Regional Economic Inequality Analysis:

A Comparative Study of the United States and China

by

Sizhe Wang

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved April 2016 by the
Graduate Supervisory Committee:

Sergio Rey, Chair
Wenwen Li
Deborah Salon

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

Economic inequality is always presented as how economic metrics vary amongst individuals in a group, amongst groups in a population, or amongst some regions. Economic inequality can substantially impact the social environment, socioeconomics as well as human living standard. Since economic inequality always plays an important role in our social environment, its study has attracted much attention from scholars in various research fields, such as development economics, sociology and political science. On the other hand, economic inequality can result from many factors, phenomena, and complex procedures, including policy, ethnic, education, globalization and etc. However, the spatial dimension in economic inequality research did not draw much attention from scholars until early 2000s. Spatial dependency, perform key roles in economic inequality analysis. The spatial econometric methods do not merely convey a consequence of the characters of the data exclusively. More importantly, they also respect and quantify the spatial effects in the economic inequality. As aforementioned, although regional economic inequality starts to attract scholars' attention in both economy and regional science domains, corresponding methodologies to examine such regional inequality remain in their preliminary phase, which need substantial further exploration. My thesis aims at contributing to the body of knowledge in the method development to support economic inequality studies by exploring the feasibility of a set of new analytical methods in use of regional inequality analysis. These methods include Theil's T statistic, geographical rank Markov and new methods applying graph theory. The thesis will also leverage these methods to compare the inequality between China and US, two large economic entities in the world, because of the long history of economic development as well as the corresponding evolution of inequality in US; the rapid economic development and consequent high variation of economic inequality in China.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Regional Economic Inequality

Economic inequality is always presented as how economic metrics vary amongst individuals in a group, amongst groups in a population, or amongst some regions, for instance, countries. Economic inequality can impact substantially the social environment, socioeconomics as well as human living standard. For instance, the relationships between economic inequality and crime rate (Kelly, 2000), human health (Coburn, 2000; Deaton, 2001), economic growth (Kuznets, 1955; Aghion, Caroli, & Garcia-Penalosa, 1999; Panayotou et al., 2000; Thorbecke & Charumilind, 2002), monopolization of the labor force (Castells Quintana & Royuela Mora, 2012), even biodiversity (T. G. Holland, Peterson, & Gonzalez, 2009) are investigated in the literature. Since economic inequality always plays an important role in our social environment, its study has attracted much attention from scholars in various research fields, such as development economics, social studies and political science.

Economic inequality can result from many factors, phenomena, and complex procedures, which include (but are not only limited to) policy (Smeeding, 2005, s1), globalization (Stiglitz, 2002; Navarro et al., 2007; Von Braun, Díaz-Bonilla, Pinstrup-Andersen, et al., 2008), education (Arrow, Bowles, & Durlauf, 2000; Becker & Murphy, 2007; Keller, 2010), labor market (Katz et al., 1999; Carter, 2013), gender (Korpi, 2000; Seguino, 2000), ethnics (Raudenbush & Kasim, 1998; Ostby, Nordas, & Rod, 2009), and geographical and spatial factors. However, the spatial dimension in economic inequality research did not draw much attention from scholars until early 2000s (Rey, 2001). According to Novotnỳ, spatial factors, i.e. spatial dependency, perform key roles in economic inequality analysis (Novotnỳ, 2007).

The spatial econometric methods do not merely convey a consequence of the characters of the data exclusively. More importantly, they also respect and quantify the spatial effects in the economic inequality (Anselin, 2001).

Though there could be many causes of economic inequality, in this thesis, how the economic inequality is affected by spatial factors will be the emphasis. Some scholars use analogous shapes, a reversed "U", to depict the expected evolution of economic inequality that started at low level, gradually increased and restored to low in the end (Kuznets, 1955; Williamson, 1965). In recent years, research that highlights the geographical factors as the reason contributing to economic inequality becomes attractive, since the spatial dimensions of these dynamics are largely untouched in those trajectories. They do provide the dynamics statistically of some overall or whole map, but miss the spatial effects of those dynamics (Rey, 2015). Comparing with other causes, the spatial factor is important because it indeed does have an impact on economic inequality. For instance, Rey and Janikas (2005) find that the different relationships between spatial clustering in state income levels and national economic growth based on income when using different spatial scales; and a larger income inequality is shown in U.S. as the regions are further decomposed (Rey, 2004).

There is a large literature on regional inequality analysis, in which various strategies and methods are proposed and examined by scholars and researchers. Traditional approaches that investigate the regional inequality sometimes have difficulty in demonstrating the spatial agglomeration and the importance of regions in shaping trends of regional inequality. Wei and Ye (2009) leveraged the exploratory spatial data analysis (ESDA) and geographically weighted regression (GWR) with a combined top-down and bottom-up strategy to discover the driving forces and trends of regional inequality. An exploratory analysis method which integrates inequality indices, mobility indices, kernel density estimation, spatial autocorrelation statistics and scale variances was created by Yamamoto (2007), and this method enables the exploration of the average per capita income across different spatial scales. Rey and Sastré-Gutiérrez (2010) also utilized ESDA to examine how the spatial clustering and heterogeneity impact the evolution of regional inequality.

Theil's Index and the decomposition of inequality measures, as well as the intra-group and inter-group inequality is another effective and widely utilized framework for inequality analysis. Ye and Wei (2005) investigated the multi-scalar pattern in regional development and emerging clusters with inequality analysis from both intra-provincial and inter-county perspectives. Yildirim, Öcal, and Özyildirim (2009) investigated regional income inequality and the convergence dynamics in Turkey for the time period 1987—2001, and applied Theil coefficient of concentration index to study diverse convergence process and decomposition of inequality. Theil's index also helps to identify geographical heterogeneity of the inequality. Paredes, Iturra, and Lufin (2014, ahead-of-print) focused on the relationship of individuals and spatial inequality, and proposed a spatial decomposition based on Theil framework for the inequality within diverse geographical scales, which includes regional, provincial and county levels.

Another hot spot of performing regional inequality analysis is to leverage the Markov framework, especially its spatial variation. In 2001, Rey (2001) suggested some new empirical strategies for spatiotemporally investigating the evolution of regional income distribution. These strategies are developed based on extensions of the classical Markov methods and provide a more comprehensive view on transition dynamics in spatial dimensions. Subsequently, Rey (2013) introduced rank Markov framework to mitigate the issue arising from the discretization of classic Markov methods. These methods and extensions have also applied in very recent works. For instance, Rey and Gutiérrez (2015) employed the spatial Markov chain and spatio-temporal mobility measures to examine the variation of regional economic inequality dynamics for U.S. and Mexico - two adjacent national systems.

Additionally, recent attempts in the literature also include the integration of graph theory into inequality analysis. Kets, Iyengar, Sethi, and Bowles (2011) investigated how the structure of a graph can impact the degree of inequality that the graph represents, and how this inequality can be preserved depending on different network structures. The authors found that it was the cardinality of the largest independent set, which is also known to be the key network property, that determined the inequality. Another work that contributes to this field was done by Palestini and Pignataro (2015). They pointed out that it could be a complex process to decide whether some policies (e.g. budget, fiscal choice, political will) can be assessed together with inequality analysis. To address this issue, an approach based on the connected network was proposed. In this network, income distribution resulted from policies are represented by the vertices, and the possibilities of applying some policies are indicated by edges. Although the graph-based inequality analysis started to proliferate in literature, the applications of graph theory on regional economic inequality analysis are still not thoroughly studied.

1.2 The Case of US and China

The evolution of economic inequality in both China and the U.S. has been extensively explored and studied. Autor, Katz, and Kearney (2008) investigated spatial trends in the U.S. of wage inequality and found an "episodic" character of the inequality trend; Socio-economic inequality among US adolescents was examined by Zhang and Wang (2007); and the work of Meng, Shen, and Xue (2013) contributed to the field in that they linked the changes in earnings inequality with measurable structural and institutional changes. Although the research on inequality in the two countries has been performed based on various themes, individual groups, time periods, and the spatial dependence is seldom considered and integrated into inequality analysis. There are always economic gaps between countries, and also between different regions within the same country. In 1932, the per capita income of the richest state (New York) in U.S. was 5.32 times larger than the poorest one (Mississippi). Such regional inequality is far more severe in China. Investigating per capita GDP (Gross Domestic Product) of China in 1978, this number was 14.20. As two biggest economic entities in the world, US experienced the Great Depression and is now highly developed. There is a long history of economic development as well as the corresponding evolution of inequality. China, in contrast, is one of the countries that have the highest economic growth rate in the world after the reforms of 1978. Based on preliminary studies, such rapid economic development resulted in high variation of economic inequality. Because of their representativeness, the regional inequality in both countries will be explored in this thesis and the differences in the economic inequality patterns will be quantified and systematically compared. Specifically, the comparison between the inequality of the early development in US after the Great Depression and in China during the post-reform period will be conducted. Chapters 3-4 describe these studies in details.

1.3 Research Objectives

As aforementioned, although regional economic inequality starts to attract scholars' attention in both economy and regional science domains, corresponding methodologies to examine such regional inequality remain in their preliminary phase and are quite immature, which need substantial further exploration. This thesis aims at contributing to the body of knowledge in the method development to support economic inequality studies by exploring the feasibility of a set of new analytical methods in use of regional inequality analysis. These methods include geographical rank Markov and methods applying graph theory. The thesis will also leverage these new methods to compare the inequality between China and US, two large economic entities in the world.

1.4 Organization

The regional economic inequality of U.S. and China will be examined in three aspects in this thesis. In Chapter 2, I will utilize a classic index Theil's T to reveal the economic inequality characteristics of the two countries as a reference for the following applications of novel methods on spatial dependence analysis. Theil's T is utilized for providing the overall economic inequality conditions in our studied areas based on regional decomposition. Chapter 3 introduces the geographical rank Markov method and its application. Different from the global statistic proposed in the former chapter, this method allows for the spatial decomposition of inequality levels and therefore helps to analyze the economic inequality in different subareas of China and the US, respectively. Besides, based on this ability of acquiring the interregional economic inequality, this method is also leveraged to identify whether are there any clusters, in which adjacent subareas share the similar economic inequality level, distributed in our studied countries or not. In Chapter 4, graph theory will be applied to study regional inequality. Specifically, Normalized Total Degree and Degree Centrality are applied to the economic inequality analysis. This method allows not only the spatial decomposition but also the temporal decomposition of time-series economic data in studying the regional inequality. Though Chapter 3 answered the question of whether the spatial autocorrelation exists or not, in this chapter, the newly developed method is used to further depict the spatial autocorrelation statistically during the whole time period as well as some specified partial time spans. Chapter 5 summarizes the results and the contribution of this thesis and offers directions for future research.

Chapter 2

GLOBAL INEQUALITY ANALYSIS

2.1 Introduction

In this chapter, the global economic inequality patterns within both US and China are studied and revealed because the results will provide a baseline study to the regional analysis of inequality in later chapters. Theil's T, which is an indicator that shows the global inequality at a certain point of time, is applied to compare the economic inequality between the two countries, because of its effectiveness and popularity in inequality studies, and because of its capability of showing the variation of global economic inequality patterns. In previous works of inequality analysis, a bell analogy of inequality variation patterns over time was introduced. Such an analogy is also known as the inverted U pattern, posited by Kuznets (1955) for social, or personal income, inequality, and Williamson (1965) for regional inequality, in that the shape of the bell portrays the expected evolution of each type of inequality as an economic system develops. Taking advantage of time series data, the "bell" patterns of two countries' inequality can be reproduced. Comparing those two patterns will provide us some valuable evidence to support the exploration of new spatial analytical methods in later chapters.

To calculate the Theil's Ts in each year of the two countries, and visualize the inequality variation patterns of different countries respectively, some tools and packages will be utilized. PySAL (Python Spatial Analysis Library) (Rey & Anselin, 2007) is one of the most important tools adopted in this thesis to handle the statistical calculations.

2.2 Theil's T

In this paper, the Theil index (Theil, 1967) is utilized to produce global inequality. Because there is no spatial decomposition involved in the analysis, this method is also called a global method. Among a large amount of models developed to measure inequality, Theil's inequality index is one of the best known and most popular measures (Cowell, 2011). Its mathematical equation can be expressed as follows:

$$T = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\overline{y}}\log(\frac{y_i}{\overline{y}}) \tag{2.1}$$

where n is the number of observed regions, y_i is the economic index of observation i, and $\overline{y}$ is the mean of all y_i. In current context, economic index would be either per capita income or per capita GDP (Gross Domestic Product).

This measure derives from entropy concept, which characterizes the "degree of disorder" of a system. When all the observations hold the same value, the system gets completely disordered, and entropy reach its maximum. Meanwhile, this situation presents the perfect equality. Theil subtracted actual entropy from that max value and produced the equation 2.1. Theil's T using zero, the minimum T to reflect the lowest degree of inequality. On the other hand, as T becomes larger, the degree of inequality also increases.

