geographical analysis

Geographical Analysis (2020) **52**, 537–557

Special Issue

A Visual Analytics System for Space–Time Dynamics of Regional Income Distributions Utilizing Animated Flow Maps and Rank-based Markov Chains

Sergio Rey, Su Yeon Han, Wei Kang, Elijah Knaap, Renan Xavier Cortes

Center for Geospatial Sciences, University of California, Riverside, CA

Regional income convergence and divergence has been an active field of research for more than 20 years, and research papers in this field are still being produced at a prodigious rate. Despite their importance for the study of dynamics of income distribution, interactive visualization tools revealing spatiotemporal dimensions of the income data have been sparsely developed. This study introduces a visual analytics system for the space–time analysis of income dynamics. We use state-level US income data from 1929 to 2009 to demonstrate the visual analytics system and its utility for exploring similar data. The system consists of two modules, visualization and analytics. The visualization module, a Web-based front-end called Rank-Path Visualizer (RPV), draws inspiration from the cartographic technique of flow mapping, originally developed by Tobler and embodied in his canonical Flow Mapper application.

Introduction

Tobler's innovative work on the software package Flow Mapper (Tobler 1987, 2004) integrated two strands of his previous research: the study of movement (Tobler 1981b) and the science of cartography. Because flows tie together pairs of places at two different points in time, they provide a powerful mechanism to integrate space and time. Tobler's contributions have inspired the widespread adoption of flow mapping across many scientific domains as well as a vibrant strand of research addressing the technical issues in flow mapping (Stephen and Jenny 2017; Yang et al. 2019).

One of the prominent motivating applications Tobler developed was the representation of the flow of federal monies within the continental United States (Tobler 1981a). Federal expenditures in, and tax receipts from, each of the 48 lower states were considered to develop a net transfer value for each state, with a positive (negative) sign indicating the state that received more

Correspondence: Sergio Rey, Center for Geospatial Sciences, University of California, Riverside, CA
e-mail: sergio.rey@ucr.edu

Submitted: May 24, 2019. Revised version accepted: May 5, 2020.

doi: 10.1111/gean.12239
© 2020 The Ohio State University

(less) in expenditures than it contributed in taxes. Interstate flows were then imputed by setting these net transfer value in a potential field, and then aggregating the flows that cross state boundaries. The final US map renders the visible movement of US Federal Reserve through arrows whose width is proportional to the average flow that connects the centroids of adjacent states.

In this article, we draw inspiration from, and extend, Tobler's flow mapping of federal funds framework to provide a new visual-analytical approach to study spatial income distribution dynamics. We explore the question of how the recasting of flows to the case of income distribution dynamics may reveal novel insights on the evolution of spatial income inequality. Our approach is based on the concept of a Rank Path that portrays the "migration" of ranks of the income distribution over time. The concept of a rank path emerges from work in the distributional dynamics literature that employs Markov chain frameworks.

The article makes three contributions. First, we develop an analytical framework that combines the notion of a rank path with that of a flow map to provide a new approach to distributional dynamics. Second, while the notion of a rank path has been introduced by Rey (2014b), to date its impact has been limited to a proof-of-concept. By embedding the rank path in the visual analytic framework, we feel the rank path concept may now be applicable across a wide number of domains dealing with longitudinal spatial data. Finally, we develop the Rank-Path Visualizer (RPV), an interactive web-based visual analytic system, and illustrate its functionality with the case of US income distribution dynamics.

The article is organized as follows. Section "Literature review" provides a discussion of previous work on flow mapping and spatial income distribution dynamics research to set the context for our approach. This is followed by Section "Methods," where we introduce the notion of rank-based Markov chains and the associated rank path, and detail the design and implementation of the RPV. We provide an illustration of the framework applied to US per capita incomes over the 1929–2009 period in Section "Illustration: Spatial dynamics of US income inequality". Section "Conclusion" closes the article with key findings and directions for future research.

Literature review

Mapping flows through time and space

Distributive flow maps represent "the movement of commodities, people, or ideas between geographic regions" (Slocum et al. 2008, p. 360). They have straight or arc-shaped flow lines connecting each pair of origin and destination, where line width denotes the magnitude of the flow and arrows at the end of lines represent the directions of flows. Flow mapping is one of the oldest known examples of cartographic data visualization (Tufte 2001). Ten short years following the publication of John Snow's infamous cholera map, Minard (1862) "gave quantity as well as direction to the data measures located on the world map in his portrayal of the 1864 exports of French wine" (Tufte 2001, p. 24) and the first flow map was born. For more than a century thereafter, flow maps were used as powerful data visualizations representing the movement of people, commodities, and services between various geographic regions around the world. Prior to the advent of computerized cartography, however, such maps were highly labor intensive, requiring both considerable artistic skills and meticulous manual measurement to design maps that were both aesthetically pleasing and quantitatively accurate. The digital era, naturally, upended this tradition. Indeed, at the outset of the nascent personal computing revolution, early Big Data pioneer Tobler (1987) imagined a computational framework that would introduce the concepts of scalability and reproducibility into the process of flow map design. After developing the first

proof-of-concept on early consumer hardware in 1987, Tobler went on to design software capable of ingesting any arbitrary origin–destination matrix and generating a flow map encoding the spatial interactions embedded in the data.

Almost immediately the value of flow mapping as a visual analytics system was demonstrated by Clark and Koloutsou-Vakakis (1992) who used Flow Mapper to investigate more thoroughly the concept of migration vector fields, the motivating theoretical construct behind Tobler's initial design. Focusing on Dutch migration patterns, their flow maps revealed "an important change in the system of flows between 1975 and 1985. In 1975, there is a general movement toward the south and east with dominant flows out of the metropolitan areas of Amsterdam and the Hague. In 1985, the net exchanges are not only smaller but the only substantial out-migration is the one from Amsterdam to the neighboring region (Zuid-Ijsselmeer Polders)" (Clark and Koloutsou-Vakakis 1992, p.114). While these same patterns could have been found through a thorough examination of the input matrix, the speed and intuition with which the flow map presents this complex result obviates the need for tabular exploration. Following this initial work, many others have improved and transformed flow mapping in visual analytics to reveal the unique complexity in spatial interaction data to great effect; notable examples include Boyandin, Bertini, and Lalanne (2010) who extend flow mapping to the temporal case, Chen et al. (2015) who apply flow mapping techniques to social media data, and Wood, Dykes, and Slingsby (2010) who use a unique treemap visualization to avoid aggregation or generalization.

However, despite continuous development in the literature and constant attention from a group of analytical cartographers, flow mapping lags behind many other geovisual techniques in adoption by professionals and the wider public. As Rae 2011, p. 791) describes, this lack of adoption outside academia is troublesome, since flow mapping is the "visual approach, based on careful spatial data analysis and analytical decision making, that best conveys [the relationships between] functional spatial structures. It is possible to convert the complexity of large spatial interaction matrices into simple geovisualisations, so the hope now is that others, and particularly national statistical agencies, will move towards this kind of approach in their dissemination activities."

Nearly 20 years after its initial prototype, Flow Mapper was implemented in modern computer architecture and re-released as "Tobler's Flow Mapper" (Tobler 2004). As computers have grown more powerful, scholars and software developers have enhanced Tobler's original concept by adding more advanced functionality and more appealing visual effects. For example, Jenny et al. (2018) provides a set of aesthetic principles for designing better flow maps, and Demšar and Virrantaus 2010, p. 1538) extend classic flow mapping using modern graphics cards to permit visualization of "3D space-time density of trajectories". Buchin, Speckmann, and Verbeek (2011) and Stephen and Jenny (2017) offer algorithmic solutions for better flow map layouts, the latter of which is available as a web-based application. [1] Indeed, today there is a variety of flow mapping software packages available for both desktop and web-based environments. Other web-based tools include work by Rae (2009) visualizing residential mobility in large migration data and the Hubway Trip Explorer that visualizes human movements by quantitating bicycle rides (Woodruff 2019). Even the US Census bureau provides its own proprietary flow mapping tool for visualizing its "County-to-County Migration Flows" data set.[2]

For desktop environments, there are flow mapping extensions to popular GIS packages as well as standalone platforms (Kim et al. 2012). Guo (), for example, developed a decision-support tool for modeling pandemics that includes flow-mapping capability, including a new functionality for large data sets that gains efficiency by aggregating data from neighboring geographic

regions. Boyandin, Bertini, and Lalanne (2010) also developed a desktop-based flow mapping tool called JFlowMap that facilitates "dynamic queries for filtering out flows by their quantities or their length" and the exploration of the temporal changes in flows.

As technology continues to progress, a new generation of flow mapping software that leverages movement and animation has begun to appear. These versions continue the natural evolution of flow mapping in the digital era to elucidate the temporal *change* in flows. For example, Koblin (2006) created animations of air traffic flow. CartoDB provides a function called Torque that enables creating the animated flow representing the temporally changing movements of objects (CARTO 2017). By using Torque, the flows of naval ships were animated to show the yearly changing movements (Rogers 2012). Ho *et al.* (2011) developed a web-enabled flow mapping tool containing a function to play animation of flows across time and various charts to visualize additional information other than flows. Han, Clarke, and Tsou (2017) also developed a web-based animated flow mapping tool, verifying its visual effectiveness through a human subject opinion survey.

