Research on the Spatial Structure and Dynamics of Socio-Economic Systems

by

Jian Gao

Submitted to the School of Computer Science and Engineering in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Software and Theory at the University of Electronic Science and Technology of China

Abstract

Socio-economic systems are an important branch of complex systems, which involves the complex interactions between people's economic activities and the social environment in which they live. With the constant change of cognition and behavior, people's subjective decision-making process greatly affects the operation of socio-economic systems. To accurately and timely perceive socioeconomic situation and to reveal and understand the law of socioeconomic development have great theoretical and practical values. Revealing the status of socioeconomic development in many aspects and predicting the development trends with desirable accuracy can greatly help to guide socioeconomic decision-making. Uncovering the socioeconomic behavioral patterns of individuals can contribute to gradually realizing predictive management. Quantifying the macro socioeconomic structure can help to explore the path of economic development. How to effectively analyze the structure and evolution of socio-economic systems is an important scientific issue in the interdisciplinary research field, and it has recently received great attention from many related disciplines including computer science, network science, complexity science, statistical physics and socioeconomics.

Traditional socioeconomic research relies mainly on qualitative or semi-quantitative methods, which makes it difficult to understand relevant issues at the mechanism level. The process that calculates macroeconomic indicators based on traditional census data not only consumes substantial resources, but also follows a long-time delay. Besides, traditional analytical methods have difficulty in tracking the structural transformation of economic development, fail to quantify the complexity of economic development and are lack of predictive power on development trends. The recent simultaneous development of hardware and technology is driving a new wave of big data, which has brought unprecedented opportunities and changes to socioeconomic research. The advances in methods of data acquisition have increased the availability of large-scale socioeconomic data, and the increases in the size and diversity of data have contributed to the transformation of socio-economic analytical tools and methodologies. The application of novel data and methods has gradually increased the level of quantification in socioeconomic research and led to the emergence of a new scientific branch, named Computational Socioeconomics. Under the framework of computational socioeconomics, this dissertation will investigate the status inference and structural modeling of socio-

economic systems from the micro, meso and macro levels, and explore the evolution of economic structure and the optimal strategy for economic development through theoretical and empirical studies. In particular, studies at different levels are based on the similar theoretical basis of network spatial structure and dynamics. The main contents and major contributions of this dissertation are summarized as follows:

(1) At the micro level, the predictive management of socio-economic systems was studied based on unobtrusive behavioral data. By analyzing data recorded by anonymized campus cards, we proposed a novel orderliness measure to quantify the regularity of individual behavior. Orderliness is significantly correlated with student academic performance, and it can largely improve the performance of learning-to-ranking algorithm on predicting student academic performance. Based on the analysis of two employee networks built on data from an enterprise socialization platform, we found that the locations of employees in both networks are predictive to the possibility of their promotion and resignation. In particular, action network has stronger predictive power than social network, and predicting resignation is easier than predicting promotion. Moreover, by analyzing large-scale online platform data, we revealed some socio-economic phenomena in a quantitative way, including keeping team size below 8 can improve employee's communication and performance, the size of Chinese social circle is also around Dunbar's Number 150, and there are height premium and gender inequality in the workplace.

(2) At the meso level, the ranking of socio-economic systems was studied based on online user rating data. To solve the of problem reputation ranking, we proposed a group-based reputation ranking (GR) method. Instead of relying on the traditional assumption of product quality, GR method calculates user reputation based on the size of rating groups. Experiments based on real-world datasets showed that GR method outperforms benchmark methods in the accuracy of ranking users by their reputation. By introducing an iterative process into the GR method, we further proposed an iterative group-based ranking (IGR) method. Considering both the number and the reputation of users when calculating the group size, GR method exhibits better accuracy and robustness in reputation ranking. To solve the problem of object ranking, we proposed a novel vertex similarity measure, named CosRA index, based on which we developed a CosRA-based recommendation algorithm that exhibits better performance. Further, we proposed a trust-based recommendation algorithm, named CosRA+T, and found that relying too much on trust relations among users is detrimental to recommendation performance.

(3) At the macro level, socio-economic structures were quantified and analyzed based on large-scale real data. Using firm registration information data, we quantified China's regional economic complexity. We found that ECI index and Fitness index exhibit comparable predictive power for China's regional economic development, and economic complexity is negative correlated with income inequality. Using labor and firm data, we built Brazil's and China's regional industry space, respectively. We found that both industry spaces exhibit a ``core-periphery'' structure, where industries with high and low level of sophistication occupy the core and the periphery of the industry space, respectively. Moreover, China's regional industry space has a ``dumbbell'' structure, and its time evolution has regional competitions. Based on Weibo and resume data, we built information flow and talent mobility network, respectively. We found that regional economic status can be inferred from the structure of both networks. In particular, talent

mobility network exhibits a stronger predictive power, and combining the structures of both networks can explain about 84\% of the variance in GDP.

(4) In economic development and structure evolution, the path of economic evolution and the strategy of industrial upgrading were studied based on spatial networks. By leveraging the spatial network model and the spreading process, we revealed the effects of the spatial structure of networks on information diffusion. We found that the distribution of long-range links of spatial networks can change the phase transition of bootstrap percolation, where the exponent -1 of the distribution of long-range links is a critical value for the presence of a double phase transition with two nearly constant critical points. For industry space and geographical adjacent networks, we proposed the inter-industry learning and the inter-regional learning for economic development, respectively. We found that both collective learning channels can increase the probability of development new industries, while they exhibit an alternative effect. Moreover, we explored the optimal strategy for economic development using both theoretical and empirical analyses. We found that reducing geographical distance can enhance the collective learning effects, introducing high-speed rail can increase regional industrial similarity and productivity, and both collective learning channels have optimal strategies for industrial development.

Computational socioeconomics is an emerging research branch, and it faces new challenges and opportunities in both data and methods. In future studies, it is worthwhile to further explore the spatial structure and dynamics of socio-economic systems, and to improve the perception of socioeconomic situation and the understanding of the law of development. In the long run, data-driven research paradigm will become the mainstream methodology for solving social and economic problems and will profoundly change the landscape of socioeconomic research.

Keywords: complex networks, socio-economic systems, ranking method, economic complexity, network structure

Thesis Supervisor: Tao Zhou
Title: Professor of Computer Science and Technology
University of Electronic Science and Technology of China