Visualizing Migration Dynamics Using Weighted Radial Variation

Clio Andris and Frank Hardisty

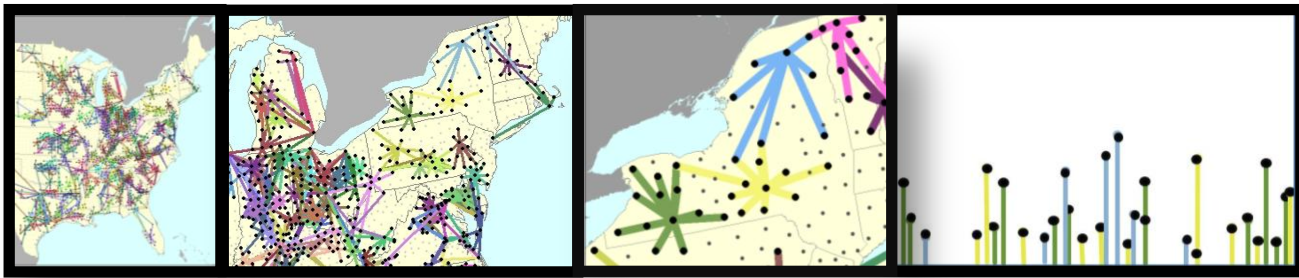

Fig. 1. Migration stars in two different zoom views, followed by a focus on three outgoing migration star patterns in New York State, and their respective weighted radial variation signals.

Abstract—Directional flows created from an origin/destination matrix have been traditionally difficult to visualize because of the number of flows to be rendered in a small cartographic space. Because visualizing geographic flow dynamics are useful for understanding the complex dynamics of human and information flow that connect non-adjacent space, techniques that allow for visual data mining or static representations of system dynamics are a growing field of research. Here, we use a Weighted Radial Variation (WRV) technique to classify places based on their group's radially-emanating vector flows. Each entity's vector are syncopated in terms of cardinality, direction, length, and flow magnitude. The WRV process unravels each star-like entity's individual flow vectors on a 0-360° spectrum, to form a unique signal whose distribution depends on the flow presence at each step around the entity, and is further characterized by flow distance and magnitude. The signals are processed with a supervised classification method that clusters entities with similar signatures or trajectories in order to learn about types and geographic distribution of flow dynamics. We use U.S. county-to-county human incoming and outgoing migration data to test our method.

Index Terms— Movement, Flow mapping, Geovisualization, Feature reduction, Graph structures

1 Introduction

We present a novel way to classify geographic entities (origins and destinations) in a flow system while preserving the individual characteristics (flow magnitude, weight and direction). Using these classifications, we can better visualize, and thus better understand the nature and dynamics of large complex geographic flow systems. This research is driven by the increasing availability of large datasets: Data from cell phone traces, traffic sensors, flight schedules and telephone records, and government digital collections are now becoming more common sources for analysis and problem solving for fields such as transportation, logistics and operations, geography and civil engineering.

Computational methods for matrix datasets have already helped researchers in these fields learn more about human and communication transactions across the built environment. The dynamics of large, multi-scale flow systems are often measured with summary factors—like a node's degree (number of neighbors), or a hub's centrality in a whole flow system [1-3].

While these methods continue to inform spatial system dynamics and benefit from cutting-edge complex network analysis (CNA) techniques, their progress has been largely unaccompanied by spatial visualization techniques. The need for good characteristic-reducing techniques for multi-featured spatial systems, like complex flows, are an important component of the static and dynamic, user-interactive, geovisualization tools that currently support visual spatial data mining. [4][5] A recent focus towards visualizing flows [6][7] and object movement [8][9] and space-time dynamics [10] has demonstrated the benefits of these kinds of methods.