While it is clear that technology for constructing and displaying flow maps has improved continuously since Tobler's first implementation, novel applications of flow mapping as a technique have failed to keep pace. Inherently, flow mapping is a technique for visualizing the space–time dynamics of *things* that move between places. Traditionally, these *things* have represented physical or material objects, such as humans, commodities, or money. But the notion of visualizing movements between places over time can also be abstracted to immaterial objects.

One such abstraction is the notion of a rank within a particular distribution. More specifically, consider an ascending ranking of US state per capita incomes at one point in time. The rank assigned to a particular state in that moment in time tells us where in the income distribution the state would be located, with states having low ranks in the left tail of the distribution, and rich states with high ranks in the right tail of the distribution. Over time, as states move within the income distribution, their ranks would change. From the perspective of a particular rank, however, its "geographic" location would also change. Flow mapping using this abstraction could visualize how a particular rank passes through geographic units over time. Thus, although classic flow mapping can show migration patterns by representing the quantities of people moving between two areas, adopting the rank path abstraction could visualize, for instance, the "movement" of the most populous city in a country over the course of a particular time period. In the following sections, we describe a new visualization tool: the RPV that expands upon the animated flow mapping tools mentioned above, but adopts this novel approach designed to visualize the spatio-temporal movements of rank data across regions over time.

The concept of rank-based geographic mobility was first discussed by Rey (2014a) who developed a methodology for investigating whether geographic partitioning or spatial relationships exist within the rank distribution of a particular variable or a portion of its value distribution. In this article, we develop a novel method for visualizing such mobility in an interactive platform that lets researchers explore any part of the rank distribution or geographic scope of their choosing. This technique is useful because selecting and visualizing an arbitrary rank allows scholars to examine whether certain parts of the rank distribution are more volatile than others, or whether spatial units of an arbitrary rank move conterminously with their neighbors or get trapped in space (Bloom, Canning, and Sevilla 2003). This can be informative for applications such as regional economic policy development.

If we observe that rank 20 moves from city A to city B to city C, and the three cities are proximate to one another, then we can infer that some variety of spatial process is taking place

that keeps rank 20 isolated in a particular region. Different portions of the rank distribution may be influenced by different varieties of spatial effects (for instance, spatial poverty traps versus urbanization agglomeration). Our interactive visual analytics system lets researchers explore these dynamics at will, to investigate how the relationships between geography and rank dynamism vary for different portions of an attribute distribution (or a particular portion of the study area).

GIS applications of Markov chains

Formally, a Markov chain is a sequence of discrete random variables $X_0, X_1, X_2, \ldots, X_t$ where the probability of arriving at a given Markov state x at time $t + 1$ depends only on the immediate past value X_t:

$$P(X_{t+1} = x \mid X_0, X_1, \ldots, X_t) = P(X_{t+1} = x \mid X_t) \tag{1}$$

Since this mathematical framework can model the dynamics of different types of phenomena, the applications in a geographical context are quite vast including land use changes (Muller and Middleton 1994), spatial criminology (Rey, Mack, and Koschinsky 2012), migrations (Pan and Nagurney 1994), and weather (Jordan and Talkner 2000), among others.

Interactive visualization of Markov Chain analysis in GIS can be dated back to (Logsdon, Bell, and Westerlund, 1996) where a Graphic User Interface is used to present transition probability matrices for land type data using ARC/INFO Macro Language. The IDRISI-GIS Analysis Tools of TerrSet (Eastman 2015) also provides a suite of tools for predictive land cover change modeling that include the Land Change Modeler extension for ArcGIS, (Johnson 2009) which uses Markov Chain analysis to project the expected quantity of change in a competitive land allocation model to determine scenarios for a specified future date. Another data visualization tool that takes advantage of Markov Chain analysis in GIS is CrimeVis[3] (Cortes et al. 2017) developed using the `R` language's Shiny package (Chang et al. 2018). This web application permits a detailed analysis of how Markovian transition models can help reveal temporal and spatiotemporal crime dynamics in the city of Rio Grande do Sul state, BR. The variable under analysis is a dichotomous dummy variable indicating whether the municipality had at least one occurrence of a particular crime. Following Rey, Mack, and Koschinsky (2012), the application applies Markov chains to analyze the probabilities that a municipality transitions across Markov states during the initial period and end period; it also facilitates analyses of joint spatiotemporal analysis by stratifying the sample in cities that had (or not) contiguous municipalities with a crime. Several maps, probabilities matrices, odds ratios, and hypothesis tests for temporal and spatial homogeneity are provided in an interactive interface.

Despite their widespread application in the study of regional income inequality dynamics,[4] the use of visualization for Markov chains has been very limited. Instead, the emphasis has been on the estimation of transition probabilities and related derived summary measures. Some exceptions to this are early work on the package Space-Time Analysis of Regional Systems (STARS) (Rey and Janikas 2006) that has many features to deal with transitional dynamics of distributional attributes through the use of classic Markov and spatial Markov techniques. One of the innovations of STARS was interactive transition tables that allow the user to select specific cells of the estimated Markov transition matrix, which, in turn, highlight the areas on a choropleth map that made transitions associated with the selected cell.

Since the original release of STARS, a number of new analytics for spatial Markov chains have been introduced including rank-based Markov chains (Rey 2014b), joint and conditional

tests for spatial dependence in Markov chains (Rey, Kang, and Wolf 2016), and local indicators of mobility association (Rey 2016). Each of these new measures can be coupled with new visual analytics, and in this article we describe one particular development in this regard, the integration of rank paths from rank-based Markov chains, together with the Tobler's flow mapping framework.

Methods

We have developed a web-based visual analytics system for exploring the movements of the ranks of data across regions over time. It consists of two modules: RPV and Geographic Rank Markov (GRM). These two modules together provide insight into the spatiotemporal movements of ranks of data that are difficult to reveal using conventional visualization methods. To illustrate this framework, we use annual (1929–2009) per capita income data from the United States; however, the methods and visualization tools developed in this study can be used for any spatiotemporal data.

GRM and rank paths

The application of classic Markov chains to the study of regional income distribution dynamics has been criticized for the ad-hoc nature of discretization methods used to estimate the transition probability matrices (Magrini 1999). As these matrices are used to draw inferences about dynamics, convergence, and inequality, the sensitivity of the probability estimates to the choice of the quantization method has been seen as a central concern.

Rank-based Markov chains were suggested as a way to deal with the issues surrounding the arguably ad-hoc nature of discretization used in the study of regional income distributions. Rather than discretizing the usually continuous variable of interest (e.g., income) where the resulting Markov state space is contingent on the choice of discretizing scheme, numeric ranks are assigned beginning with 1 for the smallest value to n for the largest value given n observations.

Two forms of rank-based Markov chains were introduced in Rey (2014b): *Full Rank Markov* and *GRM*.[5] The first form defines the state of the discrete Markov chain from (1) on the *ranks* of the variable under consideration. In the case of regional convergence research, the focus is on the ranks of regional per capita income. More formally, define $\omega_{i,l,m}^{t,t+1} = 1$ if a geographical unit i held rank l in period t and rank m in period $t + 1$, 0, otherwise. From this, form the $(n \times n)$ Full Rank Transition Matrix $P(r)$:

$$P(r) = \begin{bmatrix} p(r)_{1,1} & p(r)_{1,2} & \cdots & p(r)_{1,n} \\ p(r)_{2,1} & p(r)_{2,2} & \cdots & p(r)_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ p(r)_{n,1} & p(r)_{n,2} & \cdots & p(r)_{n,n} \end{bmatrix} \tag{2}$$

with elements $p(r)_{l,m} = \frac{f_{l,m}}{f_{l,\cdot}}$ where: $f_{l,m} = \sum_{i=1}^{n} \sum_{t=1}^{T-1} \omega_{i,l,m}^{t,t+1}$ and $f_{l,\cdot} = \sum_{q=1}^{n} f_{l,q}$.

The second form of a rank-based Markov, the GRM, is the dual of the Full Rank form. Rather than modeling the transitional dynamics of individual regions moving across discrete income classes where the choice of the discretization is essential but unresolved, the *GRM* focuses on the transitional dynamics of income ranks moving across regions over time. It does so by embedding the states of the Markov chain in geographical space. Letting $\omega_{r,i,j}^{t,t+1} = 1$ if rank r was

in region i in period t and region j in period $t+1$, 0 otherwise, we can similarly form the $(n \times n)$ Geographic Rank Transition matrix $P(g)$:

$$P(g) = \begin{bmatrix} p(g)_{1,1} & p(g)_{1,2} & \cdots & p(g)_{1,n} \\ p(g)_{2,1} & p(g)_{2,2} & \cdots & p(g)_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ p(g)_{n,1} & p(g)_{n,2} & \cdots & p(g)_{n,n} \end{bmatrix} \tag{3}$$

with elements $p(g)_{i,j} = \frac{f_{i,j}}{f_{i,\cdot}}$ where: $f_{i,j} = \sum_{r=1}^{n} \sum_{t=1}^{T-1} \omega_{r,i,j}^{t,t+1}$ and $f_{i,\cdot} = \sum_{q=1}^{n} f_{i,q}$.