Equation 2.1 give us a global view of inequality. However, it is important to decompose the measure into multiple components. It is because that even though we retrieve the relatively low inequality degree globally, the situation could be bad in each of the objects we observed. They have the possibility to be polarized, when merely relying on a global measure. In literature, to decrease such effect, the global measure is often decomposed into two parts, "between-group" and "within-group", given as:

$$T = \frac{1}{n}\sum_{i=1}^{m}\frac{n_i y_i}{\overline{y}}\log(\frac{y_i}{\overline{y}}) + \frac{1}{n}\sum_{i=1}^{m}\frac{y_i}{\overline{y}}\sum_{j=1}^{n_i}\frac{y_{ij}}{y_i}\log(\frac{y_{ij}}{y_i}) \tag{2.2}$$

or simply put:

$$T = T_B + T_W \tag{2.3}$$

where m is the number of groups, n is the total number of observations, while n_i is the number of observations within group i; y_{ij} is the economic index of observation j in group i, y_i is the average value among group i, while $\overline{y}$ is the overall average.

In the right part of equation 2.2, the term in left side of plus sign is the "between-group" component, denoted as T_B, represents the intergroup inequality; the right side term is the "within-group" one, denoted as T_W, measures inequality within the same group. Theil's T measure is a useful device to quantify the inequality degree. But, as mentioned above, the geographical information is utilized in this method, but the spatial dependency is still under the veil, making it difficult to assess the spatial effects in the inequality patterns.

2.3 Empirical Results and Discussion

Because of the limitation in the data we utilized (the China data does not contain subdivisions of economic variables on county- or state- level), the "within-group" part of the inequality is ignored, when calculating Theil's T. We applied the Theil's model on the province level per capita GDP data of China with 30 mainland provinces, autonomies, and municipalities, and the U.S. per capita income at state level. The results are shown in Figure 2.1.

In this figure, the red curve indicates the variation of calculated Theil's Ts based on China per capital GDP over time, and the blue line shows the Theil's T values calculated using per capital income in the US. In both curves there is a "bell" pattern that shows an increase in first and a followed up decrease of Theil's T, which indicates the occurrence of a rapid economic development. US data starts in 1929 and ends in 2009, while China's data lasts from 1978 to 2012. The figure shows a very fluctuating curve of inequality changes in China. It starts at a high inequality status, drops down sharply around 1990, slightly goes up, and continues reducing after 2000. On the other hand, in comparison with China, there is a considerably

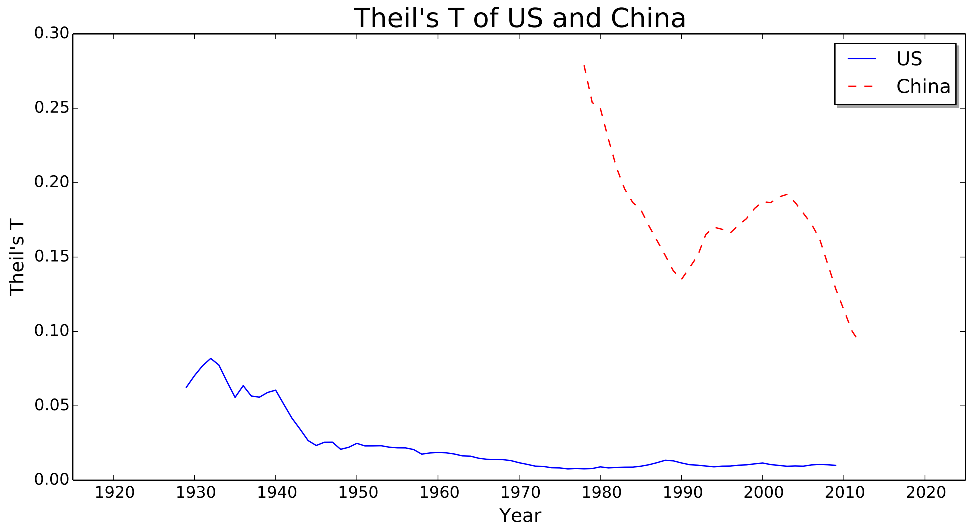

Figure 2.1: Economic Inequality Variation of US and China

distinctive pattern for the US. First, China suffers from a much more higher economic inequality over the same period of time. Additionally, unlike the great variation of economic inequality happened in China during the recent 30 years, the curve of US shows sort of flat shape after 1950. But, ignoring the big difference in scale, the pattern of China greatly resembles the pattern of US during early 1930s to middle 1940s, indicating that China is going through similar development patterns of the US between 1930s and 40s. Therefore, some time-based decomposition is planed to be applied to the method that will be introduced in Chapter 4 for the further investigation and comparison on the similar patterns.

Theil's T provides a convenient and intuitive metric to measure the overall regional economic inequality with the regional decomposed data. It also equips us with the option of further decomposing subareas to obtain more accurate inequality measures by utilizing the within-group partition of Theil's T. However, the missing spatial dimension of Theil's T makes it hard to determine interregional economic inequalities. This problem will be solved in next 2 chapters by introducing two new methods, which emphasize the spatial dimension and geographical factors.

Chapter 3

EXAMINATION OF GEOGRAPHIC RANK MARKOV AND ITS EXTENSIONS

3.1 Introduction

The first law of geography according to Tobler (1970) says "Everything is related to everything else, but near things are more related than distant things." This law can also be applied to studies of regional economic inequality. However, the role of spatial dependence in studies of regional inequality has been largely ignored (Rey, 2004) until recent years.

Non-spatial inequality analysis reveals the overall inequality of a study area. On the contrary, geographic rank Markov generalizes all time series data into one thematic map and discovers the interregional inequality, and helps us to understand the distribution of clusters of different classes of economic development. In this chapter, geographic rank Markov model will be used to create the First Mean Passage Time (FMPT) matrices of both China and US. Each of the metrices can then be used to generate a map showing the possibilities of a certain state gaining the ranks from others. A top level, a middle level, and a base level region of economic development status in the two countries will be chosen as the targets, and then compared and analyzed using the FMPT maps.

These FMPT maps will help to identify the space-time autocorrelation of a specific region. The space-time autocorrelation here means a region is more likely to gain the ranks of its neighbors, or its rank is more likely to be acquired by its neighbors. Therefore, when investigating the patterns of a rich or a poor region, if there shows a strong pattern of space-time autocorrelation, we may use it as evidence of economic growth inequality. This is because space-time autocorrelation indicates that a part of our study area remains poor or rich in the whole time period of study. However, there may be many sub-regions to investigate in a study, therefore data within multiple rows and columns in the FMPT matrices will be needed. Mapping matrices on a single row or column would be a waste of time. To provide a panorama view of regional inequality, 48 maps, each of which showing the space-time autocorrelation (by gaining ranks) of a state with all other states, are generated for US, and 30 maps are generated for China.

3.2 Data Preprocessing

In order to select a representative set of states (at top-level, middle-level and base level) for an in-depth analysis of regional inequality, a preprocessing phase is developed. This process involves ranking, normalizing, and averaging the economic variables, i.e. per capita income or GDP over time to obtain the top 3 richest and 3 poorest states, as well as a set of states falling within the middle-level of GDP or per capita income. The specific steps are listed below:

- Step 1: Assume the time series data is placed in a table whose columns indicate the time and rows represent the regions.
- Step 2: Normalize the economic data in each column by the maximum value in that column.
- Step 3: Sum up all the normalized values by row.
- Step 4: Divide the sums by the number of columns in the original table.

After this proprocessing, the ranked data for US and China are produced and shown in table 3.1 and table 3.2. As seen, the richest state over the time period of study in the US is Connecticut, followed by New York, and then New Jersey. The poorest three states are South Carolina, Arkansas, and Mississippi. In China, Shanghai, Beijing and Tianjian are the states having the highest averaged GPD; Yunnan and Guizhou, two states in the southwest

Connecticut	0.98050	New Hampshire	0.73905	Idaho	0.62441
New York	0.91355	Oregon	0.73668	Utah	0.61776
New Jersey	0.90921	Minnesota	0.72773	North Dakota	0.60466
Delaware	0.88877	Wisconsin	0.71338	Oklahoma	0.59646
California	0.88849	Indiana	0.69375	South Dakota	0.59638
Nevada	0.88322	Missouri	0.68651	Georgia	0.58307
Massachusetts	0.85998	Florida	0.68461	New Mexico	0.56946
Illinois	0.84589	Kansas	0.68409	North Carolina	0.56884
Maryland	0.83469	Virginia	0.68391	Louisiana	0.56562
Washington	0.79210	Nebraska	0.68030	Tennessee	0.56530
Rhode Island	0.77981	Iowa	0.67983	West Virginia	0.54963
Michigan	0.77475	Montana	0.66225	Kentucky	0.54446
Pennsylvania	0.75303	Arizona	0.66175	Alabama	0.51967
Ohio	0.75082	Texas	0.64957	South Carolina	0.51701
Wyoming	0.75065	Vermont	0.64577	Arkansas	0.49430
Colorado	0.74867	Maine	0.63633	Mississippi	0.44694

Table 3.1: Normalized Economic Data in US

of China, and Gansu, which are located in the northwest of China, are ranked as the poorest states.

3.3 Geographic Rank Markov for Inequality Analysis

3.3.1 Geographic Rank Markov and FMPT

Geographic rank Markov method is considered as a novel spatial analysis method for regional inequality analysis (Rey, 2013). Unlike most of the inequality measures which derive from entropy theory, geographic rank Markov method is under the framework of Markov

Shanghai	0.99672	Neimeng	0.27858	Shan3xi	0.19599
Beijing	0.76756	Jiling	0.26689	Heinan	0.19140
Tianjin	0.62596	Hebei	0.25034	Tibet	0.18201
Zhejiang	0.41692	Xinjiang	0.24761	Jiangxi	0.17813
Liaoning	0.40218	Hubei	0.23098	Anhui	0.17696
Guangdong	0.39938	Shan1xi	0.22829	Sichuan	0.17696
Jiangsu	0.39850	Qinghai	0.21571	Guangxi	0.16828
Shandong	0.31978	Ningxia	0.21398	Yunnan	0.16112
Fujian	0.31868	Chongqing	0.20137	Gansu	0.15872
Heilongjiang	0.29716	Hunan	0.19912	Guizhou	0.11429

Table 3.2: Normalized Economic Data in China

chain, as its name points out. Similar to the Markov method, it operates on time series data, classifies data values into several states, and statistically depicts the relationship between the data at two adjacent timestamps (status at each timestamp can be mapped to a state) using a probability transition matrix. More than a classic Markov, geographic rank Markov decomposes the data values to its maximum extent. In another word, every partition can hold only one value. This process can be accomplished by simply ranking all the values to measure and taking each rank as a state. Geographic rank Markov also turns around to place emphasis on the movement of ranks across different regions (in our application, the unit of regions is state/province). The probability transition matrix $P(g)$ between states can be estimated as:

$$P(g) = \begin{bmatrix} p(g)_{1,1} & p(g)_{1,2} & \dots & p(g)_{1,n} \\ p(g)_{2,1} & p(g)_{2,2} & \dots & p(g)_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p(g)_{n,1} & p(g)_{n,2} & \dots & p(g)_{n,n} \end{bmatrix} \tag{3.1}$$

where

$$p(r)_{i,j} = \frac{\sum_{t=0}^{T-1} \Omega_{i,j}^{t,t+1}}{\sum_k \sum_{t=0}^{T-1} \Omega_{i,k}^{t,t+1}} \tag{3.2}$$

with:

$$\Omega_{i,j}^{t,t+1} = \begin{cases} 1, & \text{if } r_{i,t} = r_{j,t+1} \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

$r_{i,t}$ is defined as the rank of a region i in time period t. $f(g)_{i,j}$ indicates the total number of times that a rank of region i switches to that of region j over all timestamps.

The Geographic Rank Probability Transition Matrix has a dimension of n by n. Each element in the matrix represents the possibility that one region gains the rank of another region over a period of time. A useful analysis tool, *First Mean Passage Time (FMPT)* matrix, has been developed by Rey (2013) to derive additional information from the transition matrix for further inequality analysis. The FMPT is the average length of time that the chain requires to pass from region i to region j. An estimate of the FMPT is shown as follows:

$$F = (1 - Z - E \cdot Z_{dg})(P(g)_{n \to \infty}^{-n})_{dg} \tag{3.4}$$

where

$$Z = (1 - P(g) + P(g)_{n \to \infty}^{n})^{-1} \tag{3.5}$$

$P(g)$ is the Geographic Rank Probability Transition Matrix shown in equation 3.1, E is the matrix with all entries 1, and A_{dg} represents the matrix that set all off-diagonal entries of matrix A equal to 0. Same as the transition matrix, the FMPT matrix of geographic rank Markov method shows the time costs of a region obtaining the other regions' ranks. Row i in the matrix indicates the possibilities that other regions gain the rank of region i. Column j indicates the possibilities that region j gains the ranks of other regions. To illustrate the information in FMPT intuitively, data in each column is extracted to create an individual FMPT choropleth map. From these maps, static and dynamic regions could be easily identified.

3.3.2 Rank-based Markov Test

As introduced by Rey (2013), there are two kinds of rank-Markov-based tests for space-time autocorrelation. One is the global version, which shows the overall space-time dynamics of a study area. This global test determines whether or not the transition of ranks are randomly distributed in space. Another one is the local version, which carries out the space-time autocorrelation within each region (subarea) of the entire study area.