Though rendering spatial systems is a growing topic of interest, the nature of complex geographic networks' geographic flow intersections and overlaps, present a natural visualization problem, as point-to-point datasets can have many links, yielding a 'haystack' of links—from which little analysis can be performed. (Figure 2) Early efforts to spatialize flow dynamics include Tobler's computer mapping, where flows were rendered as aggregate arrows in order to fit in a cartographic space, and also woven into vector surfaces that resembled magnetic fields. [11-13] More recently, Andrienko and Andrienko echo the benefits of aggregating flows, adding that multi-scale analysis is now possible. [8] Similarly, Woods et al take a unique approach to flow aggregation by assigning characteristics of OD vectors to cells, instead of the more traditional line summaries. [14] One method that fixes the haystack problem and the aggregation problem (sidesteps summarizing or averaging values, distances or flow direction.) is an interactive system where nodes selection, so that certain links are shown instead of all links. This provides a clearer picture of small-scale behavior, but at the expense of losing the 'all data in one view' advantage, so pre-selected views must be stored and retrieved by memory.

• *Clio Andris is with the Massachusetts Institute of Technology and the Pennsylvania State University GeoVista Center, E-Mail: clio@mit.edu*

• *Frank Hardisty is with the Pennsylvania State University GeoVISTA Center and the John A. Dutton e-Education Institute, E-Mail: hardisty@psu.edu.*

Manuscript received November 5, 2010
For information on obtaining reprints of this article, please send email to: clio@mit.edu

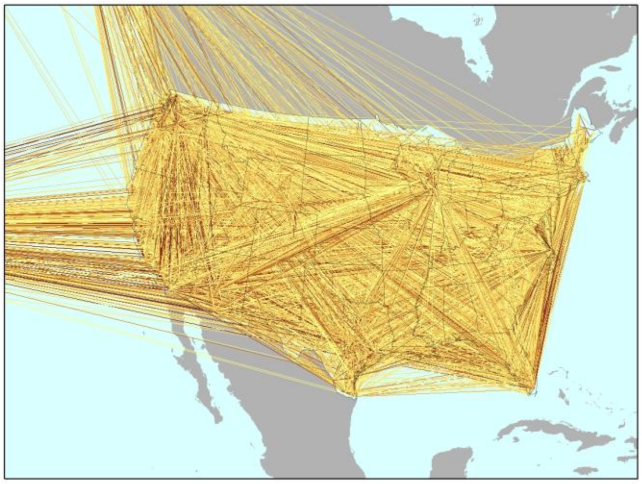

Fig. 2. Example of layered flow lines, where darker lines indicate more migrants moving from an origin county to a destination county.

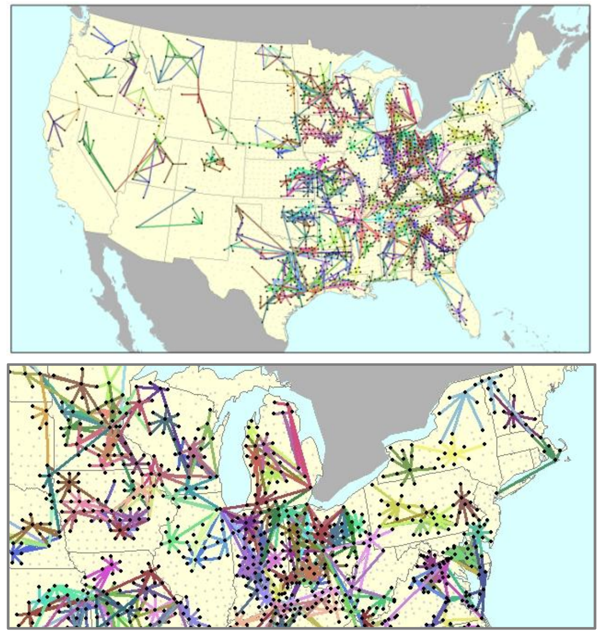

Fig. 3. Migration Patterns for a certain distance-class of U.S. counties, representing 10% of the total origin counties. Each migration star shows a total distance of 600-100 kilometers of straight-line travel.