Two mobility measures can be constructed based on the transition probability matrix $P(g)$ (equation (3)), shedding light on the interactions between regions in a dynamic setting. The first, Mean Sojourn time, defined for a given region is the expected time any income rank spent in this region before leaving it[6]:

$$s_i = \frac{1}{1 - p(g)_{i,i}}. \tag{4}$$

The larger the Mean Sojourn time is for a region, the less likely this region is to interact with other regions in terms of exchanging ranks in the income distribution. The second mobility measure, mean first passage time (MFPT), is defined for a pair of regions. For example, the MFPT from region i to region j is the expected time any income rank takes to move from i to j for the first time. The mean first passage time matrix is estimated as[7]:

$$MFPT = (I - Z + EZ_{dg})D \tag{5}$$

where Z_{dg} is the matrix formed from the diagonal elements of the fundamental matrix $Z = (I - (P - A))^{-1}$, with A the limiting matrix of the chain. D is the diagonal matrix with elements $d_{i,i} = 1/a_i$ with a_i the limiting probability of the chain being in state i. E is a matrix of ones.

While mobility statistics provide insight into the overall dynamics of the system, the GRM affords a particularly interesting analytic in the form of a *Rank Path* that can provide a spatially explicit view of these dynamics. The path for each rank consists of the consecutive source and destination regions that the rank moves between. The rank path, thus, provides a view of *rank migration*. Because there is a path for each rank in the variate distribution, comparisons between different rank paths can provide insight into the spatial structure of the distribution's evolution. In what follows, we develop a new visual analytic that exploits this opportunity, together with new approaches to visualizing the information provided by the MFPT and Geographic Rank Transition Matrix.

A web-based visual analytics systems for exploring rank mobility

Our visual analytic framework consists of two tools, a RPV and a Geographic Rank Markov Explorer (GRME).

An overview of the RPV is provided in Fig. 1. The system is built with a collection of open-source JavaScript libraries and presents to the user five interactive elements comprising the web-based interface. Starting at the top left, there are controls allowing the user to specify what ranks and time periods are to be examined. This affords different types of filtering that we demonstrate more fully in the case study below. Below the controls is a map canvas, which portrays the rank

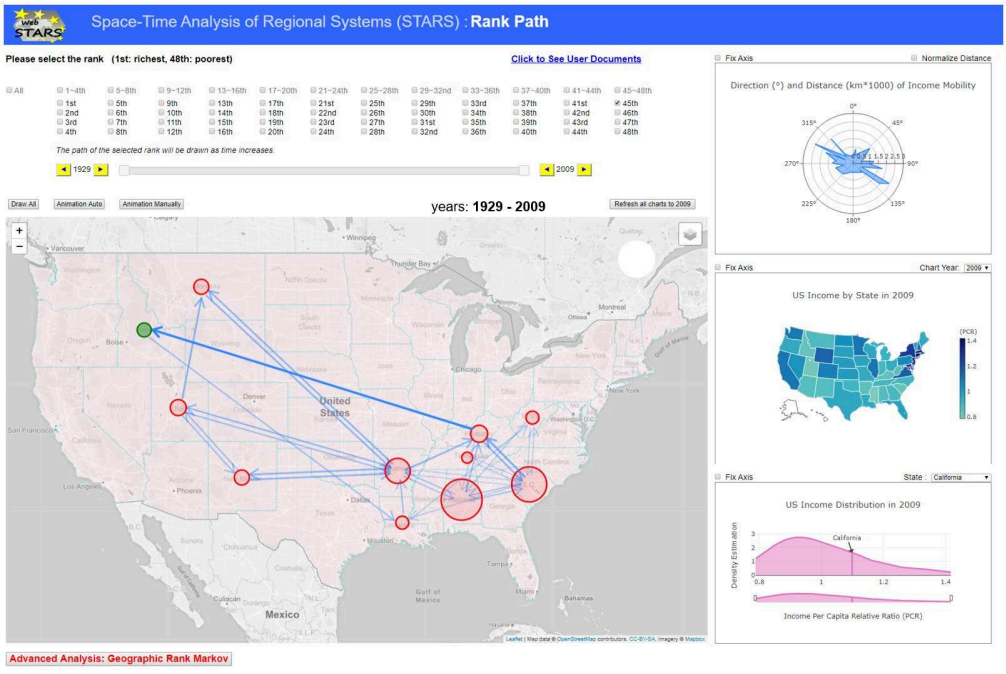

Figure 1. Rank Path Visualizer: The large map canvas plots movement of the 45th rank from 1929 to 2009. Above the map are button controls that allow the selection of ranks, and a slider bar that controls the time window. On the right column, a radar plot reports the angular direction of the rank migrations, on top of a choropleth map for per capita income values in the last year of the sample. The bottom right figure contains two views of the density of per capita incomes for the current year: the top focuses on a selected portion of the density and the bottom has the control which allows the user to select portions of the density for examination.

path for the currently selected rank. The next three components are nested in the right column, beginning with a radar chart at the top of the right panel to visualize the direction and distance of flows. Below the radar chart is another choropleth map which portrays the distribution of per capita incomes at one point in time. Finally, a pair of distributional views are included that display the statistical distribution of the incomes for the same point in time. The bottom distribution is an overview of the full distribution which has a slider. The slider provides two knobs that can be moved horizontally to zoom-in the distribution between the two knobs. The top density shows the section of the distribution focused by the slider.[8] As the user hovers the pointer over the distribution, the state name appears to show its location in the distribution. A zoomable density distribution is especially useful in the high-density area of the graph to help the user locate the state in the distribution. The five elements in the interface are interconnected and enable interactive exploration of the space–time dynamics of the attribute of interest. Each of the elements is dynamically linked with one another such that user interaction in the choropleth map will also be reflected in the density charts and radar charts. For example, using a lasso select on the map will trigger a recalculation of the density estimate based on the selected states only.

Fig. 2 provides an overview of the GRME. To construct our GRM Chain models and associated analytics, we leverage the open-source Python package "giddy" (Kang et al. 2019) part

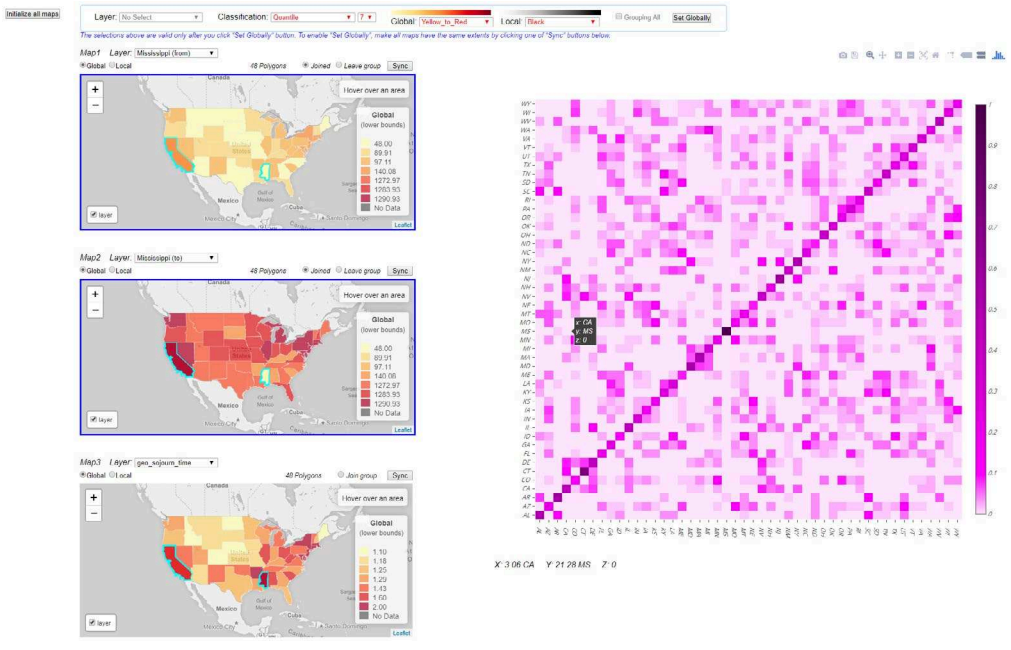

Figure 2. Geographic Rank Markov Explorer: Top controls for selection of color map and classification scheme to map mean first passage time (MFPT) and sojourn times for ranks. The top map displays MFPT for ranks leaving Mississippi; the middle map shows MFPT for ranks entering California; the bottom map displays sojourn time for ranks in all US states. On the right is a heatmap matrix of the MFPT for ranks between all states, with a selected cell highlighting California and Mississippi on the three maps.

of the Python Spatial Analysis Library (PySAL) ecosystem (Rey and Anselin 2007). Similar to the RPV, the GRME has five interface elements based upon Han et al. (2019)'s Adaptive Choropleth Mapper. The toolbar along the top allows users to set parameters used to draw the set of choropleth maps below. Users can select from a list of summary statistics describing mobility from or to each region, including sojourn times and mean first passage times by clicking on the dropdown box above each map. On the right side of the interface, a heatmap displays transition rates, defined in (3) between each pair of Markov states. The heatmap is interactive such that hovering over each cell in the matrix displays a popup with the summary statistic for that cell. The left side displays three maps whose content can be conveniently adjusted by the users. The maps in Fig. 2 showcase (1) the MFPTs from Mississippi (2), the MFPTs to California, and (3) the Sojourn times for all US states.