The global statistics of space-time autocorrelation can be calculated using the equation below:

$$GR = \sum_{t=0}^{T} \sum_{i} \sum_{j} W_{i,j} \Omega_{i,j}^{t,t+1} \tag{3.6}$$

in which, $W_{i,j}$ is an element in the spatial weight matrix, indicating whether or not region i and region j are neighbors. If they are neighbors $W_{i,j} = 1$, if not $W_{i,j} = 0$. The spatial weight matrix can be obtained by utilizing *rook_from_shapefile* or *queen_from_shapefile* or some other similar functions in PySAL to process the shape file of our study area. $\Omega_{i,j}^{t,t+1}$ was defined in equation 3.3. Using equation 3.6 to test the global space-time autocorrelation, the null hypothesis is that there is no spatial clustering in the transition of ranks. We use the Monte-Carlo method (O'Sullivan & Unwin, 2003) to simulate the null hypothesis, and compare them with the real statistics to get the pseudo p-value. If the p-value is smaller than the significance level (0.05), we reject the null hypothesis and confirm that there is clustering in transition of ranks in space.

Besides, the local statistics of space-time autocorrelation takes the following forms:

$$GR_i = \sum_{t=0}^{T} \sum_{j} W_{i,j} \Omega_{i,j}^{t,t+1} \tag{3.7}$$

and

$$GR_{.,i} = \sum_{t=0}^{T} \sum_{j} W_{i,j} \Omega_{j,i}^{t,t+1} \tag{3.8}$$

The equation 3.7 is an origin-based local statistic, which means the space-time autocorrelation only considers the situation that other regions gain the rank of region i. On the other hand, equation 3.8 is the destination-based local statistic, which means the space-time autocorrelation only considers the situation that region j gains the rank of other regions. In practice, determining the direction of the movement of ranks is usually less important than determining whether or not there is a transition between regions. To capture this transition information, a new version of local statistic – the *"hybrid"* version of local statistic, $GR(*)_i$, is developed. Mathematically,

$$GR(*)_i = \sum_{t=0}^{T} \sum_{j} W_{i,j} \Omega(*)_{i,j}^{t,t+1} \tag{3.9}$$

where

$$\Omega(*)_{i,j}^{t,t+1} = \begin{cases} 1, & \text{if } r_{i,t} = r_{j,t+1} \text{ or } r_{j,t} = r_{i,t+1} \\ 0, & \text{otherwise} \end{cases} \tag{3.10}$$

Following similar procedure in a global test, the pseudo p-values of all regions can be obtained using equation 3.7, equation 3.8, or equation 3.9. The regions are then classified into two categories according to the p-values associated with them for generating the choropleth map. The first category contains regions with p-values below the significant level, indicating strong space-time autocorrelation. The rest regions are put into the second category. Because their p-values are above the significant level, there is no significant space-time clustering.

To handle all analysis tasks in this section, a geographic rank Markov module that helps generate geographic FMPT matrices and accomplish the rank-Markov-based tests are developed based on PySAL.

3.4 Empirical Result

The proposed method is applied to data in both U.S. and China to demonstrate the space-time autocorrelation on rank transitions. For U.S., per capita income at state level is used for local statistics. Figure 3.1 to figure 3.4 map out destination-based FMPT of all 48 states in contiguous U.S. Taking the first map for example, estimation of the time cost of rank transition from Connecticut to the others is displayed. Darker (blue) color on the maps means longer time and light color (more towards yellow) means the time cost is smaller. These maps are presented in the order of the rank of the states shown in 3.1. Several spatial clusters can be identified from the maps. One is located in southeast U.S. and it is composed by states that are relatively poorer. The other one is composed by California and Nevada. These two states are ranked high in terms of their richness, and they are marked as dark blue color on the second half of the maps, indicating the longer FMPT it would take for the poorer states to gain their ranks. The third cluster, composed by states near Great lakes, can also be identified from the second half of the maps, meaning that they have higher levels of development. These clusters indicate there are strong spatial autocorrelation on rank transitions. Because similar FMPT values or similar rank transition possibilities indicate the alike economic inequality degrees, it can be inferred that a spatial autocorrelation of similar economic inequality degree exists in U.S. as well. Taking a closer look at the maps, it can also be observed that the top 3 richest states – Connecticut, New York, and New Jersey present very similar patterns in terms of rank exchange with other states because they are at similar development levels. The same pattern can be observed amongst the poorest states. This clear pattern of rank transition across space and time reflects the fact that there exists some degree of regional inequality in the U.S, because otherwise, the rank transition maps will present a very diverse color distribution.

As a comparison to U.S., the same method is applied to data for Mainland China. The results of the 30 provinces, autonomies, and municipalities of destination-based FMPT are presented in figures 3.5 to 3.8. Again, we started from the FMPT maps of the richest regions. According to table 3.2, the top three are Shanghai, Beijing and Tianjin. Interestingly, the patterns presented on these three maps are almost identical. The only distinguishable elements are the legends of the maps, which indicate estimated years to obtain ranks from other regions. In compliance with the average transition time, the sequence of the three regions, conforms to their ranks in table 3.2. This high comparability is not a coincidence. Investigating the intermediate results which show the rank variation of all regions, we found that Shanghai, Beijing and Tianjin are very different from all the other regions.This may be the reason causing the relatively long estimated transition time and the consistency of the patterns. On the other hand, the patterns themselves still convey lucid information. There is a large cluster in the Southwest China that is composed by regions filled with the darkest color indicating the longest time to give out the rank. On the opposite, the light color regions do not typically cluster together, which is the situation that happens in U.S. On these maps, regions that are relatively easier to send out their ranks are all along or close to the east coast. These regions form a long strip, starting from Liaoning province in the North, and ending in Guangdong province in the South.

Utilizing the destination based FMPT maps for the richest or the poorest regions, the divergence in the trend of economic development during the whole time period can be conveniently identified. Spatial clusters of regions with relatively high or low economic development can be detected from the FMPT maps for both US and China. In China, the clustering pattern of the two kinds of regions is very clear and remains unchanged. This distribution demonstrates a strong relationship between economic development and geographical conditions in China. The highly developed regions are all closed to the coastline. On the contrary, the bottom ranked regions are almost all distributed in mountainous areas in Southwest China. However, such kind of distribution in US is less uniform. Although, the bottom ranked regions are still clustered geographically, the highly developed ones are scattered in various regions in the U.S, and are not necessarily along the coast. Additionally, the FMPT also reveals the economic dynamics over time. The averages of destination based FMPT of all regions in China and US are calculated and listed in table 3 and 4 respectively.

Sampling the averages of 3 regions in the top class, the middle class, and the bottom class, China demonstrates relatively low economic dynamics in general, especially in top ranked regions. Because of this, the rank transition is very limited over time between regions, causing higher degree of unequal development across different regions. The average FMPT in U.S. is lower though, indicating that regions are more equally developed than in China.

3.5 Summary

In this chapter, we first utilized destination-based FMPT maps to examine the interregional economic inequality and identify spatial clusters at different economic development levels to help understand the spatial patterns of an economic system. Then, by the comparison of the FMPT matrix and destination-based averages between China and U.S., the different levels of economic dynamics in two countries were revealed. This analysis can to some extent reflect the inequality situation across different subregions of a study area in comparison to using global approaches such as Theil's T, discussed in Chapter 2. Here, the economic dynamics, indicated by averaged FMPT, provided us a new perspective to study regional inequality. The larger the average of FMPT is, the higher chance that the investigated region is capable to keep its rank. This pattern could occur in both the richest and poorest regions, indicating overall inequality situation in the economic system. More importantly, leveraging the FMPT map the clusters composed by the adjacent states or provinces that share the similar economic inequality degrees can be conveniently identified, which help us to determine whether is there a spatial autocorrelation in a country or not. As a conclusion, FMPT is a proper index to investigate regional economic inequality and spatial distribution of such inequality. In next chapter, the spatial distribution of economic inequality degree will be further investigated by using a new developed method.