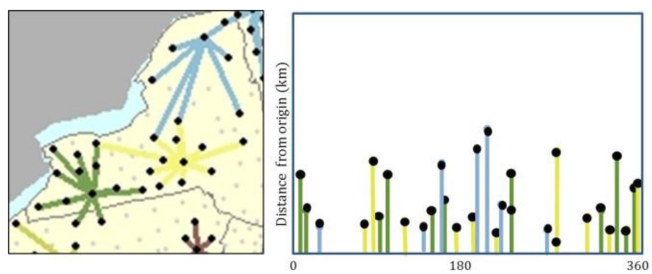

Fig. 4. Each element of the three 'star' patterns above are separated and spread out by their degree on the x axis of the graph above. Each of the blue, yellow and green county stars are then treated as three different signals and classified, as part of the 3500 signal system, with an eigenvector decomposition algorithm.

2 METHODOLOGY

We use data on U.S. county-to-county migration, collected by the IRS, to simulate a node/edge flow system in geographic space. We start with a matrix of 3140 x 3140 entries, where each column and row represents a U.S. county centroid. We do not count self-nodes, when movers choose a new home within the same county, but note that these values are typically high for each place. We add a distance matrix, listing the distance in kilometers between each county centroid and a matrix of angles between each node pair, where the vector head is at the origin and tail at the migrant's destination. (These can, of course, be reversed, if the destination is the node in question, but here, we concentrate on outgoing flows only.)

Our goal is to characterize individual places by their out migration features: for each county, we extract 'stars,' where the node in focus is the center, and the flows leaving the node are attached to the central county centroid, and treated as part of the star. (Figure 3)

The star method has been used before for multidimensional data visualization, where each spoke from the center has a length equivalent to a specified quantitative feature of the entity. [15][16] Others have taken the tool a step further, stressing interactivity [17] and evaluating the effectiveness of different star symbologies [18]. The difference between this glyph-type entity visualization technique and our geographic case is that *each* vector in the radial system represents three (or more) characteristics instead of a single characteristic, as we have measures of (1) distance, (2) direction and (3) magnitude, for each. Also, our vectors are "tacked" to geographic space, meaning that an arc's radial direction is a variable with syncopated occurrences around the 0 – 360° radius, instead of an evenly-spread series of a pre-determined number of spokes in non-spatial star glyphs.

Noting that geographic stars (like graph structures) have a nearly infinite number of possible configurations, our probability of having a certain graph occur could be calculated by the convolution of 4 continuous variables: ray cardinality, magnitude, distance, and angle. To manage and group these e use an 'unraveling' technique to characterize the radial dynamics of each county's graph structure, to measure *Weighted Radial Variation.* For each county, we create a signature vector comprised of an edge weight (number of migrants) and distance value for each angle circling around the node from 0 – 360°. This signature vector is laid out as a signal over the radial steps to show similarities and differences between counties. (Figure 4) The weighted signals are then clustered using a supervised eigenvector decomposition algorithm, and a typology is created for each cluster, where the number of desired resultant cluster classes can be chosen. [19] This type of signal processing and pattern recognition analysis has been successful in a geographic context for classifying and understanding space. [20]

3 RESULTS

When these typologies of county graph types are visualized in geographic space, we are able to compare which counties are similar in their migration behavior, look at regional variation, and join demographic information. We also find that this method is robust with respect to including many edge weights per flow instead of a single measure of total migrants from county i to county j. For example, we can use our method to cluster stars where each flow has a measure of female migrants and male migrants, or migrants by age group. These typologies are then visualized via a single-variable cartographic representation that still represents the anisotropic 'spread/reach' of people migrating from different kinds of locales. From this representation, we can answer questions that would be difficult to answer otherwise, for example: how far flows travel, to what geographic direction are the flows travelling, and the magnitude of movers from each locale.

4 Conclusion

Our aim for this new visualization technique was to preserve individual, disaggregate characteristics of flow data while allowing the information to be explored in a single view. By extracting and clustering different geographically-tacked graph configurations, we are better able to understand the distribution of human movement patterns in space, while using disaggregated data. This method is not limited to migration patterns, but can be used for other datasets where origin “reach” is a metric of interest like commuting flows, phone call volume, or temporal vacation/leisure flows.