Illustration: Spatial dynamics of US income inequality

We demonstrate the core functionality of our visual analytics system using US per capita income data annually from 1929 to 2009, from the US Bureau of Economic Analysis.[9] The focus is on the 48 conterminous states. We first illustrate the use of the RPV, followed by a demonstration of the GRME. The tools are highly interactive, so it is important to stress the static images reported

below do not capture the full dynamics of the system. Interested readers can explore the interactivity in more detail online.[10]

Rank-Path Visualizer

The departure point for our illustration is the overview of all rank paths portrayed in Fig. 3. Such a view is clearly overly complex as the vast number of paths makes the identification of pattern difficult. To allow for more refined inspection of the rank paths, a number of different types of filtering are supported by the tool: filtering by rank; filtering by space; filtering by time; and filtering by combinations of rank–space–time.

The map of the rank path was designed to visualize where the selected rank path moves in each different year and how often they move or stay among the 48 lower US states in each year from 1929 to 2009. The path of a rank is represented using distinctive colors to differentiate each different rank path (e.g., Fig. 3). As seen in the legend area of Fig. 3, nine visually distinctive colors are used to differentiate the first nine rank paths from one another, and the nine colors are repeated to color the rest. Each rank path can be toggled on and off, which helps differentiate two or more rank paths represented in the same color. The legend is also collapsable, it opens when the user mouse hovers over the layers symbol on the top right corner of the map.

The location of the circles represents the origins and destinations of each movement. In the case when a user isolates a single rank path, such as Fig. 1, the origin is represented in red, and the destination is represented in green. The radii of the circles increase with the frequency at which the state obtained a specific rank over the time series. In Fig. 1, Alabama ranked 45th 26 times while Louisiana held the same rank only 3 times, hence Alabama's circle is much larger.

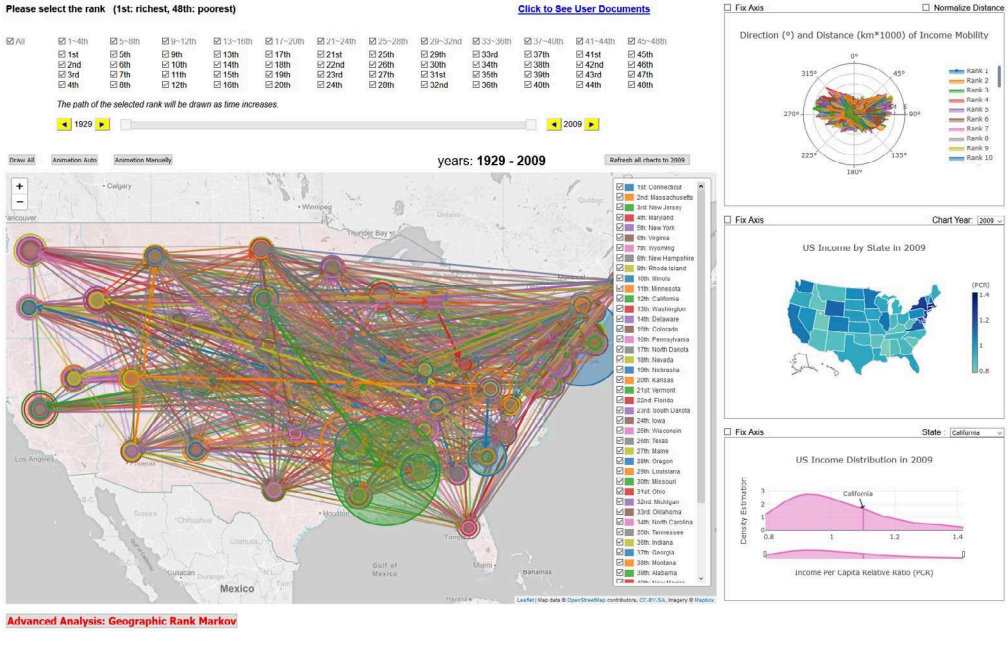

Figure 3. Rank Path Visualizer: Overview with all rank paths selected for all 48 conterminous states. See Fig. 1 for details on components.

When multiple rank paths are visualized, the color of the circles is the same as the color of the lines to differentiate clearly the paths of each rank (e.g., Fig. 3).

Filtering by rank

Fig. 4 demonstrates the case comparing the rank paths for three parts of the distribution: Rank paths 1, 24, and 48. Several visual cues are used to represent the temporal dimension of the paths. First, each rank path is distinguished by a unique hue, and the temporal ordering of each segment of the path is represented by an intensity that declines with time. In other words, more recent edges of a path are darker relative to edges that appear earlier in the path. Edges that repeat take on the intensity associated with the latest period the edge was traversed, and the width of the edge is used to represent the edge's frequency. The terminal node for each path is identified by increasing the saturation of the hue filling the node.

Filtering by rank reveals how spatial dynamics vary over the two tails and the center of the distribution. At the extremes of the distribution, the rank paths display strong return and staying tendencies focusing on Connecticut (Rank Path 1) and Mississippi (Rank Path 48), respectively. In contrast, the path associated with the middle of the income distribution (Rank Path 24) shows

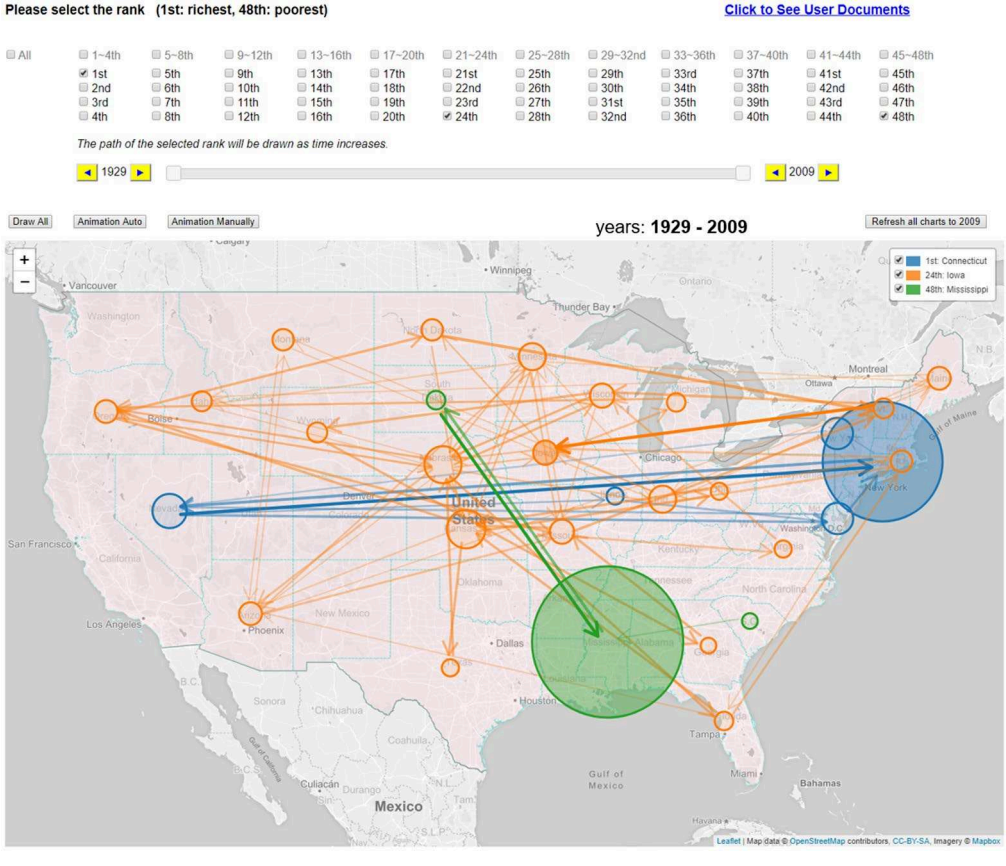

Figure 4. Filtering by Rank: Selection of Ranks 1, 24, and 48 over all time periods. The radius of graduated circle reflects the number of times the rank was held by that state. States that held the selected rank most frequently are labeled in the upper right.

a much more dispersed pattern over the entire sample period. This filtering by rank provides a mechanism to identify spatial polarization in the rank distribution.