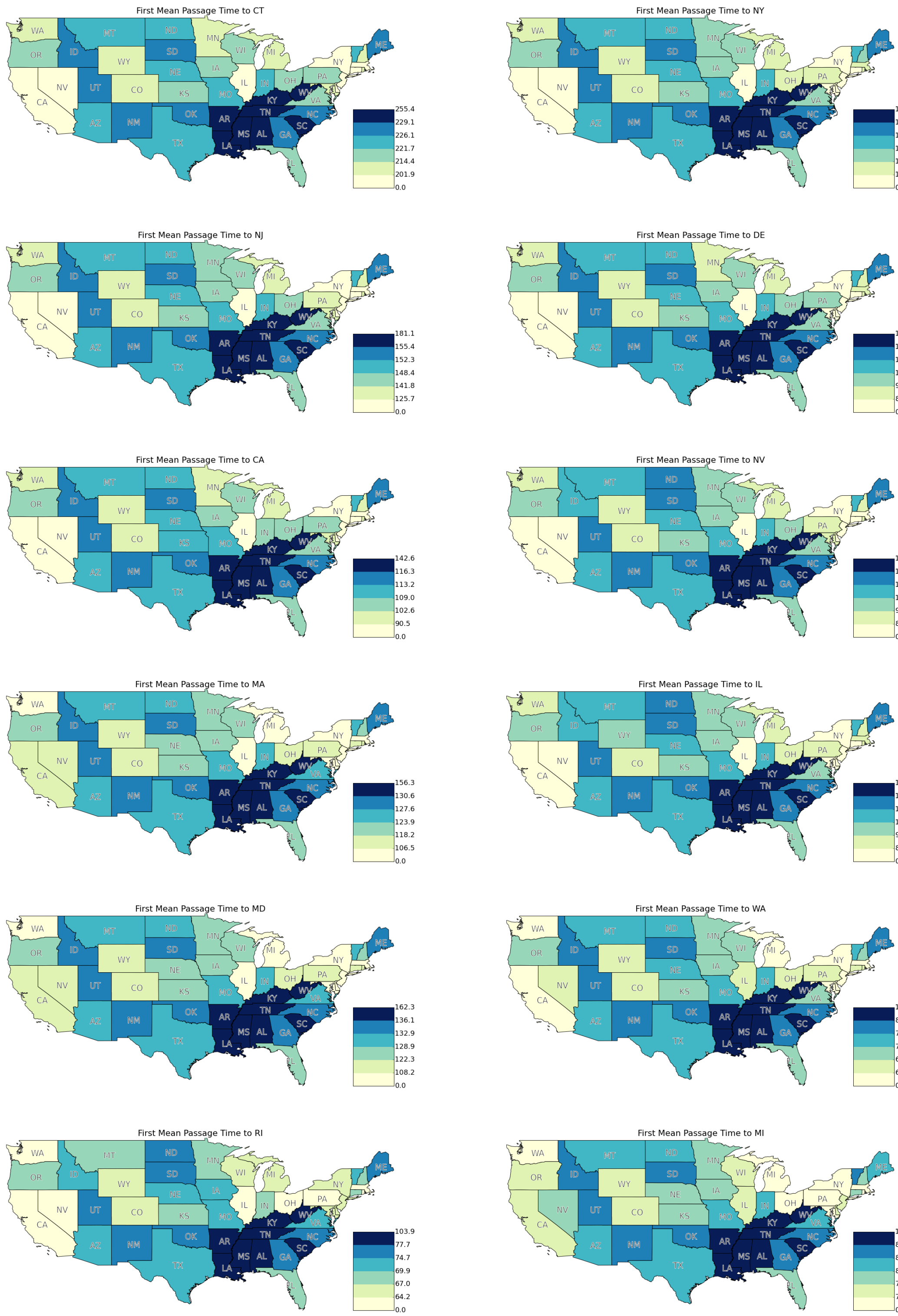

Figure 3.1: The First Mean Passage Time of US - 1

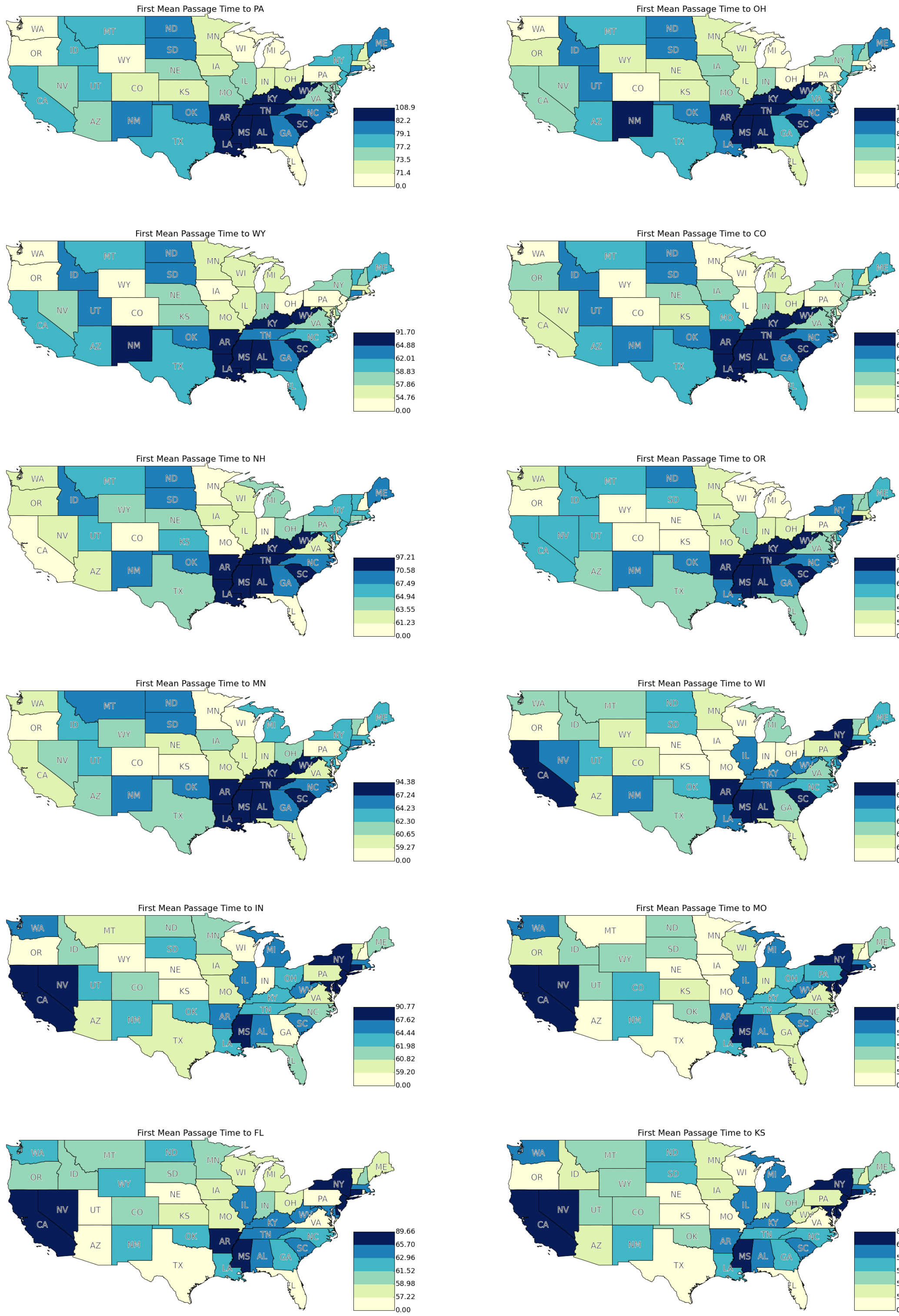

Figure 3.2: The First Mean Passage Time of US - 2

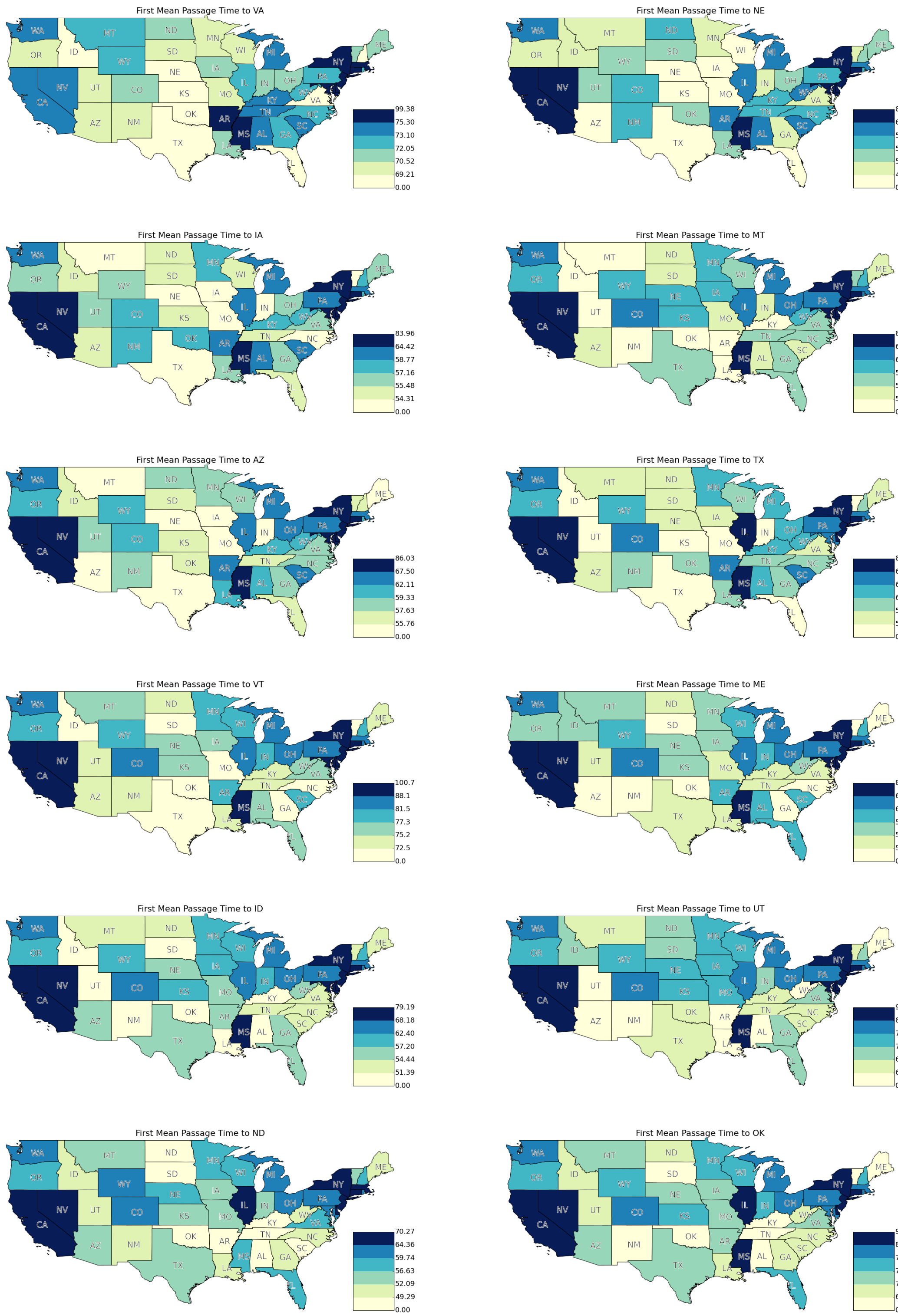

Figure 3.3: The First Mean Passage Time of US - 3

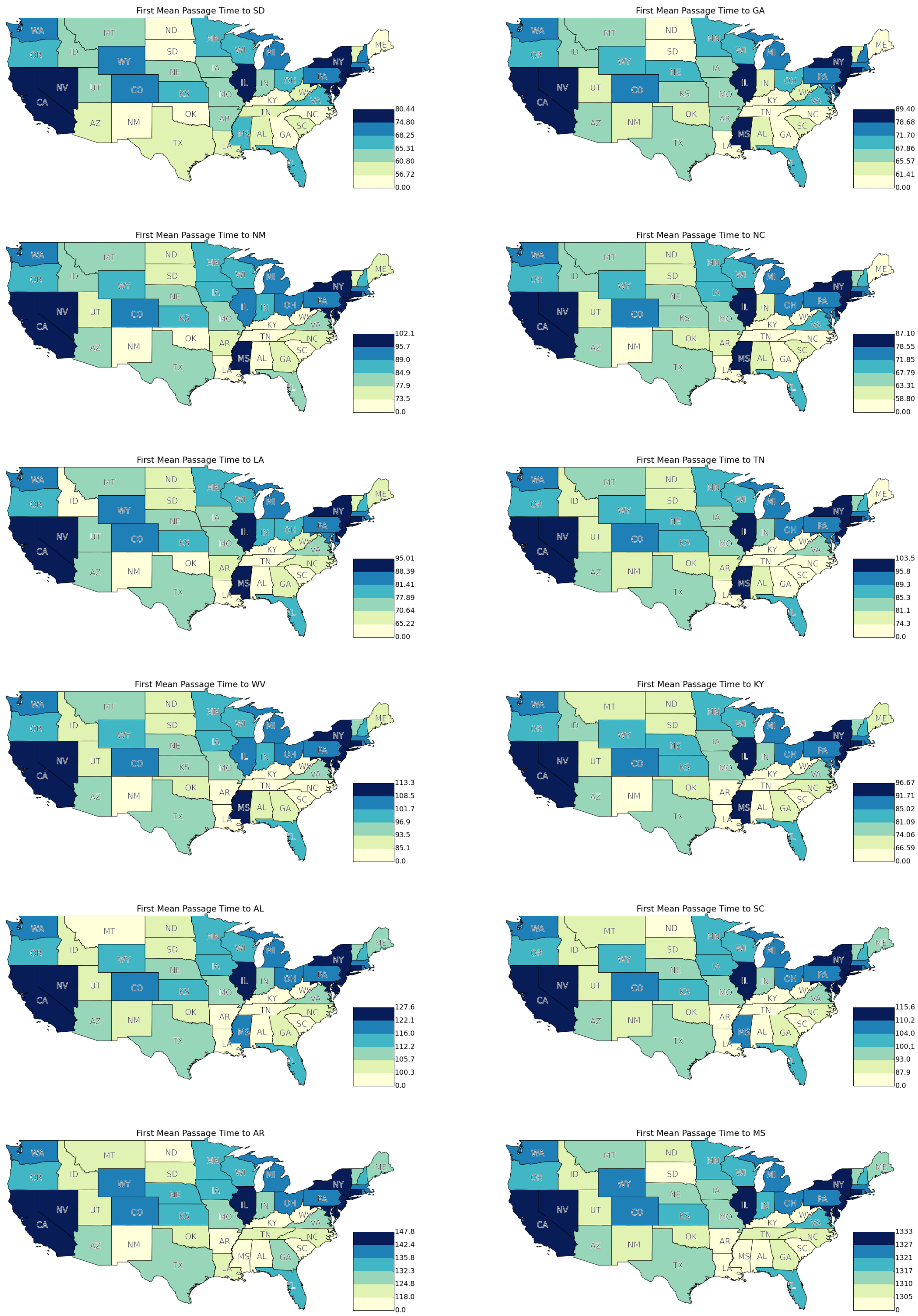

Figure 3.4: The First Mean Passage Time of US - 4

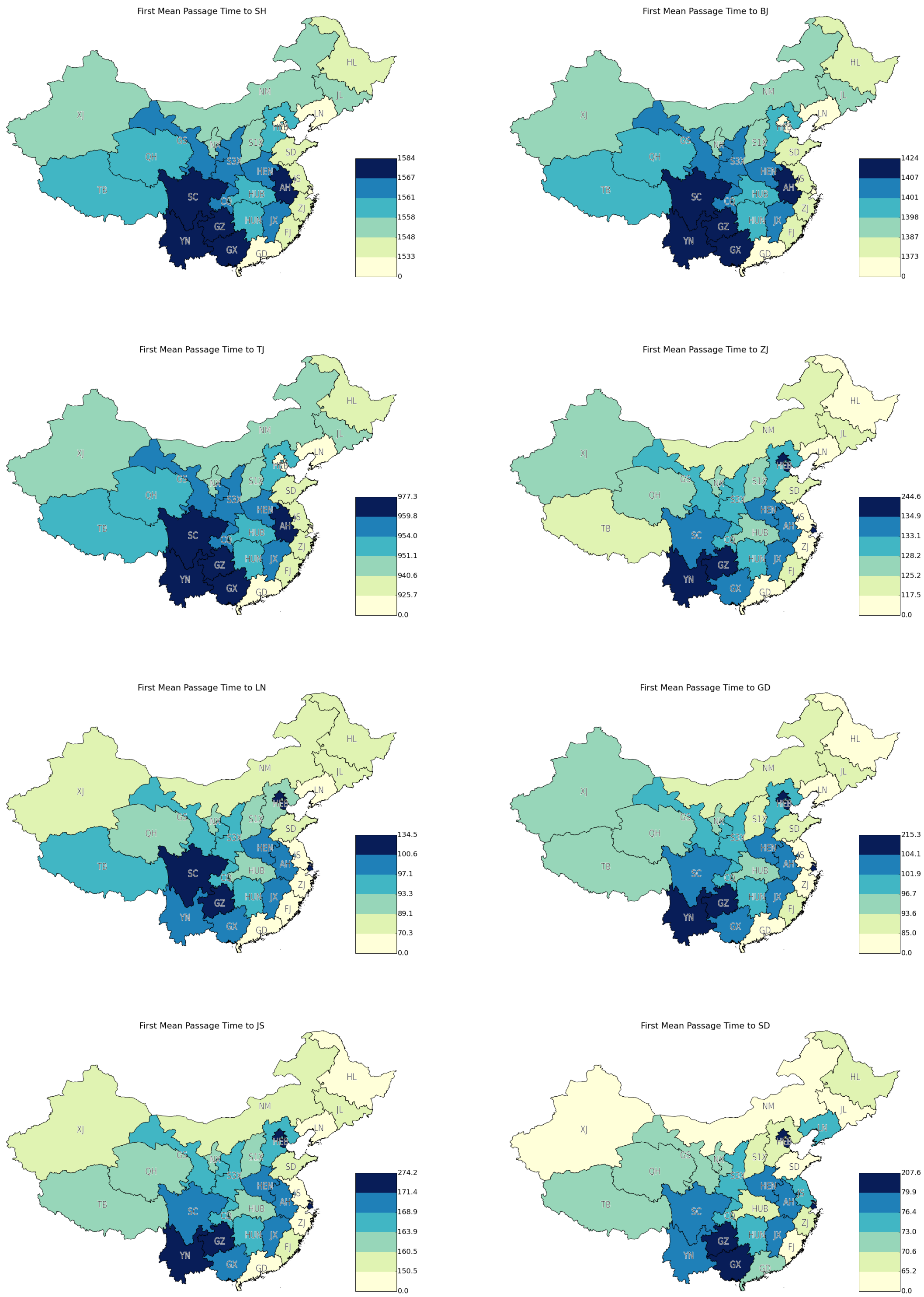

Figure 3.5: The First Mean Passage Time of China - 1

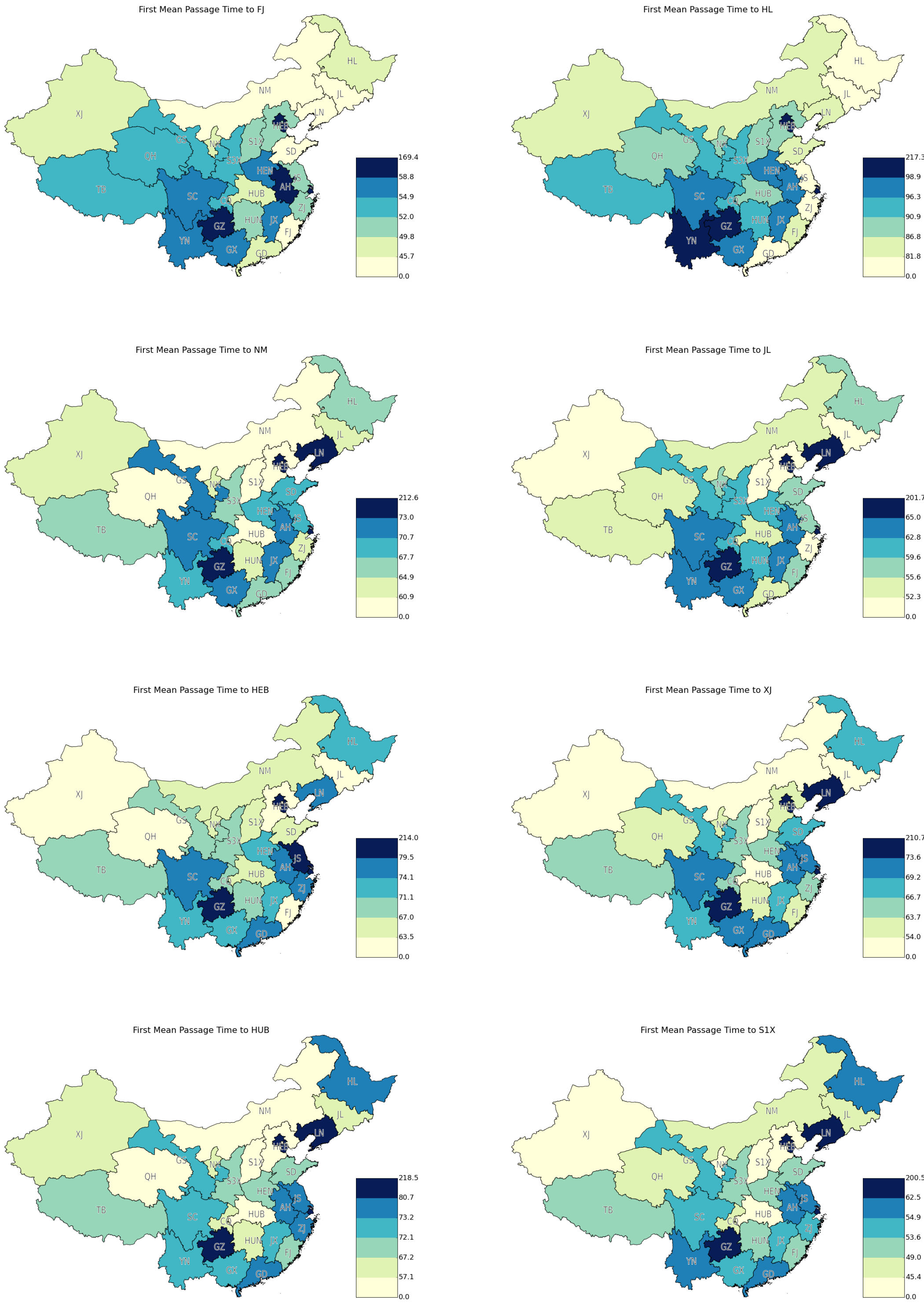

Figure 3.6: The First Mean Passage Time of China - 2

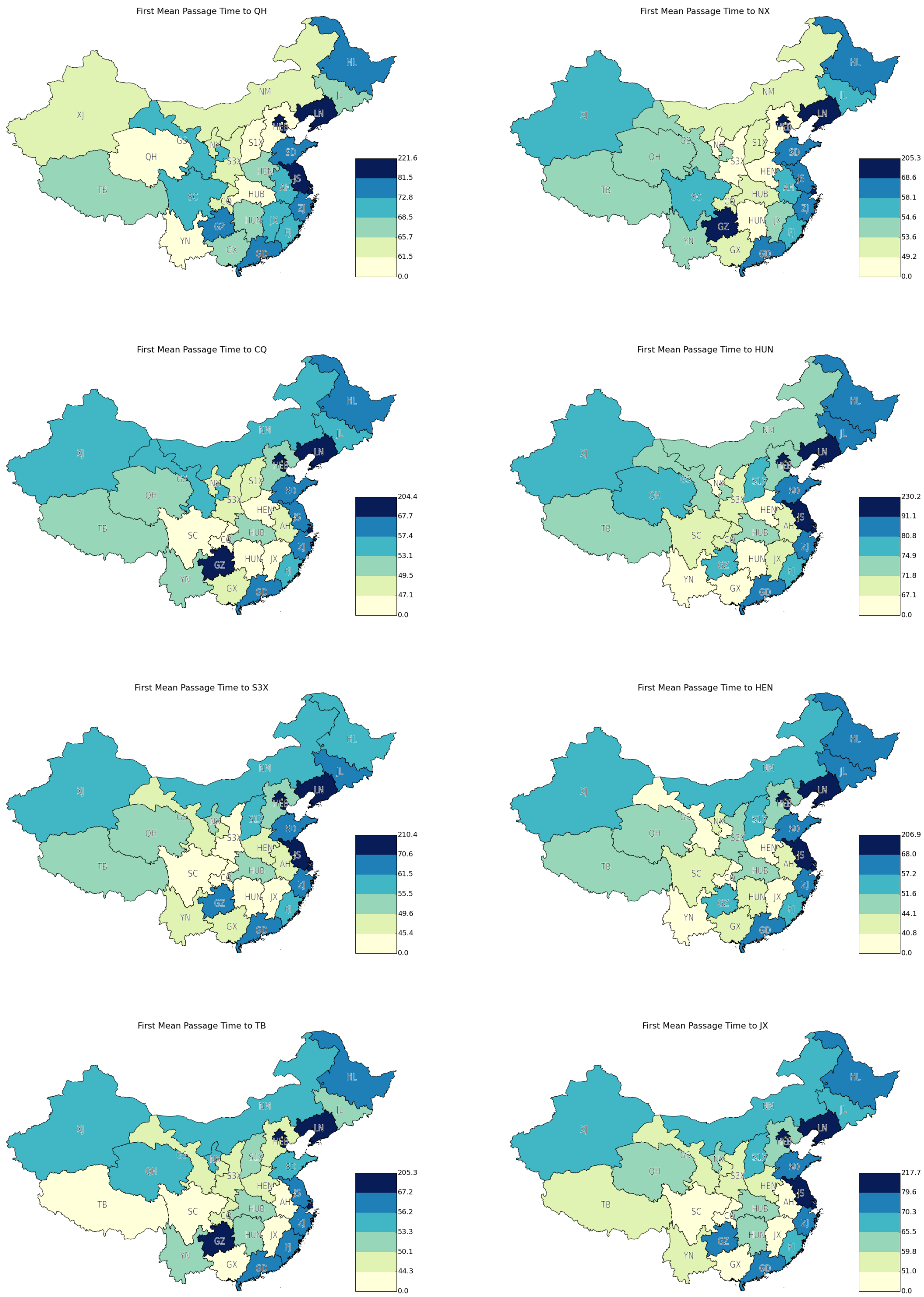

Figure 3.7: The First Mean Passage Time of China - 3

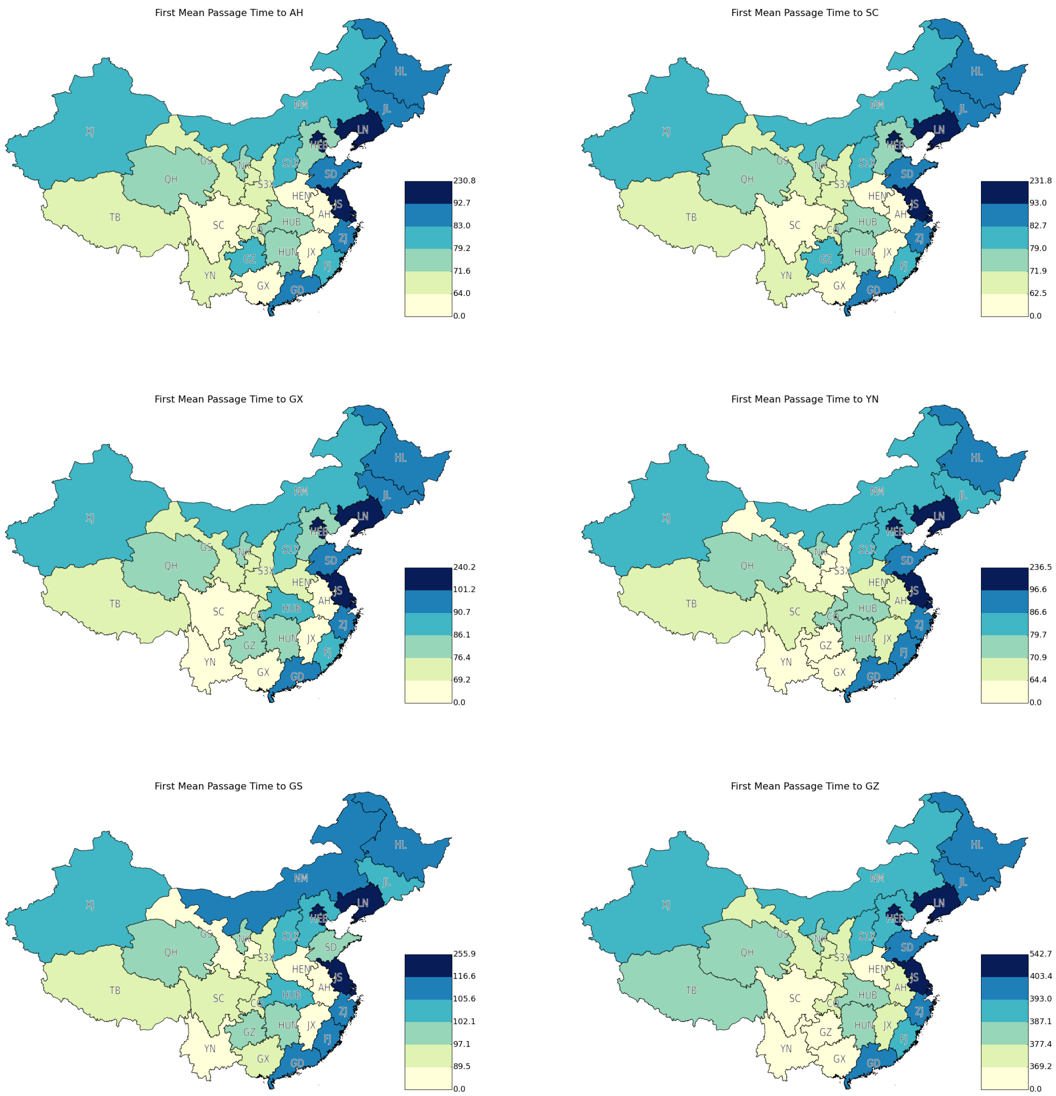

Figure 3.8: The First Mean Passage Time of China - 4

Chapter 4

EXPLORATION OF GRAPH INDICATOR BASED TEST

4.1 Introduction

Chapters 2 and 3 introduce two well-developed techniques to analyze regional inequality from both a global and local perspective. In this chapter, I plan to examine the feasibility of applying graph theory into inequality analysis, which is believed to be a brand new research direction. Inspired by the Rank-Markov-Based Test introduced by Rey (2013), I developed this analytical method by applying graph theory and named it as "Graph Indicator Based Test". The Rank-Markov-Bast Test examines the spatial autocorrelation of rank transitions. Comparatively, Graph Indicator Based Test provides more statistics and perspectives to view how ranks move, in order to investigate the regional economic inequality in more aspects. Although very few attempts has been made to integrate graph models into inequality analysis, there is a large literature that apply this theory to modeling problems raised in physical science, engineering, and economic analysis (Chen, 2012). Goldberg and Harrelson (2005) utilized two basic lemmas introduced in classic graph theory to build the algorithm to compute the shortest path. Similar to this, Derrible and Kennedy (2009) conducted network analysis of world subway system by calculating coverage, directness, connectivity and other basic graph(network) indicators. In these works, graph theory is proved to be highly effective in solving location-based problems.