Acknowledgements

This work is supported in part by the ASEE National Defense Science and Engineering Fellowship Program, the VACCINE Center (Visual Analytics for Command, Control, and Interoperability Environments, Grant 2009-ST-061-CI0001), and the Contextual influences on the category construction of geographic-scale movement patterns (ConCat) grant from the National Science Foundation (Grant 0924534). The Vaccine Modeling Initiative grant from the Bill & Melinda Gates Foundation (Grant 49279) is gratefully acknowledged, however no endorsement is implied.

References

[1] M. O'Kelly. A geographer's analysis of hub-and-spoke networks. Journal of Transport Geography, 2:171—186, 1998.

[2] Z. Xu and R. Harriss. Exploring the structure of the US intercity passenger air transportation network: a weighted complex network approach. GeoJournal, 73:87—102, 2008.

[3] B. Jiang and C. Claramunt. Topological analysis of urban street networks. Environment and Planning B, 31:151—162, 2004.

[4] L. Anselin. Exploring Spatial Data with GeoDA: A Workbook Spatial Analysis Laboratory. Department of Geography, University of Illinois at Urbana-Champaign, Center for Spatially Integrated Social Science, 2005.

[5] A. MacEachren and M. Kraak. Exploratory cartographic visualization: advancing the agenda. Computers and Geosciences, 23:335—343, 1997.

[6] D. Marble, Z. Gou, L. Liu, and J. Saunders. Recent advances in the exploratory analysis of interregional flows in space and time. 'Innovations in GIS 4: selected papers from the Fourth National Conference on GIS Research UK (GISRUK), CRC, 75, 1997.

[7] D. Guo. Flow mapping and multivariate visualization of large spatial interaction data. IEEE Transactions on Visualization and Computer Graphics, 15:1041—1048, 2009.

[8] N. Andrienko, G. Andrienko, and P. Gatalsky. Supporting visual exploration of object movement. Proceedings of the Working Conference on Advanced Visual Interfaces (Palermo, Italy), V. Di Gesù, S. Levialdi, L. Tarantino eds., ACM Press, pages 217—220. 2000.

[9] G. Andrienko, N. Andrienko, and S. Wrobel. Visual analytics tools for analysis of movement data. ACM SIGKDD Explorations Newsletter, 9:38—46, 2007.

[10] J. Dykes and D. Mountain. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. Computational Statistics & Data Analysis, 43:581—603, 2003.

[11] W. Tobler. Automation and cartography. Geographical Review, 526—534, 1959.

[12] W. Tobler. Migration fields. Population, 215—232, 1978.

[13] W. Tobler. Experiments in migration mapping by computer. Cartography and Geographic Information Science, 14:155—163, 1987.

[14] J. Wood, J. Dykes, and A. Slingsby. Visualisation of origins, destinations and flows with OD maps. The Cartographic Journal, 47:117—129, 2010.

[15] M. Noirhomme-Fraiture. Visualization of large data sets: the zoom star solution. International Electronic Journal of Symbolic Data Analysis, 0, 2002.

[16] E. Kandogan. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates'. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 107—116, 2001.

[17] S. Teoh and K. Ma. StarClass: Interactive visual classification using star coordinates. Proceedings of the 3rd SIAM International Conference on Data Mining, 2003.

[18] A. Klippel, F. Hardisty, R. Li, and C. Weaver. 'Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome? Cartographica: The International Journal for Geographic Information and Geovisualization, 44:217—231, 2009.

[19] A. Jain, M. Murty, and P. Flynn. Data clustering: a review. ACM Computing Surveys, 31:264—323, 1999.

[20] J. Reades, F. Calabrese, and C. Ratti. Eigenplaces: analysing cities using the space-time structure of the mobile phone network. Environment and Planning B: Planning and Design, 36:824—836, 2009.