Filtering by time

A second type of filtering can be adopted to contrast the dynamics over temporal subsets of the sample. An example can be seen in comparing Figs. 5 and 6 splitting the data into two temporal regimes: 1929–1969 and 1970–2009. For each of the three rank paths, there is a change in the greater concentration of rank migrations between fewer nodes, with the pattern most pronounced for the two extreme ranks.

The visulization lays bare a vivid spatial poverty trap in which Mississippi is prominent during the second period; the state never escapes the bottom of the income distribution. This is in contrast to the previous period where there was some migration of the 48th rank. Interestingly, even in the earlier period, the migration for the 48th rank follows a circular path, briefly escaping to South Dakota before returning to Mississippi. At the other end of the distribution, there is a similar circulation of the top rank between Connecticut and Nevada in the second period, whereas more states were visited by the migration of rank 1 in the first period.

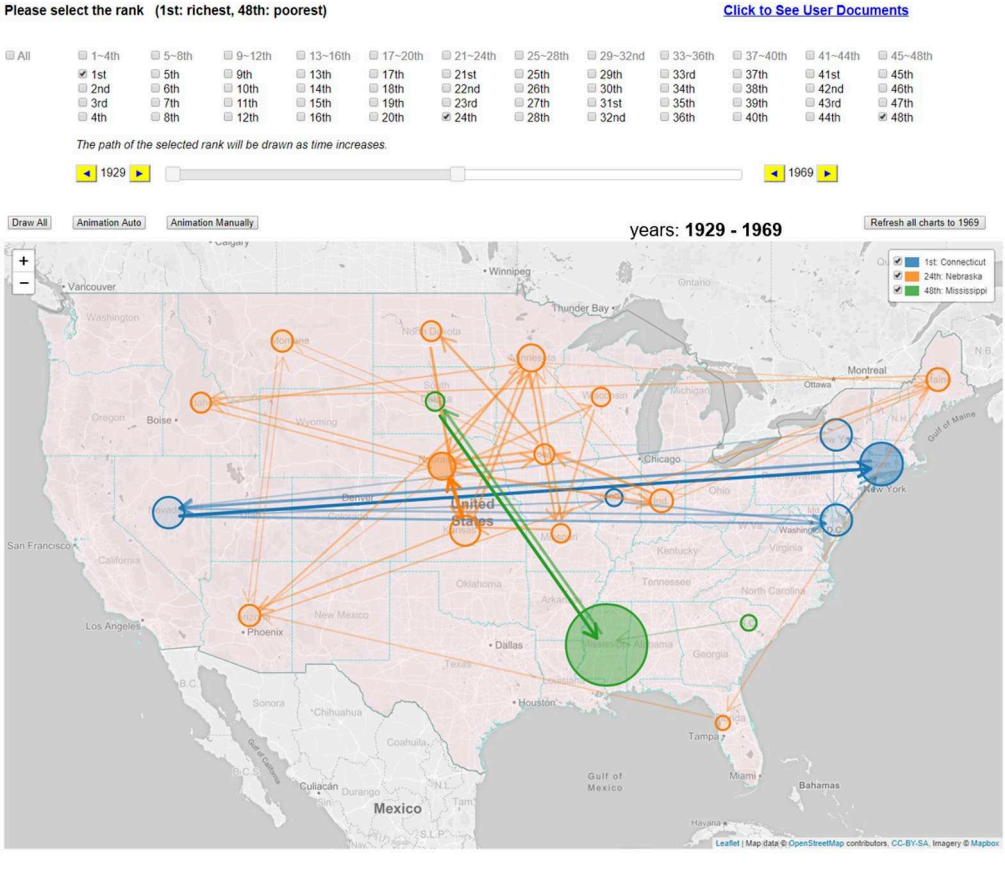

Figure 5. Filtering by Time and Rank: Rank Paths 1, 24, 48 over 1929–1969. See Fig. 4 for details on components.

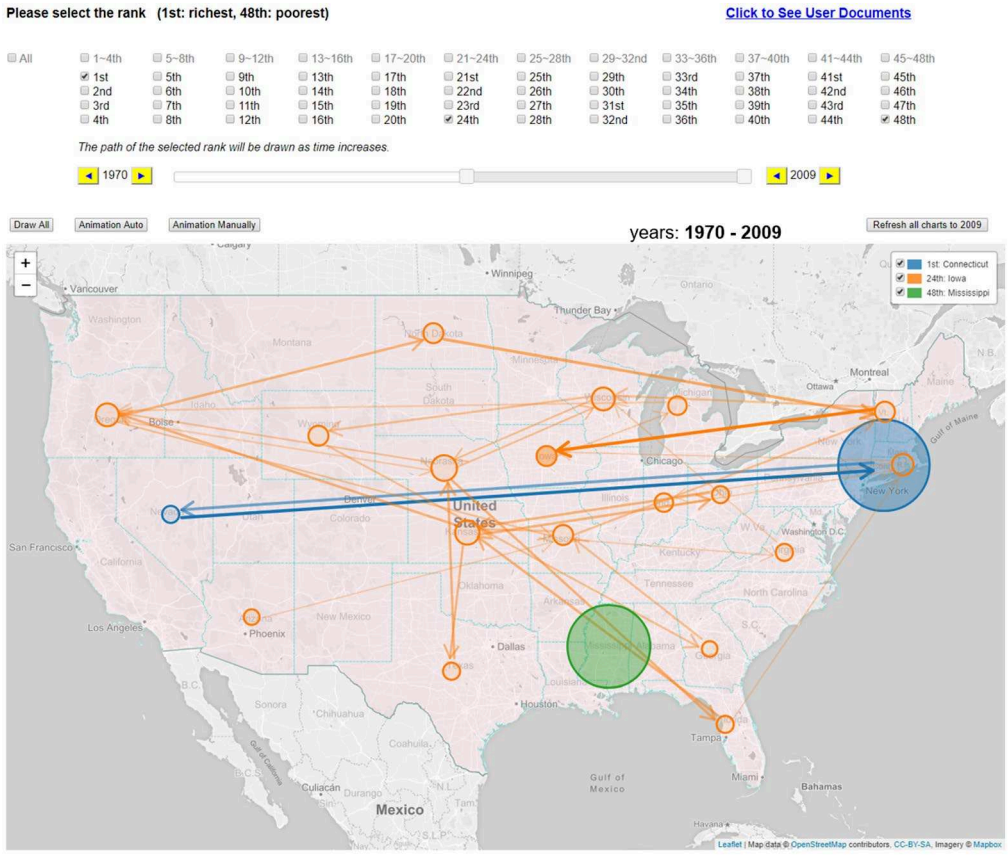

Figure 6. Filtering by Time and Rank: Rank Paths 1, 24, 48 over 1970–2009. See Fig. 4 for details on components.

Filtering by space

Filtering the data by region allows a user to focus on the dynamics within the selected spatial units only. For instance, after selecting n_1 spatial units, numeric ranks are reassigned to these n_1 selected observations beginning with 1 for the smallest value and n_1 for the smallest value to create “local” or “conditional" ranks, in the sense that ranks are recreated to be locally adaptive or conditioned on the selected regions. Filtering by space can be achieved by applying the lasso tool to the middle map on the right of the RPV. Fig. 7 illustrates this user case where the user has changed to focus to the south, having selected 10 states ($n_1 = 10$) that are highlighted in the main map. This triggers a change in the interface for rank filtering (top of tool) to only reflect the internal rank migrations between the selected 10 states, where the ranks now become local, or conditional, to the selected sets. Here the user is now examining the movement of the first and ninth ranks within this conditional set of southern states. Doing so reveals that Florida used to be the richest state in the south, however, it has been surpassed by Virginia which assumes the top local rank at the end of the period. Conversely, the rank path of 10th local rank moved back and forth between Alabama and South Carolina until 1982, moved to West Virginia in 1983 before returning to South Carolina where it remained.