As known, a graph can not only represent discrete points (as nodes), it can also represent the interactions between different points (through edges). Following this principle, if we model sub-regions (e.g. states in US, provinces in China) as nodes, and the transitions of ranks as edges, a graph can be constructed. I call this graph *"graph of rank path"* in the rest of the thesis. Once the graph model is constructed, we can adopt different graph indicators for data analysis and visualization. This strategy may offer new perspectives on regional inequality studies. Next section introduces important graph indicators to be investigated in this chapter.

4.2 Graph Indicators

Graph indicators are a set of tools that are able to generalize some characteristics of the whole or part of a graph. For convenience, in this thesis, indicators that capture characteristics of the entire graph are named as *global indicators*; those that apply on part of a graph, especially on a single node, are called *local indicators*. The two kinds of indicators are both useful. Global indicators provide a panorama view to investigate a graph of rank path. Local indicators, on the other hand, perform spatial analysis to investigate the interaction amongst nodes, which are location sensitive, such as a state or province. However, in most cases, the indicators cannot be interpreted independently. Instead, reference data are needed. In this chapter, permutation tests (also called randomization tests or re-randomization tests) will be adopted to generate reference data. A permutation test is a type of statistical significance test in which the distribution of the test statistic under the null hypothesis is obtained by calculating all possible values of the test statistic under rearrangements of the labels on the observed data points (Fisher, 1935). If a null hypothesis applied in a certain graph is rejected, it can be told that the graph carries opposite characters that the hypothesis represents. In this case, we can conclude that the graph indicator is suitable for use in the statistic test. Using this interpretation method, the following graph indicators will be examined:

Global indicators:

- Average Clustering (global clustering coefficient) (Wasserman & Faust, 1994): The local clustering of each node in a graph is the ratio between number of triangles that actually exist and that of all possible triangles in its neighborhood. The average clustering coefficient of the graph is the mean of all local clusterings (Schank & Wagner, 2004).

- Average Shortest Path Length (Dreyfus, 1969): the indicator calculates the ratio between the sum of shortest distance between every two nodes in graph A and the degree sum of a complete graph with same number of nodes in A.
- Diameter (Geodesic distance) (Bouttier, Di Francesco, & Guitter, 2003): The diameter is the maximum eccentricity, which is the maximum distance from a node to all other nodes in a graph.
- Normalized Total Degree: the ratio between sum of degrees of all nodes in a graph A, aka, the degree of the graph, and the degree sum of a complete graph with same number of nodes in A.

Local indicators:

- Average Neighbor Degree (Barrat, Barthelemy, Pastor-Satorras, & Vespignani, 2004): the average degree of the neighborhood of each node.
- Betweenness Centrality (Freeman, 1977, 1979, 1980): the indicator of a node v whose value is the sum of the fraction of all-pairs' shortest paths which pass through v.
- Closeness vitality (Koschützki et al., 2005): the indicator of a node whose value is the change in the sum of distances between all node pairs when excluding that node.
- Clustering Coefficient (P. W. Holland & Leinhardt, 1971; Watts & Strogatz, 1998): the indicator of a node whose value is the fraction of the number of all possible triangles through the actual number that the node possesses.

- Degree Centrality (Freeman, 1979): degree values of a node that are normalized by the maximum possible degrees of a complete graph, the node number of which is 1 less than that in the given graph.

- PageRank (Page, Brin, Motwani, & Winograd, 1999): indicator that ranks all nodes in a graph based on the structure of the incoming links. It was originally designed as an algorithm to rank web pages.

4.3 Methodology

In order to examine whether the rank mobility is spatially autocorrelated or not, two graph indicators are chosen based on the exploration among all the proposed indicators in former sections. The two chosen indicators have both global and local versions. The first one is degree, whose global version is *Normalized Total Degree*, and the local version is *Degree Centrality*. The other indicator is a measure of clustering. *Average Clustering* and *Clustering Coefficient* serve as the global and local indicators respectively.

4.3.1 Global Graph Indicator Test

The graph indicator test will be performed from both global and local perspectives. Global statistic focuses on answering whether the rank transitions will be more likely to take place between neighbor regions within a country or not. In contrast, the local one decomposes the overall statistic and presents the spatial correlation discovered by the graph indicators in each region.

The global graph indicator statistic for rank transition presents as follows:

$$GGIT = GI\_Func(A \circ W) \tag{4.1}$$

where:

$$A_{i,j} = \sum_{t=1}^{T-1} \Omega_{i,j}^{t,t+1} \tag{4.2}$$

with:

$$\Omega_{i,j}^{t,t+1} = \begin{cases} 1, & \text{if } r_{i,t} = r_{j,t+1} \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

In the equations above, $r_{i,t}$ defines the rank of region i in time period t. W is the weight matrix, where a certain element $W_{i,j}$ utilizes binary to indicate the relationship of adjacency between two regions, i and j. If they are neighbors, the element shows 1, otherwise 0. GI_Func represents the global graph indicator that will be applied to the test. Its input, the weight filtered matrix, $A \circ W$ is the *Hadamard product* of A and W. It is also known as the *element-wise product*, that is: $(A \circ B)_{ij} = A_{ij}B_{ij}$.

To perform the graph indicator test using equation 4.1, the null hypothesis assumes that rank migration will be independent of any spatial effect, no matter which indicator is applied to the statistics. In another word, when a region j gives its rank to another region i, region i does not need to be geographically dependent on region j. An example of such dependency can be geographic neighbors. Based on a spatial random permutation, the statistical significance can be evaluated by counting the number of results given out by equation 4.1 that is more extreme than the observed statistic under the null. The specific graph indicator will decide which side, above or under the null, should be sampled. For instance, when the indicator is *Normalized Total Degree* or *Average Clustering*, the lower tail of the distribution under the null will be examined. On the contrary, we count the part of distribution that is above the null while using the *Average Shortest Path Length*. Here the random permutation originates in a random data matrix by spatial shuffling, which is presented as follows:

$$D(R) = I(R) \cdot D \tag{4.4}$$

with:

$$I(R)_{i,j} \in \{0, 1\}, \forall i, j \tag{4.5}$$

and

$$\sum_{q}^{n} I(R)_{i,q} = 1, \forall i \tag{4.6}$$

and

$$\sum_{p}^{n} I(R)_{p,j} = 1, \forall j \tag{4.7}$$

where D is the original data matrix with the dimensions of m by n, indicating that there are m time periods and n regions. $I(R)$ is an n by n random binary matrix, with each row and column has and only has one cell assigned value 1.

4.3.2 *Local Graph Indicator Test*

The overall spatial dynamics on rank transitions are expected to be demonstrated by the global graph indicator statistic. The local test decomposes the global one and provides more detailed statistic on each region. The local version of the graph indicator test is carried out as:

$$GIT_i = GI\_Func_i(A \circ W) \tag{4.8}$$

Similar to the global statistic, here it takes the same function argument, the weight filtered matrix. GI_Func_i represents the local graph indicator that is applied to calculate the statistic of region i. Since the local statistic is the decomposition of the global one, their relationship can be typically represented in the following forms:

$$GGIT = \sum_{i=0}^{n} GIT_i \tag{4.9}$$

or

$$GGIT = \frac{\sum_{i=0}^{n} GIT_i}{n} \tag{4.10}$$

4.4 Empirical Results

To examine whether or not the graph indicators help to discover the economic inequality, the *Normalized Total Degree* and *Degree Centrality* are applied to the data on per capital GDP of the 30 provinces, municipalities, and autonomous regions in Mainland China, and the per capital income of the 48 states in the contiguous US.

4.4.1 *China Results of Degree Statistics*

The degree statistics are examined in 4 different perspectives. We applied the Global statistic to the entire data, the temporal decomposed data, the one using spatial decomposition respectively, and applied the local indicator to the entire dataset.

The Global Graph Indicator Test for Rank Path

To obtain overall degree of centrality of the data during the entire time period, the global graph indicator test is applied, with *Normalized Total Degree* to investigate the total times of rank transitions between adjacent regions to show its spatial autocorrelation. Based on 999 partially random permutations, figure 4.1 displays the real value and the distribution of those permutation values. The X axis indicates the Normalized Total Degree, while the Y axis tells the frequency that values are laid on a certain range. The red vertical line indicates the observed value calculated from data, separating the graph into two parts - a slightly larger left side, and smaller right side (by area). Since the greater value of transition times between neighbors means higher spatial autocorrelation, this entire data shows an almost random pattern, and also a bit positive autocorrelation.

The above analysis provides summary statistics. In order to gain a more detailed spatial autocorrelation and contrastive views, we decompose the entire dataset both spatially and temporally, and then apply the graph indicator test to them respectively.

Data decomposition

First, the data is decomposed by time. The time period is approximately evenly divided into 17 years from 1978 to 1994, and 18 years from 1995 to 2012. The results are shown in figure

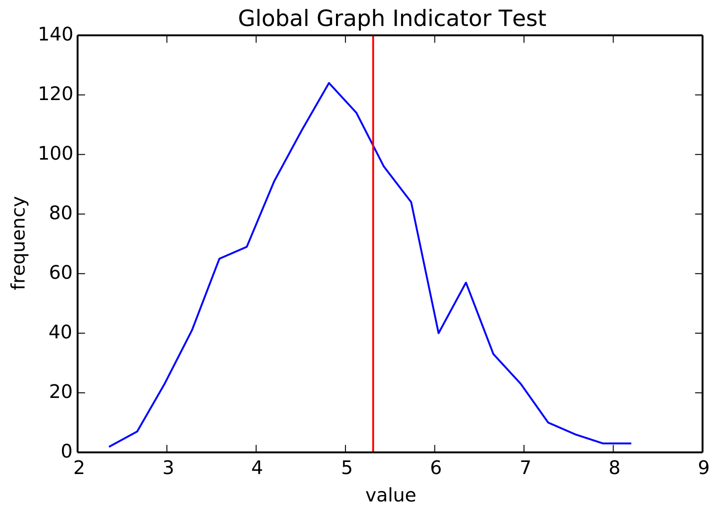

Figure 4.1: Global Graph Indicator Test for China

4.2. Comparing with figure 4.1, the observed data values both decreased, because of the decomposition. However, the more noticeable fact is the variation from the first half time period to the second. The first one resembles the graph presented in figure 4.1 on its pattern. When it comes to the second half time period, the entire graph moves to left side, showing a decrease in the transition time. Moreover, relative to the permutation distribution, the red vertical line goes further to a lower value. Synthesizing these results, we can obtain that rank transition shows weaker spatial autocorrelation as time passes by and an increase in the economic development.

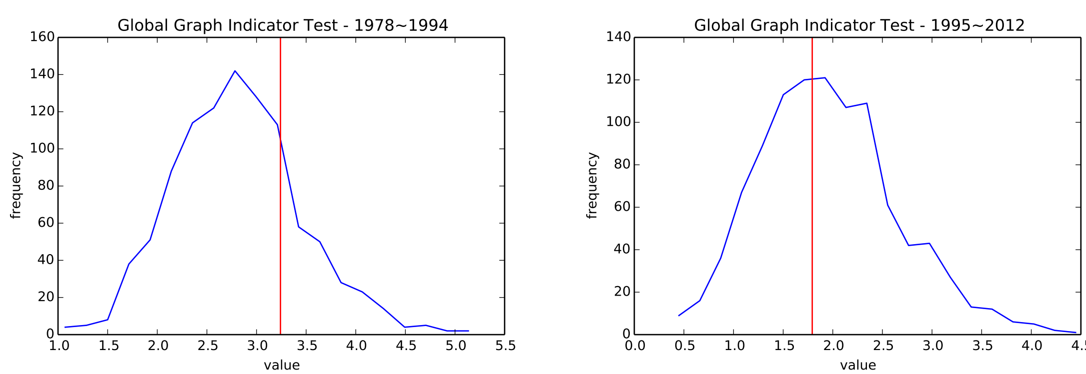

Figure 4.2: Global Graph Indicator Test for China, Decomposed by Time

Second, we utilized the decomposition to classify regions into two classes. One contains the poorest 15 regions, according to the accumulative per capital GDP during the entire time period. On the contrary, the other one consists of the richest 16 regions. Figure 4.3 presents the results of the tests applied to the two classes. The two graphs in figure 4.3 share similar patterns of permutation distribution and position of the real value. They both indicate a near-random tendency to transit the ranks to neighbors. In addition, some negative spatial autocorrelation is displayed. Additionally, same as the former case, as the economy develops, from the poorest regions to the richest regions, the total number of the transition time is dramatically dropped.

