Applied Geography 96 (2018) 77–85

Contents lists available at ScienceDirect

Applied Geography

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

The rich-club phenomenon of China's population flow network during the country's spring festival

Ye Wei[a], Wei Song[b,*], Chunliang Xiu[c], Ziyu Zhao[d]

[a] School of Geographical Sciences, Northeast Normal University, Changchun, Jilin 130024, China
[b] Department of Geography and Geosciences, University of Louisville, Louisville, KY 40292, USA
[c] College of Jang Ho Architecture, Northeastern University, Shenyang, Liaoning 110169, China
[d] College of Earth Sciences, Jilin University, Changchun, Jilin 130000, China

ARTICLE INFO

Keywords:
Population flow network
Weighted networks
Rich-club coefficient
Assortativity

ABSTRACT

This paper provides a structural analysis of the rich-club phenomenon of China's population flow network (PFN) during the country's 2015 Spring Festival. We define the PFN as the weighted network where nodes are cities and links account for the population flows. The global weighted rich-club coefficient, local weighted rich-club coefficient, assortativity and community detection are used to quantitatively characterize the rich-club tendency of the PFN. Results of global rich-club coefficient uncover the existence of a significant rich-club phenomenon in the PFN. Six cities including Beijing, Shanghai, Suzhou, Guangzhou, Shenzhen and Dongguan are identified as rich-club members. And, the value of assortativity (−0.216) shows that the network is a dissortative network. We therefore conclude that the network has typical oligarchy characteristics. Subsequently, our local rich-club coefficient analysis reveals that most (77.66%) of the nodes in the network prefer to connect to rich nodes rather than non-rich nodes. Community detection demonstrates that the network presents obvious geographical agglomeration characteristics and rich-club cities also play a dominant local role in their own "responsible" communities. These findings suggest that China's PFN during the country's Spring Festival is a specific kind of hub-and-spoke network where the rich-club cities are hubs and their interconnections form the backbones. We also identify some specific problems (e.g., the community in southwest China lacks its own rich-club cities resulting in its less integration into the national transportation and migration network). Being the first to discuss rich-club issues of population flow networks in China, findings of this research may be useful for a wide spectrum of applications ranging from population migration, urbanization and traffic infrastructure planning in China.

1. Introduction

Population flow is the collective result of individual human behavior. Like other flows, population flow has origins, destinations and flow paths that constitute a complex population flow network (PFN). Although the origins and destinations in a PFN have specific physical locations, due to data collection limitations and constraints each is often linked to a territorial area. For instance, origins and destinations may be countries, provinces, cities or smaller administrative districts. In other words, population flows build networks between territorial areas (Conti, Guarneri, & Tucci, 2010), and "forces the researchers to consider both origin and destination contexts and the relationship between them" (Bakewell, 2014). In this case, although the PFN exists in a physical space, the often-discussed PFNs are just types of projections of a real network in certain spatial scales. PFNs can be divided into two basic types according to their spatial scales: (1) those operating at urban and zonal levels, namely commuting flow networks (Patuelli, Reggiani, Gorman, Nijkamp, & Bade, 2007; Reggiani, Bucci, & Russo, 2014), and (2) those occurring at the regional level, that is, migration flow networks (Chun & Griffith, 2011; Consterdine & Everton, 2012). Studies on migration flow networks, in contrast to commuting flow networks, are found less frequently in the literature mainly due to the shortage of reliable and timely data. Therefore, every single case study is valuable to the richness of theory and the development of migration flow network research methodologies. Because migration flow networks could be regarded as complex networks, kinds of network metrics (e.g., degree distribution, average path length, transitivity and community structure analysis) are tentatively applied on the structural analysis of migration flow networks (Davis, D'Odorico, Laio, & Ridolfi, 2013). Such metrics play a major role unfolding the characteristics and nature of migration flow networks. More new network matrics (e.g., the rich-club coefficient) are expected to emerge and to be applied to the

* Corresponding author. Department of Geography and Geosciences, University of Louisville, Louisville, KY 40292, USA.
E-mail addresses: weiy742@nenu.edu.cn (Y. Wei), wei.song@louisville.edu (W. Song), xiucl@nenu.edu.cn (C. Xiu), zyzhao14@mails.jlu.edu.cn (Z. Zhao).

https://doi.org/10.1016/j.apgeog.2018.05.009
Received October 2017; Received in revised form 8 May 2018; Accepted 11 May 2018
0143-6228/ Published by Elsevier Ltd.

structural analysis of complex networks.

The phenomenon of influential people becoming friends and forming groups is known as society's rich-club phenomenon. In network sciences, the concept of 'rich-club' refers to the subgroups of important or influential (rich) nodes that preferentially and intensely interact with one another (Colizza, Flammini, Serrano, & Vespignani1, 2006; Zhou & Mondragón, 2004). Rich-club analysis is quite different from the typical methods used for identifying well-connected leaders or important subgroups, such as “power centrality” or “community detection” (Ansell, Bichir, & Zhou, 2016; Newman, 2004b). “Rich” nodes not only form a cohesive group among themselves but they also maintain connections to “poor” nodes. Previous studies reveal that the rich-club phenomenon exists in many networks including international trade, aviation and traffic networks, the human connectome and the research-writing network (Van den Heuvel & Sporns, 2011; Colizza, Pastor-Satorras, & Vespignani, 2007). Rich clubs can serve as a network's backbone for optimizing the routing between peripheral nodes (Van den Heuvel, Kahn, Goñi, & Sporns, 2012), and can substantially affect other key properties of the networks, including clustering, assortativity and transitivity (Alstott, Panzarasa, Rubinov, Bullmore, & Vértes, 2014; Xu, Zhang, & Small, 2010). Research has demonstrated that the small number of rich nodes play a core role in static and dynamic processes on complex networks, such as targeted attack, cascade failure, and disease spreading (Albert, Jeong, & Barabasi, 2000; Motter & Lai, 2002; Small, Tse, & Walker, 2006; Xu et al., 2010). The rich-club phenomenon has also been used to signify the dominance of an oligarchy of the rich in complex networks (Ansell et al., 2016; Smilkov & Kocarev, 2010), and the rich-clubness thus has a high interpretative ability on the structure of hub-and-spoke and core-periphery networks (Cinelli, Ferraro, & Iovanella, 2017; Xu & Harriss, 2008). Comparatively, not much empirical research has been conducted to uncover the existence of rich club phenomenon in human migration. It is therefore of great theoretical and practical importance to extend the rich-club framework to the study of human migration networks.