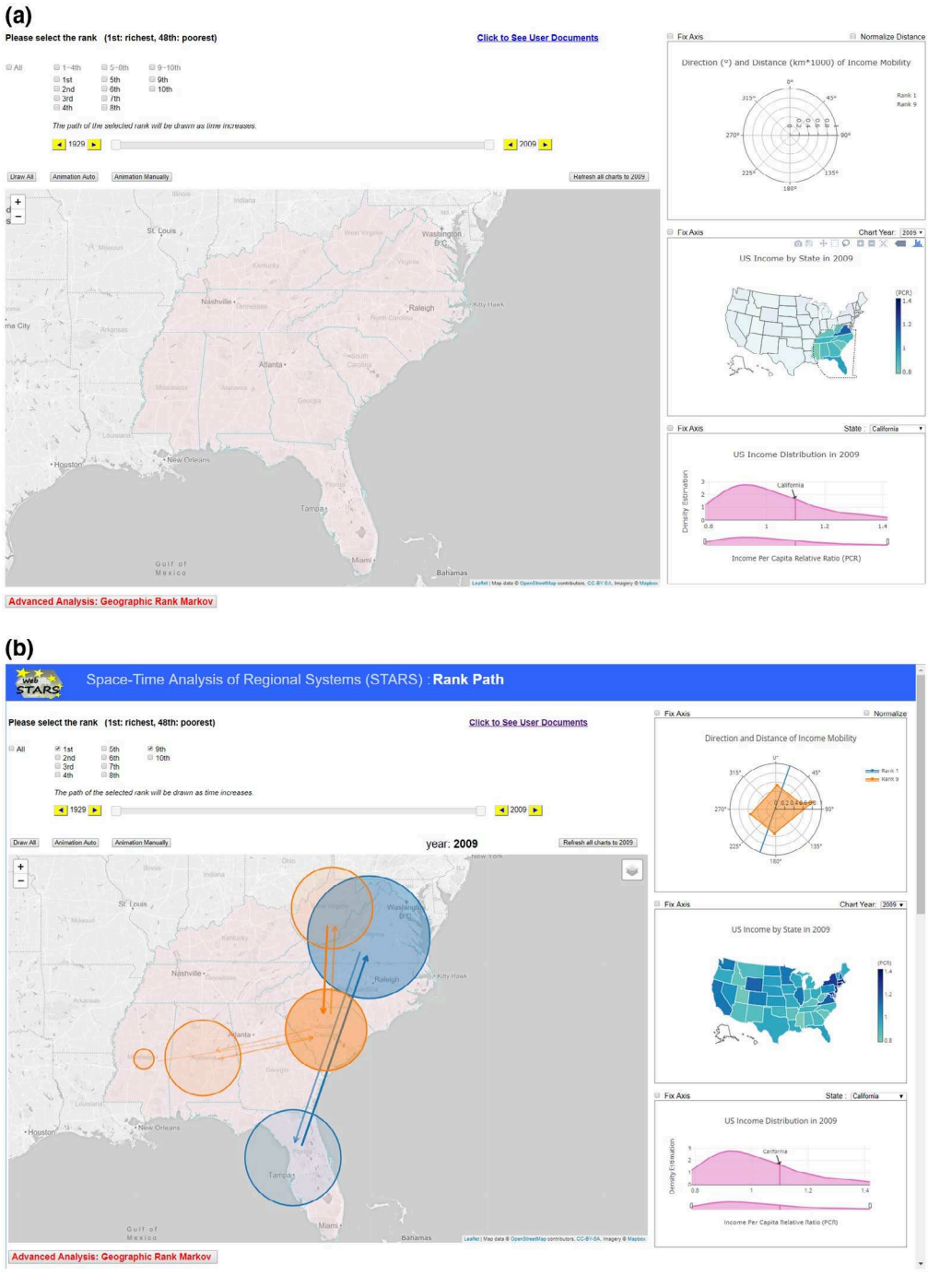

Figure 7. Spatial filtering: (a) A lasso tool to select regions; (b) The movement of the first and the last rank in a subset of 10 southern states.

GRM explorer

The dynamics of the income rank migrations across space and time are modeled with GRM, which provides a suite of summary statistics to reveal interactions between regions. As discussed earlier, the combined use of RPV and GRME offers insights which cannot be revealed using only one tool in isolation.

GRM overall

Fig. 2 provides the visualization of the GRM mobility statistics, serving as a concise summary complementary to the rank path visualization in Fig. 3. More specifically, a GRM transition probability matrix $P(g)$ was estimated and visualized as a heatmap on the right side. Each cell presents the probability of ranks migrating from the corresponding US state (indicated by the column of state names to the left of the heatmap) to another (indicated by the row of state names to the bottom of the heatmap). The diagonal cells show the probabilities of income ranks staying in correspondent states.

It is clear from the heatmap that the darkest cell lies in the row and column of Mississippi. Utilizing the convenient interactive functionality of GRM Explorer, we can obtain the value of this transition probability by hovering over the cell to reveal the probability of income ranks staying in Mississippi over a year interval as 96.5%. As shown in the third map in Fig. 3, the Sojourn time for Mississippi is 26.67 years, meaning that it takes an average of 26.67 years for an income rank to leave Mississippi once it has entered the state. Recall the visualization of the path of income rank 48 (the poorest) in Fig. 4 which tends to return to and stay at Mississippi, we can verify and complement our visual inspection with quantitative evidence by turning to the first two maps which visualize the expected number of years a rank takes to migrate from Mississippi to any other state and from any other state to Mississippi.[11] From these two maps, we can see that it generally takes a much longer time for an income rank to move to Mississippi than moving from Mississippi. For instance, it takes a rank 1297 years to move from California to Mississippi while it takes 147 years for a rank to move from Mississippi to California. It would seem that Mississippi tends to be trapped in the bottom of the income distribution, lacking upward mobility.

GRM—Temporal structural breaks

Following the filtering by time option in RPV, we can also utilize GRME to model the dynamics of income ranks across space and time for selected periods. In Fig. 8, we have partitioned the data into the same temporal regimes used in Section “Filtering by time”. At first glance, a visual comparison of the MFPT maps and transition probability matrix heat maps between the two periods does not reveal any obvious differences in the spatial dynamics over the two periods. However, closer inspection reveals that there is an important structural difference between the two periods. In the second period, the graph that is obtained from the rank paths is not fully connected as there are two connected components, one consisting of 47 states, and the second component is the isolate of Mississippi. This is clearly evident from Fig. 6 where the rank path for rank 48 is a singleton at Mississippi.

In contrast, the graph for the rank paths in the first period is fully connected. In addition to the differences in connectivity structures, the MFPTs are generally larger in the first period. For example, the MFPT for a rank leaving California to enter Alabama is 256 years in the first period, but this time drops to 104 years in the second period of the sample. Two factors may be at work behind this change. First, the rank paths for 1929–1969 include those that circulate through

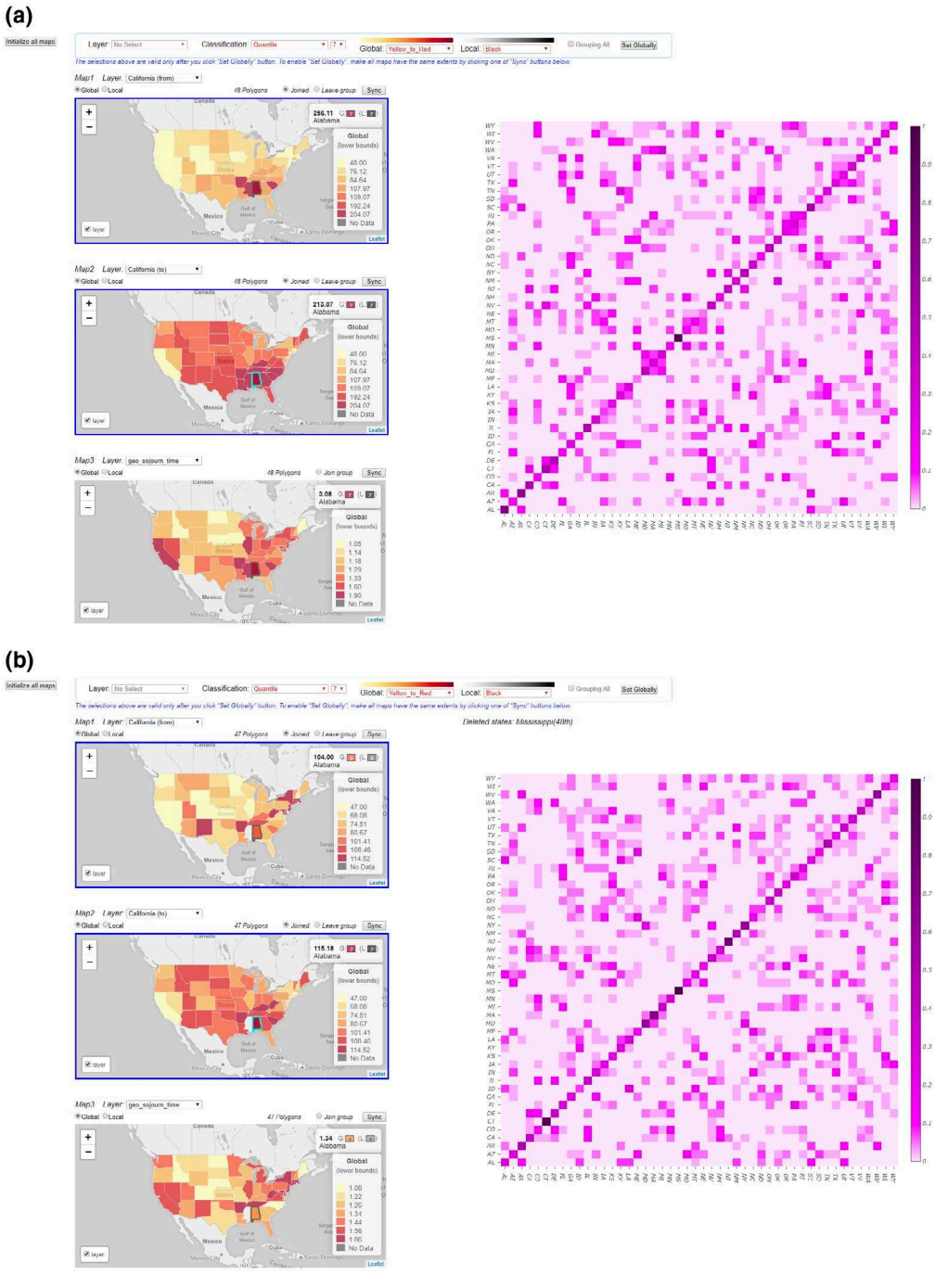

Figure 8. Temporal change of income rank mobility: (a) GRM modeling of US states over 1929–1969 (b) GRM modelling of US states over 1970–2009. See Fig. 2 for details components.