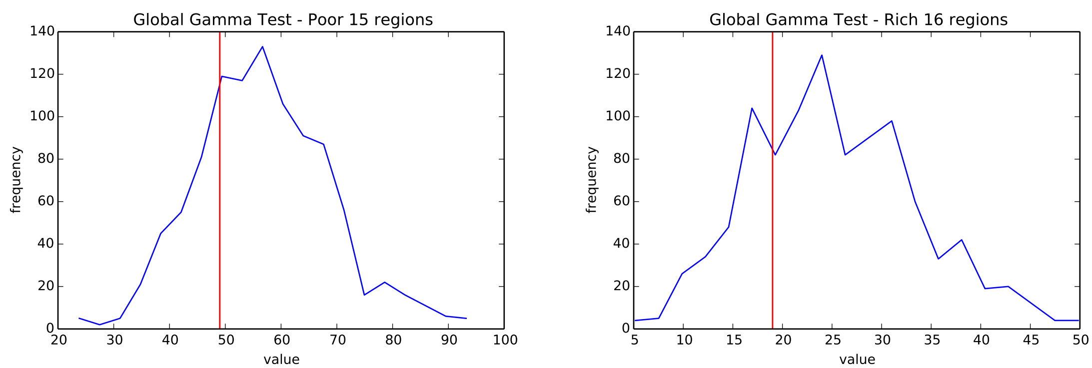

Figure 4.3: Global Graph Indicator Test for China, Decomposed by Region

The Local graph indicator test for rank path

To examine the level of spatial autocorrelation in each region, I conduct further analysis using local graph indicator test. Similar to the global test, 999 permutations were generated. From the result of figure 4.1, the data shows no clear tendency to transfer its ranks to a neighboring region or a non-neighbor one with a global view. Therefore, both of the upper-tailed test and lower-tailed test are performed with the local statistics. The upper-tailed test rejects the null hypothesis that ranks related to a certain region are independent in space, to conclude the significance of that region, indicating the positive spatial autocorrelation of rank transition. However, this local version upper-tailed test for china produced a black map, which means no region shows the strong spatial autocorrelation. On the other hand, the significant regions (red ones) shown in figure 4.4 reject the null hypothesis in lower-tailed test, indicating that these regions have negative spatial autocorrelation of rank transition. More specifically, they are more likely to transfer their ranks to their non-neighbor regions or keep the ranks by themselves. In the figure, there are three coastal regions. From north to south, they are Liaoning, Shanghai and Guangdong. They are all surrounded by the regions that are relatively less developed in the economy. Hence there is a smaller chance for them to transfer the ranks to their neighbors. This could be the reason why they show the negative spatial autocorrelation.

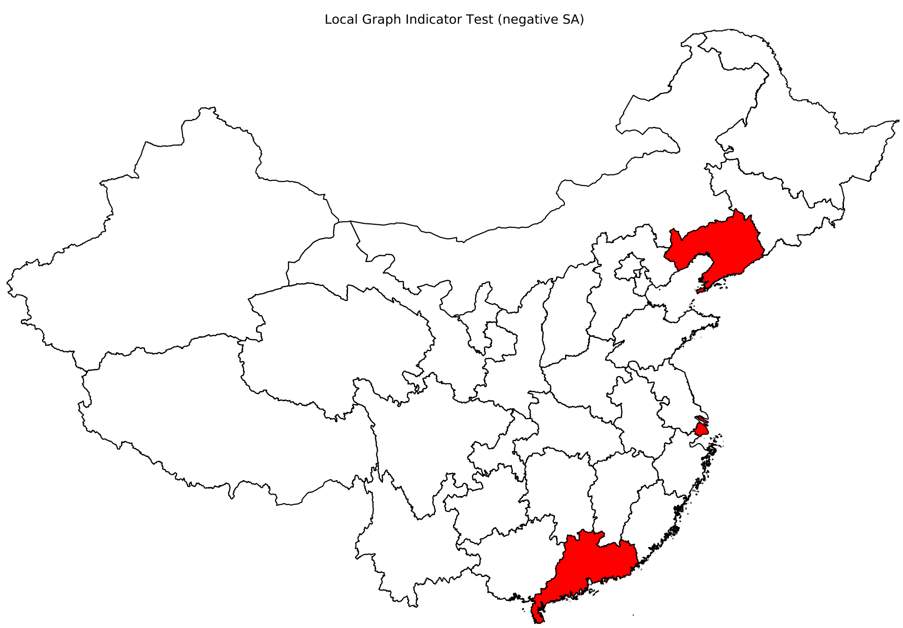

Figure 4.4: Local Graph Indicator Test for China

4.4.2 US Results of Degree Statistics

The US per capital income is investigated by "degree" statistics using similar procedure for the China analysis. Global test, temporal decomposition on global statistic and local test are performed on the U.S. dataset in sequence.

The Global Graph Indicator Test for Rank Path

Figure 4.5 displays the result of the global test using Normalized Total Degree as the graph indicator on US economic data. Different from the result of China, the overall Normalized Total Degree statistic presents conspicuous positive spatial autocorrelation. The p-value of this test is 0.039, which is less than 0.05. Therefore, the null hypothesis can be rejected. Generally, the rank path graph has a much higher Normalized Total Degree than the graph generated from spatially shuffled data, meaning the ranks are more likely to migrate to its neighbors.

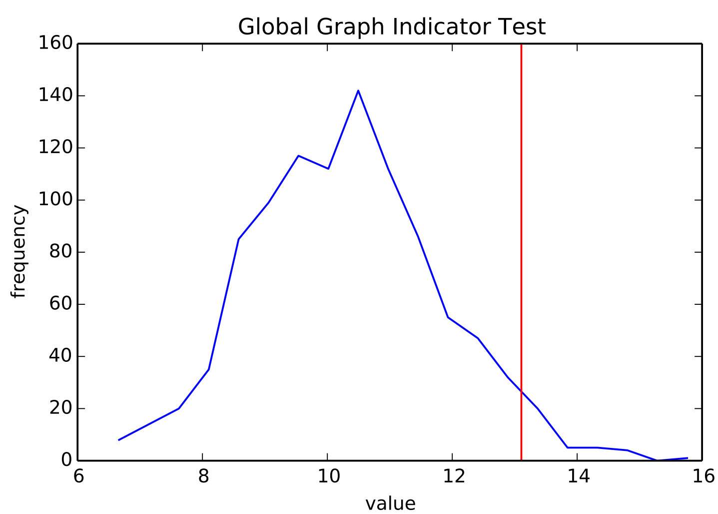

Figure 4.5: Global Graph Indicator Test for US

As the counterpart of the summary statistics given above, the temporal decomposition on the global indicator test is also performed on the US data.

Temporal data decomposition

There are 81 time frames in US per capital income time series data. Each time frame represents one year, lasting from 1929 to 2009. The data is decomposed into the first 40 years from 1929 to 1968, and the following 41 years from 1969 to 2009. Figure 4.6 displays the results. It can be observed that the first 40 years show stronger positive spatial autocorrelation than the non-decomposed test. Its p-value is 0.001. On the contrary, the result of next 41 years starting from 1969 is totally different from that using data from 1929-1968 and the non-decomposed data. The vertical line indicating the graph statistic value moves much to the left side of the figure. Rather than showing a positive spatial autocorrelation, the rank mobility demonstrates more negative spatial dependency on data from 1969 to 2009. Generally, there is a higher possibility for ranks to move between distant regions than neighbors during this time period.

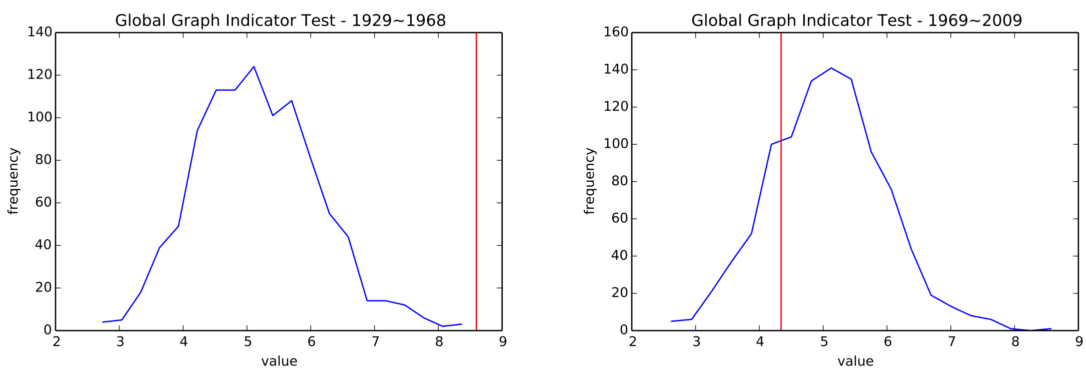

Figure 4.6: Global Graph Indicator Test for US, Decomposed by Time

The Local graph indicator test for rank path

We further propose the null hypotheses, “rank transition is positively spatially autocorrelated” and ”rank transition is negatively spatially autocorrelated”, and apply these nulls on the local graph indicator test. The results of the two hypotheses tests are shown in figure 4.7 and figure 4.8, respectively. In figure 4.7, there are 3 hot spots located in the northwest, southeast, and center of the contiguous US, indicating the tendency of rank transition across nearby regions. In the northwest and center hot spots, Montana, Nebraska, Idaho, and Kansas rank in the middle class of economic development over the whole country. Tennessee and North Carolina in the southeast hotspots have relatively low per capital income over time. The clusters in the southeast are surrounded by the regions with similar economic development levels. Thus, these clusters and its neighbors could be more likely to exchange ranks, resulting in a positive spatial dependency on degree centrality. In the map shown in figure 4.8, the regions that demonstrate negative spatial autocorrelation on degree centrality are highlighted. Indiana and Vermont are in the same situation. The relatively low per capital income values are found in these regions over time. However, they are surrounded by the highly developed areas. Therefore it is extremely difficult for them to migrate ranks to their adjacent regions. Similarly, the per capital income of Illinois and Florida is much higher than their neighbors, which also leads to low possibility of rank transition between these two states and their neighbors.

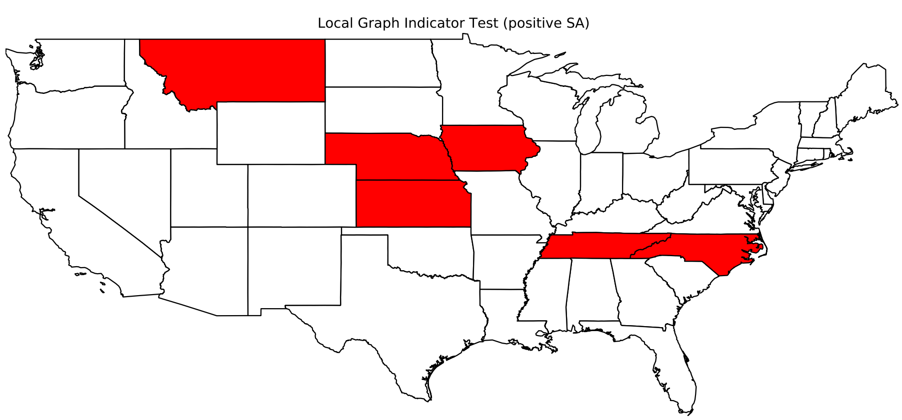

Figure 4.7: Local Graph Indicator Test for US - Positive Spatial Autocorrelation Filled

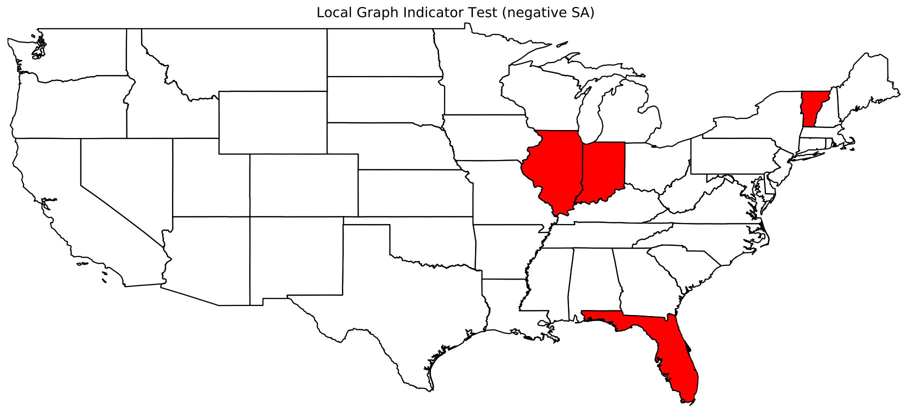

Figure 4.8: Local Graph Indicator Test for US - Negative Spatial Autocorrelation Filled

4.5 Discussion

In this chapter, the Normalized Total Degree, Degree Centrality, and Graph Indicator test are applied to the China and US economic data respectively. From both global and local perspective we examined the spatial autocorrelation on degree statistics, which reflect the tendency of rank migration between subregions (i.e. US states) and its neighbors, and further statistically depict the clustering condition of economic inequality. The global tests show that there is a conspicuous positive spatial dependency shown in US, but much less clearer pattern in China. That means rank mitigation across states in the US is easier than that in China. In the local graph indicator test, states/provinces that demonstrate negative spatial autocorrelation are identified, for both US and China, to help us justify the local economic inequality within each country. I also propose to conduct analysis on data decomposed by space and time, allowing in-depth analysis on economic inequality in different time period and subareas. From this point of view, the graph indicators differs substantially from Theil's T presented in Chapter 2 and the FMPT index discussed in Chapter 3 by providing new perspectives in understanding and analyzing regional and subregional economic inequality across space and time. As a summary, graph indicator statistic can complement existing analytical methods and contribute to the local economic inequality studies.