The rich-club coefficient, which is an effective tool to measure the rich-club phenomenon, was first put forward by Zhou and Mondragón (2004) in the context of the Internet. However, early-stage research on the rich-club coefficient was based on undirected and unweighted networks. The definition of rich nodes was measured by the degree of external connections, and the existence of the rich-club effect was verified by the existence of connections between rich nodes. It is clear, however, that this coefficient ignored the most valuable weight (connection strength) information for weighted networks. In practical applications, connection strength is often the core or essential element of many networks. Therefore, Opsahl, Colizza, Panzarasa, and Ramasco (2008) proposed a weighted rich-club coefficient to take into account the weights when selecting rich nodes in order to define the relationship between rich nodes and to perform a weighted network analysis. Colizza et al. (2006), using examples drawn from biological, social and technological domains, presented an analysis that enables measurement of rich-club ordering and its relationship with the function and dynamics of networks. Xing, Lu, and Chen (2016) found that the rich-club coefficient is very useful in their Weighted Complex Network Analysis of the Shanghai Rail Transit System. Ansell et al. (2016) confirmed the efficacy of this rich-club coefficient by analyzing and comparing several urban networks in Sao Paulo's urban infrastructure networks and the transportation policy networks in the cities of Los Angeles and Chicago. Ducruet, Cuyala, and Hosni (2016) empirically employed the rich-club coefficient to study the changing influence of city-systems on global shipping networks.

The inter-city PFN is a typical directed weighted geographical network, with unevenly distributed population flows among cities (nodes) which themselves vary greatly in influence. The crucial roles played by those “influential” cities need to be recognized in the network analysis. Spatially, these prominent cities dominate the connections within their respective local regions or communities. Collectively, they also exert enormous influence over the entire network and largely shape the network structure. The examination of interactions between the prominent nodes is thus essential for understanding the spatial and organizational structure and dynamics of the PFN. Specifically, this research was designed to identify and examine the rich-club phenomenon in China's PFN during the country's Spring Festival, based mainly on the weighted rich-club coefficient.

The paper is structured as follows. Section 2 introduces China's PFN during the Spring Festival period and data sources. Research methods are described in Section 3, followed by the presentation of analytical results in Section 4. Discussions and concluding remarks are given in Section 5.

2. China's PFN during the country's spring festival

In this paper, we define the PFN as the weighted network where nodes are cities and links account for population flows. To build the PFN, we derived raw data from the Baidu Map Spring Festival Population Migration Data (Baidu Migration Data) platform developed by the Baidu search engine. Baidu Migration Data is a population migration data project built upon location-based services (LBSs). It was launched by Baidu during China's Spring Festival. There are two salient points that illustrate the scientific basis of this data. First, during China's Spring Festival (CSF), most migrant people are active mobile internet users. Communication is a basic need during the migration process. Therefore, there is a high degree of matching between mobile internet users and the migrant population. Mobile internet users account for a large proportion of the migrant population and as a consequence, can well represent the country's migrant population. Secondly, the Baidu LBSs open platform provides location-based services to hundreds of thousands of applications. According to statistics, Baidu LBS applications cover hundreds of millions of mobile phones, accounting for about 80% of the total number of mobile phones used. With accurate positioning information and the cloud's strong data processing capability, Baidu Migration Data can comprehensively reflect the population's migration status in real-time and reveal population linkages between cities. Due to the large number of mobile internet users in China and the intensive population migration during the country's Spring Festival, billions of LBSs positioning data are collected each day and are used for computation and analysis on the cloud computing platform in order to both obtain the crowd's migration trajectory and display it in a visual form. This project enables users to have a comprehensive, real-time, dynamic and intuitive view of China's nationwide population migration trajectory and characteristics. In short, Baidu maps the migration trajectory based on mobile internet users' positioning data, visualizes it and then displays it to users.

This study focused on the PFN during CSF due to its unique significance to the nation's population and the country's socio-economic development. CSF travel refers to the phenomenon where a large number of travelers commute between locations during the country's Spring Festival. This phenomenon first occurred in the 1980's. At present, CSF travel is considered to be the largest, periodical and unique human migration in history. CSF travel is a unique socio-economic phenomenon during the country's transition period. Hidden behind the massive movement of people is the underlying problem of an imbalance between the population and regional economic development. There exists a mismatch between the population's distribution and the space for economic activities and employment. The competitiveness and attractiveness of more developed eastern coastal regions, particularly large cities, pull migrant workers from the vast less-developed central and western regions. On the other hand, the deep-rooted cultural tradition of going home for the Chinese New Year drives a large number of migrant workers from their places of work to return to their hometowns as CSF approaches, thus forming strong home-bound flows to their hometowns. After CSF, these migrants leave home and go back to their work places, forming return flows toward cities. Improvements in the