Mississippi, while in the later period, Mississippi is not a node in the larger connected component that is used to estimate the MFPTs. As we saw earlier, the rank paths involving Mississippi in the first period are highly concentrated. Thus, any path that would include edges leading to

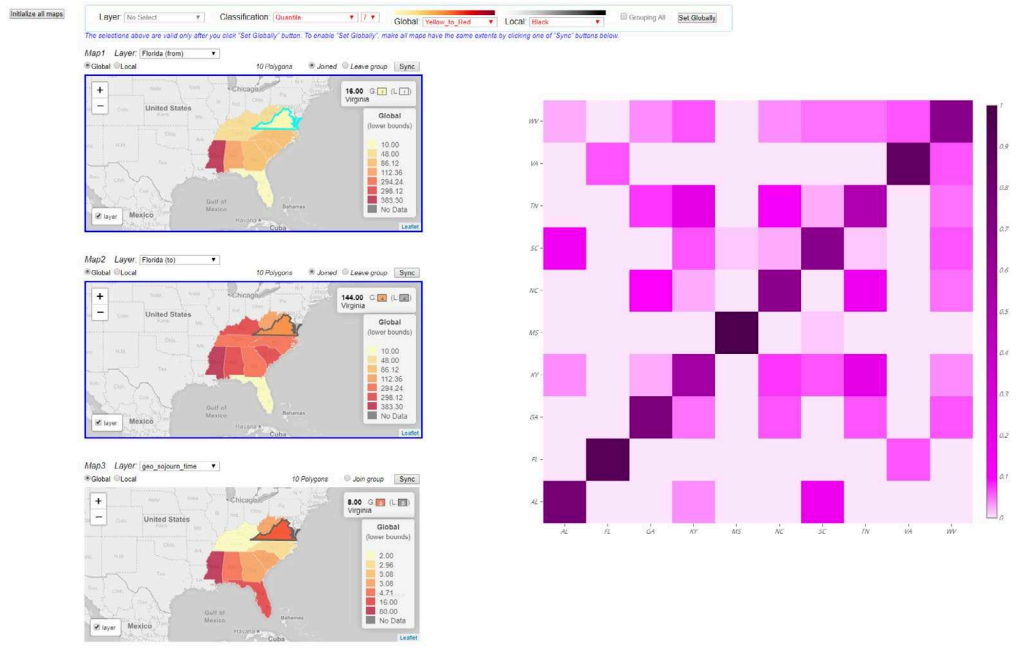

Figure 9. Filtering by Space: GRM modeling of 10 Southern US states over 1929–2009. All analytics are now recalculated for the selected states only. See Fig. 2 for details on components.

or from Mississippi will incur larger escape costs. The second factor is that the MFPTs from the two periods are based on different size-connected components, with all 48 states used in the first period, while only 47 in the second period, so the later graph is smaller and less complex.

GRM in a regional/local context

Similar to what we saw with the RPV in Section "Filtering by space", we can select a subset of states to explore spatial heterogeneity in the overall income dynamics. Fig. 9 demonstrates this for the same 10 southern states that were selected previously in Fig. 7. Here the focus is on how a number of properties of the spatial dynamics may manifest local differentiation.

The selection of a subset of states creates a new set of estimates of the GRM whose transition probability matrix is shown as a heat map in the right side of the Fig. 9. Similar to the global GRM matrix, this local matrix is diagonally dominant, reflecting the higher probability of a rank remaining in a state over a year interval rather than moving to a new state. The top two maps are used to examine the MFPT from, and to, Florida, but conditioned upon the collection of 10 states. States in the rest of the system are excluded from the estimation. Virginia is the focal state selected in the top map which reveals that the time it takes for a rank to leave Florida to arrive in Virginia is in the order of 16 years. In contrast, the second map indicates that it takes a considerably longer period for a rank that Virginia held to migrate to Florida. In other words, there is an asymmetry in the timing of these two directional rank migrations.

Conclusion

The impact of Tobler's work on flow mapping continues to grow in both the scope of the different domains adopting the framework, as well as the sheer number of applications. In this article,

we contribute to this line of research by integrating concepts from flow mapping with recently developed measures to model spatial income distribution dynamics. A new web-based visual analytic framework recasts the notion of a rank path as tracing the migration of ranks of a variable distribution over a collection of locations. Coupled with this is a geographical rank Markov exploration tool that provides summary measures on the spatial dynamics that complement the RPV. Both of these tools are highly interactive and implement different forms of data filtering which can be exploited to explore spatial, temporal, and distributional heterogeneity in longitudinal spatial data. We illustrate selected components of the framework by examining US regional income distribution dynamics over the 48 lower states from 1929 to 2001.

There are a number of enhancements to the RPV and GRM Tools that we are planning to pursue. In the original development of the rank-based Markov chains, several global and local tests for spatial dependence in the rank exchanges were developed (Rey, Kang, and Wolf 2016) that could be added to the interface to complement the visual analytics presented here. Related to this, the rank paths have a natural affinity to graph theory, and there is a wealth of graph theoretic results, summary measures, and visualization concepts that we intend to explore for incorporation into the RPV.

Space–time data come in a rich variety of forms, and our initial implementation of both the RPV and GRM tools assumes a particular kind of longitudinal spatial data where the spatial boundaries are fixed over time for the polygon units under consideration. Extending the framework to consider other types of spatial support, including marked point patterns, and various spatial networks (street networks, pipelines, etc.) remains an important future direction. Moreover, our example illustration is of a modest data size (48 spatial units over 81 periods), and we are considering modifications of the design of the filtering functionality as well as of the animation mechanisms to scale to big data contexts.

Acknowledgement

This research was supported in part by NSF Grant/Award Numbers: BCS-1759746 and SES-1733705; Renan Xavier Cortes is grateful for the support of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) foundation (Process 88881.170553/2018-01).

Notes

[1]https://usmigrationflowmapper.com/.

[2]https://flowsmapper.geo.census.gov/.

[3]Available at https://visualiza.fee.tche.br/crime/. only in Brazilian Portuguese. The Markov analysis is present in the Section of “Mapas” (Maps) and then “Cadeias de Markov” (Markov Chains).

[4]For a recent overview see Rey (2014a).

[5]A referee noted that the term “Rank” as used here should not be confused with the rank of a matrix. Instead, as explained in the text, the ranks of the attribute values are used to form the matrix.

[6]For the derivation of the Mean Sojourn time, the interested reader is directed to Ibe 2009, sec. 4.6).

[7]For the derivation of the MFPT the interested reader is directed to Kemeny and Snell 1976, p. 76). For a detailed description of the use of first passage times in the context of the *GRM* refer to Rey (2014b). All the computations were conducted in Python using PySAL/giddy (Kang et al. 2019) which has implemented *GRM* and relevant statistics.

[8]The top density is not renormalized as the slider moves.

[9]https://www.bea.gov/data/income-saving/personal-income-by-state.

[10]URL Removed temporarily to protect author anonymity

[11]These are based on the mean first passage times estimated using equation (5).

References

Bloom, D. E., D. Canning, and J. Sevilla (2003). "Geography and Poverty Traps." *Journal of Economic Growth* 8, 355–78.

Boyandin, I., E. Bertini, and D. Lalanne. (2010). "Using Flow Maps to Explore Migrations Over Time." In Geospatial Visual Analytics Workshop in Conjunction with The 13th AGILE International Conference on Geographic Information Science, 2. https://diuf.unifr.ch/main/diva/sites/diuf.unifr.ch.main.diva/files/jflowmap-geova10.pdf

Buchin, K., B. Speckmann, and K. Verbeek (2011). "Flow Map Layout via Spiral Trees." *IEEE Transactions on Visualization and Computer Graphics* 17, 2536–44. https://ieeexplore.ieee.org/document/6065021/

CARTO. (2017). Torque. https://github.com/CartoDB/torque (2019/05/09). Available at https://github.com/CartoDB/torque (2019/05/09).

Chang, W., J. Cheng, J. Allaire, Y. Xie, and J. McPherson. (2018). shiny: Web Application Framework for R. https://CRAN.R-project.org/package=shiny. R package version 1.1.0.

Chen, S., X. Yuan, Z. Wang, C. Guo, J. Liang, Z. Wang, X. L. Zhang, and J. Zhang (2015). "Interactive Visual Discovering of Movement Patterns from Sparsely Sampled Geo-Tagged Social Media Data." *IEEE Transactions on Visualization and Computer Graphics* 22, 270–79.