Chapter 5

CONCLUSION

In this thesis, I exploit a set of spatial analysis methods for use in the regional inequality analysis. Chapters 2, 3, and 4 deal with measuring inequality using the global inequality measurement, geographic rank Markov method, and rank path respectively. Each of these chapters illustrated an approach to assess inequality that aim to achieve three research objectives described in Chapter 1. The first objective is to develop an overview of inequality situation for both China and U.S. The second objective is to reveal the detail of how inequality is distributed in the whole country. The third objective is to explore how graph theory can help us to evaluate spatial inequality, and extend our knowledge about inequality at a subregion and sub-temporal scale.

The results leveraging Theil's T discovered the overall economic inequality variation in the history of US and China. An intensive fluctuation pattern was demonstrated in China's curve of Theil's T. Starting at a high inequality status, excepting for a slightly increasing trend in the middle of the whole time period, the overall inequality in China drops down sharply. On the contrary, the inequality trend of US presents a relatively stable pattern. The Theil's T in US gradually decreases to a very low level in the first half time period, and holds the level in the second half. Theil's T statistic is helpful to reveal the distinguishability of gross inequality for multiple study areas. It is not only utilizable when comparing the average inequality, or the inequality in some time frame, but also provides the trend of inequality change, when time series data is available. Though some detailed situation of inequality is still under veil, especially in terms of spatial aspects, Theil's T helps us to begin the inequality analysis by a top-down processing, and obtain some general conclusion. China suffers a much more higher economic inequality over the same time period of US, and suffers from more dynamics in inequality alteration than the US.

Utilizing the destination based First Mean Passage Time (FMPT) statistic, an FMPT map is generated for each state or province in both US and China. With FMPT maps, the clusters at different economic development levels are identified, which can be helpful to understand the spatial pattern of an economic system. Then, by the comparison of the FMPT matrix and destination based averages in China and US, we can further investigate the disagreement of economic dynamics existed in the two countries. Unlike Theil's T statistic, which can reveal the inequality situation by a series of operations on the economic data in a certain year. Here, the economic dynamics, indicated by averaged FMPT, provided us another perspective to view regional inequality. The larger the average of FMPT is, the higher chance that the investigated region is capable to keep its rank. A large value could appear in both the richest and the poorest regions, indicating the overall inequality situation in the economic system. As a conclusion, in terms of investigating economic inequality from a geographical perspective, FMPT is an appropriate statistic to test and compare the intra-regional and inter-regional economic development of our research countries.

Starting with the exploration of several different graph indicators, I tried to identify if and which graph indicators are applicable to inequality analysis. Because of their utility, the Normalized Total Degree as the global statistic, and Degree Centrality as the local one are applied to the China and US economic data respectively. The spatial autocorrelation is investigated from both global and local perspectives by the degree statistics, which reflect the tendency of rank migration between a certain region and its neighbors. The results demonstrate that there is a conspicuous positive spatial dependency in the US, whereas a much weaker dependency is found in China over the entire time period. Local analysis-wise, degree centrality is applied to examine the local economic inequality in both countries. Distinguished from Theil's T and FMPT statistics, graph indicator highlights the subareas that suffer the economic inequality. Moreover, the temporal and spatial decomposition are introduced, and applied to the graph indicator test. More detailed results are generated for further analysis of the economic inequality in different time periods and subareas. Graph indicator statistic provides a new perspective in investigating economic inequality at a local scale.

Leveraging the Geographical Rank Markov and Graph Theory to analyze the economic inequality are both relatively novel methodologies, especially the latter one. The works in this thesis demonstrate great potentials for applying both methods in the regional inequality analysis. This I believe is an important contribution to the literature in terms of both methodology development and inequality applications. While attempting to apply the graph indicators to inequality analysis, only the degree centrality and normalize total degree are used as graph indicators to perform the exploration and experiments. Some other indicator pairs are remain unexplored. For instance, clustering coefficient and average clustering would be a promising indicator pair that helps to reveal a more substantial rank exchanging amongst neighbor regions, which could be an interesting exploration topic in future research. Besides, many methods are rejected because of the limitation of those graph indicators and the graph indicator test. For example, the reason why the Average Shortest Path Length was not applicable is because there may be some isolated nodes in the weight matrix filtered graph, which is not permitted in the Average Shortest Path Length indicator. Therefore how to integrate more graph indicators into current methodology framework could be an interesting direction for future work. Modifying the filtered graph, meanwhile reserving most of its original characters might be a solution. But the detailed method and its effectiveness needs further exploration.

REFERENCES

Aghion, P., Caroli, E., & Garcia-Penalosa, C. (1999). Inequality and economic growth: the perspective of the new growth theories. *Journal of Economic literature*, 1615–1660.

Anselin, L. (2001). Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics*, *83*(3), 705–710.

Arrow, K. J., Bowles, S., & Durlauf, S. N. (2000). *Meritocracy and economic inequality*. Princeton University Press.

Autor, D. H., Katz, L. F., & Kearney, M. S. (2008). Trends in US wage inequality: revising the revisionists. *The Review of economics and statistics*, *90*(2), 300–323.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–3752.

Becker, G. S. & Murphy, K. M. (2007). The upside of income inequality. *The American*, *1*(4).

Bouttier, J., Di Francesco, P., & Guitter, E. (2003). Geodesic distance in planar graphs. *Nuclear Physics B*, *663*(3), 535–567.

Carter, C. C. (2013). The price of inequality: how today's divided society endangers our future. *Journal of Real Estate Literature*, *21*(1), 141.

Castells Quintana, D. & Royuela Mora, V. (2012). Unemployment and long-run economic growth: the role of income inequality and urbanization. *Investigaciones Regionales, 2012, vol. 24, p. 153-173*.

Chen, W.-K. (2012). *Applied graph theory*. Elsevier.

Coburn, D. (2000). Income inequality, social cohesion and the health status of populations: the role of neo-liberalism. *Social Science & Medicine*, *51*(1), 135–146.

Cowell, F. A. (2011, January 1). *Measuring inequality*. LSE perspectives in economic analysis. Oxford University Press.

Deaton, A. (2001). *Health, inequality, and economic development*. National bureau of economic research.

Derrible, S. & Kennedy, C. (2009). Network analysis of world subway systems using updated graph theory. *Transportation Research Record: Journal of the Transportation Research Board*, *2112*(1), 17–25.

Dreyfus, S. E. (1969). An appraisal of some shortest-path algorithms. *Operations research*, *17*(3), 395–412.

Fisher, R. A. (1935). The design of experiments.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, *1*(3), 215–239.

Freeman, L. C. (1980). The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity*, *14*(4), 585–592.

Goldberg, A. V. & Harrelson, C. (2005). Computing the shortest path: a search meets graph theory. In *Proceedings of the sixteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 156–165). Society for Industrial and Applied Mathematics.

Holland, P. W. & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*.

Holland, T. G., Peterson, G. D., & Gonzalez, A. (2009). A cross-national analysis of how economic inequality predicts biodiversity loss. *Conservation Biology*, *23*(5), 1304–1313.

Katz, L. F. et al. (1999). Changes in the wage structure and earnings inequality. *Handbook of labor economics*, *3*, 1463–1555.

Keller, K. R. (2010). How can education policy improve income distribution: an empirical analysis of education stages and measures on income inequality. *The Journal of Developing Areas*, *43*(2), 51–77.

Kelly, M. (2000). Inequality and crime. *Review of economics and Statistics*, *82*(4), 530–539.

Kets, W., Iyengar, G., Sethi, R., & Bowles, S. (2011). Inequality and network structure. *Games and Economic Behavior*, *73*(1), 215–226.

Korpi, W. (2000). Faces of inequality: gender, class, and patterns of inequalities in different types of welfare states. *Social Politics: international studies in gender, state & society*, *7*(2), 127–191.

Koschützki, D., Lehmann, K. A., Peeters, L., Richter, S., Tenfelde-Podehl, D., & Zlotowski, O. (2005). Centrality indices. In *Network analysis* (pp. 16–61). Springer.

Kuznets, S. (1955). Economic growth and income inequality. *The American economic review*, 1–28.

Meng, X., Shen, K., & Xue, S. (2013). Economic reform, education expansion, and earnings inequality for urban males in china, 1988–2009. *Journal of Comparative Economics*, *41*(1), 227–244.

Navarro, V. et al. (2007). *Neoliberalism, globalization, and inequalities: consequences for health and quality of life*. Citeseer.

Novotnỳ, J. (2007). On the measurement of regional inequality: does spatial dimension of income inequality matter? *The Annals of Regional Science*, *41*(3), 563–580.

Ostby, G., Nordas, R., & Rod, J. K. (2009). Regional inequalities and civil conflict in sub-saharan africa. *International Studies Quarterly*, *53*(2), 301–324.

O'Sullivan, D. & Unwin, D. (2003). *Geographic information analysis*. J. Wiley.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: bringing order to the web.

Palestini, A. & Pignataro, G. (2015). A graph-based approach to inequality assessment.

Panayotou, T. et al. (2000). *Economic growth and the environment*. Center for International Development at Harvard University.

Paredes, D., Iturra, V., & Lufin, M. (2014). A spatial decomposition of income inequality in chile. *Regional Studies*, 1–19.

Raudenbush, S. & Kasim, R. (1998). Cognitive skill and economic inequality: findings from the national adult literacy survey. *Harvard Educational Review*, *68*(1), 33–80.

Rey, S. J. (2001). Spatial empirics for economic growth and convergence. *Geographical Analysis*, *33*(3), 195–214.

Rey, S. J. (2004). Spatial analysis of regional income inequality. *Spatially Integrated Social Science*, 280–299.

Rey, S. J. (2013, August 2). Rank-based markov chains for regional income distribution dynamics.

Rey, S. J. (2015). Bells in space the spatial dynamics of us interpersonal and interregional income inequality. *International Regional Science Review*, 0160017615614899.

Rey, S. J. & Anselin, L. (2007). PySAL: a python library of spatial analytical methods. The Review of Regional Studies, *37*(1), 5–27.

Rey, S. J. & Gutiérrez, M. L. S. (2015). Comparative spatial inequality dynamics: the case of mexico and the united states. *Applied Geography*.

Rey, S. J. & Janikas, M. V. (2005). Regional convergence, inequality, and space. *Journal of Economic Geography*, *5*(2), 155–176.

Rey, S. J. & Sastré-Gutiérrez, M. (2010). Interregional inequality dynamics in mexico. *Spatial Economic Analysis*, *5*(3), 277–298.

Schank, T. & Wagner, D. (2004). *Approximating clustering-coefficient and transitivity*. Universität Karlsruhe, Fakultät für Informatik.

Seguino, S. (2000). Gender inequality and economic growth: a cross-country analysis. *World Development*, *28*(7), 1211–1230.

Smeeding, T. M. (2005). Public policy, economic inequality, and poverty: the united states in comparative perspective*. *Social Science Quarterly*, *86*, 955–983.

Stiglitz, J. E. (2002). *Globalization and its discontents*. New York.

Theil, H. (1967, January 1). *Economics and information theory*. Studies in mathematical and managerial economics. North-Holland Pub. Co.

Thorbecke, E. & Charumilind, C. (2002). Economic inequality and its socioeconomic impact. *World Development*, *30*(9), 1477–1495.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic geography*, 234–240.

Von Braun, J., Díaz-Bonilla, E., Pinstrup-Andersen, P., et al. (2008). *Globalization of food and agriculture and the poor*. New Dehli: Oxford University Press.

Wasserman, S. & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge university press.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *nature*, *393*(6684), 440–442.

Wei, Y. D. & Ye, X. (2009). Beyond convergence: space, scale, and regional inequality in china. *Tijdschrift voor Economische en Sociale Geografie*, *100*(1), 59–80.

Williamson, J. G. (1965). Regional inequality and the process of national development: a description of the patterns. *Economic development and cultural change*, 1–84.

Yamamoto, D. (2007). Scales of regional income disparities in the USA, 1955–2003. *Journal of Economic Geography*, lbm044.

Ye, X. & Wei, Y. D. (2005). Geospatial analysis of regional development in china: the case of zhejiang province and the wenzhou model. *Eurasian Geography and Economics*, *46*(6), 445–464.

Yildirim, J., Öcal, N., & Özyildirim, S. (2009). Income inequality and economic convergence in turkey: a spatial effect analysis. *International Regional Science Review*, *32*(2), 221–254.

Zhang, Q. & Wang, Y. (2007). Using concentration index to study changes in socio-economic inequality of overweight among US adolescents between 1971 and 2002. *International journal of epidemiology*, *36*(4), 916–925.