic development and industrial structure, and more interestingly, allows revealing the macro-economic structure instantaneously with nearly no cost. Beyond, this work also provides a new venue to identify risky signal in local economic structure.

© 2016 Published by Elsevier B.V.

1. Introduction

With the fall of the Iron Curtain and the upheaval of East European, the international environment has changed profoundly, in which security concerns are gradually dominated by economy instead of military considerations [1,2]. In fact, economic status not only drives the development of national defense, but more directly, affects our lives ranging from personal investment to economic regulation in government [3–5]. For instance, it is crucial for a mayor to consider the local economic structure in this city whenever he or she tends to make economic-related principles. As for ordinary people, the economic development usually accompanies with considerable improvement of the living standards. Meanwhile, our money-oriented plans or decisions explicitly or implicitly depend on the local or national economic status. How to provide an efficient and comprehensive view of the health of economy is thus of great importance for governments as well as individuals. Currently, economic census provides a straightforward way to present a clear picture of many facets of the

---

national or global economy in terms of many indices such as Gross Domestic Product (GDP), industrial structure and income growth [6–8]. However, computing these indices is usually a non-trivial task as they are always involved with considerable resources for a long time. For example, the widely accepted index, GDP, is defined as "an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production" [9]. The calculation of such index needs to collect data from distinct local governments and then integrate all data together. During this process, there are two predominant factors arisen. Firstly, the economic census requires a large quantity of manual labor, materials and other related resources. And more importantly, the procedure is time consuming and thus many economy-oriented decisions cannot be made in time. Although some new economic census techniques have been developed in recent decades to speed up the process such as sampling, they often suffer in accuracy as the statistics are derived from the partial rather than the whole data. In light of these problems, in recent years a few indirect indices have been introduced to quickly reflect the economic status. A famous index, *Keqiang Index* [10], consisting of three indicators: the railway cargo volume, electricity consumption and loans disbursed by banks, is proposed to measure the economy of China. Although the index allows reflecting the economic development, it is insufficient to provide a comprehensive overview of the economy as the measure heavily relies on industry and largely ignores agriculture and services. The Producer Price Index (PPI) is another popular index, which is used to measure the average changes in prices received by domestic producers for their output [11]. This measure manifests the economic effects on people's daily life, providing a potential hint of the stability of economy and society. Although there also exist some more complicated indices such as Consumer Price Index (CPI), Social Retail Goods (SRG) and Foreign Direct Investment (FDI) to characterize economic development or economic structure, it is a non-trivial task to obtain them due to their long-term data collection and calculation procedures. Therefore, a fast, effective and comprehensive strategy to bring deep insight into the economic status is highly desired.

Social networks, such as Facebook, Twitter and Sina Microblog, are becoming the primary venues for people to obtain and share information on a global scale [12,13]. With no doubt, the new social media is mainly driven by the advances of information technology [14]. However, many researches also have demonstrated that national economy status and policies play an important role on the growth rate, diversity and stability of social networks [15–20]. For instance, Katona et al. [15] have found that economy has significant effects on the structure of World Wide Web. Ioannides et al. [16] have studied individual outcomes in a dynamic environment in the presence of social interactions. Results show that the topology of social interactions is temporally changed once the individual outcomes vary continuously. Besides of the economists, socialists also point out that social networks and economic networks are mutually interrelated [21–23]. Actually, social networks permeate our social and economic lives and play a central role in the transmission of information about job opportunities, and are critical to the trade of many goods and services [21]. Eagle et al. [22] have showed that the regional communication diversity is strongly correlated with economic development. Furthermore, the results in Ref. [23] have showed high correlation in many of the cases revealing the diversity of socio-economic insights that can be inferred using only mobile phone call data. Moreover, Bettencourt et al. [24] have shown the strong relationships among several social and economic indices. Beyond, some companies have attempted to utilize online networks properties to reflect economic status, e.g. Taobao CPI [25]. Motivated by these studies, in this paper, we concentrate on analyzing and quantifying the latent relationship between economic status and online social activities, and propose a simple yet effective method to analyze the macro economic structure in a data mining framework. Building upon the data-driven analysis, the work has several interesting findings: (i) The online social activity shows a strong linear correlation with economic development; (ii) The macro economic structure can be well reflected by social activity; (iii) Online social activity allows analyzing the economy status instantaneously and thus supports in-time decisions from individuals to countries.

2. Materials and methods

2.1. Data acquisition and description

Here, to investigate the latent relationships between the social activities and economic status, we focus on the primary social network in China: Sina Microblog (SM), and the economic data is derived from the National Statistic Bureau of the People's Republic of China.

Sina Microblog. This data was collected from Sina Microblog (www.weibo.com), which is the leading social network in China and was launched in 2009 by Sina Corporation. Like Twitter, approximate 100 million messages have been posted each day on this platform [26,27]. Here we collect nearly 200 million online registered users from the year 2009 to 2012. For each microblogger, the registered date, verified information and location information including province and prefecture-level city are all examined. More than 97% microbloggers are ordinary users (opposed to verified users). We ransack all users' profiles in our database and get the statistic results, containing location information, gender, registered date and so on, which are available online at http://staff.uestc.edu.cn/shaojunming/files/2015/10/data.zip. The basic statistics of the data set are summarized in Table 1.

National economic data. The national economic data has been collected from the official book entitled "China City Statistical Yearbook", which is published by National Statistic Bureau (NSB) of the People's Republic of China in each year. In the statistical yearbook, major economic and social indices are reported, such as the total population, resident population, GDP, average GDP, industrial structure, to mention a few. Due to the time-consuming data collection and calculation procedures, the statistical yearbook cannot be published at the same year, but usually with about one-year delay. From