Clark, W. A. V., and S. Koloutsou-Vakakis (1992). "Evaluating Tobler's Migration Vector Fields." *Geographical Analysis* 24, 110–20. https://doi.wiley.com/10.1111/j.1538-4632.1992.tb00255.x

Cortes, R. X. (2017). Ensaios em criminalidade no Rio Grande do Sul. Pontificia Universidade Catolica do Rio Grande do Sul.

Demšar, U., and K. Virrantaus (2010). "Space–Time Density of Trajectories: Exploring Spatio-Temporal Patterns in Movement Data." *International Journal of Geographical Information Science* 24, 1527–42. https://www.tandfonline.com/doi/abs/10.1080/13658816.2010.511223

Eastman, J. R. (2015). Terrset Manual. Worcester, Massachusetts, US: Clark University.

Guo, D. (2007). "Visual Analytics of Spatial Interaction Patterns for Pandemic Decision Support." *International Journal of Geographical Information Science* 21, 859–77. https://www.tandfonline.com/doi/abs/10.1080/13658810701349037

Guo, D. (2009). "Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data." *IEEE Transactions on Visualization and Computer Graphics* 15, 1041–48.

Han, S., S. Rey, E. Knaap, W. Kang, and L. Wolf (2019). "Adaptive Choropleth Mapper: An Open-Source Web-Based Tool for Synchronous Exploration of Multiple Variables at Multiple Spatial Extents." *ISPRS International Journal of Geo-Information* 8, 509–26.

Han, S. Y., K. C. Clarke, and M.-H. Tsou. (2017). "Animated Flow Maps for Visualizing Human Movement: Two Demonstrations with Air Traffic and Twitter Data." In Proceedings of the 1st ACM SIGSPATIAL Workshop on Analytics for Local Events and News, 5. ACM. https://dl.acm.org/citation.cfm?id=3148049

Ho, Q., P. H. Nguyen, T. Åström, and M. Jern (2011). "Implementation of a Flow Map Demonstrator for Analyzing Commuting and Migration Flow Statistics Data." *Procedia-Social and Behavioral Sciences* 21, 157–66.

Ibe, O. (2009). Markov Processes for Stochastic Modeling. Amsterdam: Elsevier Academic Press.

Jenny, B., D. M. Stephen, I. Muehlenhaus, B. E. Marston, R. Sharma, E. Zhang, and H. Jenny (2018). "Design Principles for Origin-Destination Flow Maps." *Cartography and Geographic Information Science* 45, 62–75. https://www.tandfonline.com/doi/full/10.1080/15230406.2016.1262280

Johnson, S. J. (2009). An Evaluation of Land Change Modeler for ARCGIS for the Ecological Analysis of Landscape Composition. Carbondale: Southern Illinois University.

Jordan, P., and P. Talkner (2000). "A Seasonal Markov Chain Model for the Weather in the Central Alps." *Tellus A* 52, 455–69.

Kang, W., S. Rey, P. Stephens, N. Malizia, L. J. Wolf, S. Lumnitz, C. Jlaura, C. Schmidt, E. Knaap, J. Gaboardi, and A. Eschbacher. (2019). Geographical Distribution Dynamics (giddy). https://doi.org/10.5281/zenodo.2633309

Kemeny, J. G., and J. L. Snell (1976). Finite Markov Chains. New York: Springer-Verlag.

Kim, K., S.-I. Lee, J. Shin, and E. Choi (2012). "Developing a Flow Mapping Module in a GIS Environment." *The Cartographic Journal* 49, 164–75. https://www.tandfonline.com/doi/full/10.1179/174327711X13166800242356

Koblin, A. (2006). "Flight Patterns." In ACM SIGGRAPH 2006 Computer Animation Festival, 203. ACM. https://dl.acm.org/citation.cfm?id=1179221.

Logsdon, M. G., E. J. Bell, and F. V. Westerlund (1996). "Probability Mapping of Land Use Change: A GIS Interface for Visualizing Transition Probabilities." *Computers, Environment and Urban Systems* 20, 389–98.

Magrini, S. (1999). "The Evolution of Income Disparities Among the Regions of the European Union." *Regional Science and Urban Economics* 29, 257–81. https://www.sciencedirect.com/science/article/pii/S0166046298000398

Minard, C. J. (1862). Des Tableaux graphiques et des cartes figuratives, par M. Minard ··· Thunot. https://books.google.com/books?id=pJtyQwAACAAJ

Muller, M. R., and J. Middleton (1994). "A Markov Model of Land-Use Change Dynamics in the Niagara Region, Ontario, Canada." *Landscape Ecology* 9, 151–57.

Pan, J., and A. Nagurney (1994). "Using Markov Chains to Model Human Migration in a Network Equilibrium Framework." *Mathematical and Computer Modelling* 19, 31–39.

Rae, A. (2009). "From Spatial Interaction Data to Spatial Interaction Information? Geovisualisation and Spatial Structures of Migration from the 2001 UK Census." *Computers, Environment and Urban Systems* 33, 161–78.

Rae, A. (2011). "Flow-Data Analysis with Geographical Information Systems: A Visual Approach." *Environment and Planning B: Planning and Design* 38, 776–94. https://journals.sagepub.com/doi/10.1068/b36126

Rey, S. J. (2014a). "Handbook of Regional Science." Chap. Spatial Dynamics and Space-Time Data Analysis, 1365–1383. Berlin, Heidelberg: Springer, Berlin Heidelberg. doi: 10.1007/978-3-642-23430-9_78

Rey, S. J. (2014b). "Rank-Based Markov Chains for Regional Income Distribution Dynamics." *Journal of Geographical Systems* 16, 115–37.

Rey, S. J. (2016). "Space-Time Patterns of Rank Concordance: Local Indicators of Mobility Association with Application to Spatial Income Inequality Dynamics." *Annals of the American Association of Geographers* 106, 788–803. https://doi.org/10.1080/24694452.2016.1151336

Rey, S. J., and L. Anselin (2007). "PySAL: A Python Library of Spatial Analytical Methods." *The Review of Regional Studies* 37, 5–27.

Rey, S. J., and M. V. Janikas (2006). "STARS: Space-Time Analysis of Regional Systems." *Geographical Analysis* 38, 67–86.

Rey, S. J., W. Kang, and L. Wolf (2016). "The Properties of Tests for Spatial Effects in Discrete Markov Chain Models of Regional Income Distribution Dynamics." *Journal of Geographical Systems* 18, 377–98. https://doi.org/10.1007/s10109-016-0234-x

Rey, S. J., E. A. Mack, and J. Koschinsky (2012). "Exploratory Space-Time Analysis of Burglary Patterns"." *Journal of Quantitative Criminology* 28, 509–31.

Rogers, S. (2012). Britain's Royal Navy in the First World War—Animated. https://www.theguardian.com/news/datablog/interactive/2012/oct/01/first-world-war-royal-navy-ships-mapped (2019/05/09).

Slocum, T. A., R. B. McMaster, F. C. Kessler, and H. Howard (2008). Thematic Cartography and Geographic Visualization, 3rd ed. Upper Saddle River NJ: Prentice Hall.

Stephen, D. M., and B. Jenny (2017). "Automated Layout of Origin-Destination Flow Maps: U.S. county-to-County Migration 2009–2013." *Journal of Maps* 13, 46–55. https://doi.org/10.1080/17445647.2017.1313788

Tobler, W. R. (1981a). "Depicting Federal Fiscal Transfers." *The Professional Geographer* 33, 419–22. https://doi.org/10.1111/j.0033-0124.1981.00419.x

Tobler, W. R. (1981b). "A Model of Geographical Movement." *Geographical Analysis* 13, 1–20.

Tobler, W. R. (1987). "Experiments in Migration Mapping by Computer." *The American Cartographer* 14, 155–63.

Tobler, W. R. (2004). Flow Mapper Tutorial. Santa Barbara: Center for Spatially Integrated Social Science, University of California. https://www.csiss.org/clearinghouse/FlowMapper/FlowTutorial.pdf

Tufte, E. R. (2001). The Visual Display of Quantitative Information. Cheshire, CT, USA: Graphics Press.

Wood, J., J. Dykes, A., and Slingsby (2010). "Visualisation of Origins, Destinations and Flows with OD Maps." *The Cartographic Journal* 47, 117–29. https://www.tandfonline.com/doi/full/10.1179/000870410X12658023467367

Woodruff, A. (2019). Explore Hubway Trips Filtered by Time, Demographic, and Weather Factors. https://bostonography.com/hubwaymap/. Available at https://bostonography.com/hubwaymap/ (2019/05/09).

Yang, Y., T. Dwyer, B. Jenny, K. Marriott, M. Cordeil, and H. Chen (2019). "Origin-Destination Flow Maps in Immersive Environments." *IEEE Transactions on Visualization and Computer Graphics* 25, 693–703. https://doi.org/10.1109/tvcg.2018.2865192