Spatial-Social Network Visualization for Exploratory Data Analysis

Wei Luo
GeoVISTA Center
Department of Geography,
Penn State University
302 Walker Building
University Park, PA, U.S
wul132@psu.edu

Alan M. MacEachren
GeoVISTA Center
Department of Geography,
Penn State University
302 Walker Building
University Park, PA, U.S
maceachren@psu.edu

Peifeng Yin
PDA Group
Department of CSE,
Penn State University
337 IST Building
University Park, PA, U.S
pzy102@cse.psu.edu

Frank Hardisty
GeoVISTA Center
Department of Geography,
Penn State University
302 Walker Building
University Park, PA, U.S
Hardisty@psu.edu

ABSTRACT
There has been considerable interest in applying social network analysis methods to geographically embedded networks such as population migration and international trade. However, research is hampered by a lack of support for exploratory spatial-social network analysis in integrated tools. To bridge the gap, this research introduces a spatial-social network visualization tool, the *GeoSocialApp*, that supports the exploration of spatial-social networks among network, geographical, and attribute spaces. It also supports exploration of network attributes from community-level (clustering) to individual-level (network node measures). Using an international trade case study, this research shows that mixed methods — computational and visual — can enable discovery of complex patterns in large spatial-social network datasets in an effective and efficient way.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI)

General Terms
Management, Measurement, Design

Keywords
Spatial-Social Network, International Trade, Geovisualization, Geovisual Analytics

1. INTRODUCTION
Social network analysis is used to understand how relationships among actors (i.e., individuals, groups, or other social collectives) within a network affect the behaviors or attributes of themselves and other actors [18]. A growing number of researchers realize the potential application of this analysis method to geographically embedded networks and flows such as population migration [17], and international trade driven by globalization [11]. This increasing realization has called for development of integrated spatial-social tools to support spatially intuitive thinking in traditional social sciences [7].

Many network statistical algorithms exist to detect important individuals, relationships, and clusters, but it is difficult to understand their strictly quantitative outputs. Integrating network statistics with visualization methods has been effective in enabling users to make sense of data [16]. Thus, integrating visualization into spatial-social analysis systems can facilitate sensemaking and discovery of features such as distributions, patterns, and trends over social and spatial space.

The integration of computational methods into effective, interactive visualization in the context of geo-spatial environment gives birth to a new field: geovisual analytics. This new field aims to enable insights on large, complex data containing a geospatial component [1]. This paper introduces a novel geovisual analytics tool, the *GeoSocialApp*, designed to discover previously hidden patterns in complex datasets and to facilitate insight gain. The tool consists of three major components, each of which performs a specific task and can coordinate with other components to facilitate the insight gaining process. These major components include (a) network space, (b) geographic space, and (c) attribute space.

This paper is organized as follows. Section 2 reviews related literature. Section 3 presents an overview of the methods implemented. Section 4 presents the case study application of the tool. Section 5 discusses the limitations of the case study and tool and outlines the future work.

2. RELATED WORK
Current spatial-social network visualization tools can be classified into two major groups: the first group focuses on the spatializing network structures; the second group focuses on the combination of spatial analysis and social network analysis.

The first group considers spatial-social network visualization from a spatial perspective. For example, Guo [8] proposes an integrated interactive visualization framework, in order to effectively discover and visualize major flow patterns and multivariate relations from county-to-county migration data in the U.S. Wood et al.[19] propose an origins and destinations (OD) map to preserve all origin and destination locations of the spatial layout. Overall, this group develops new visualization and interaction techniques to synthesize, visualize, and discover patterns from very large spatial-social networks, but it lacks

ACM LBSN '11, November 1, 2011. Chicago, IL, USA
Copyright © 2011 ACM ISBN 978-1-4503-1033-8/11/11...$10.00

network analysis methods, which have the potential to gain deeper insight for users.

The second group of methods combines geospatial and network analysis with visualization (e.g., Sentinel Visualizer). Davis et al [5] introduce an integrated tool, connecting the Organizational Risk Analysis tool with Geospatial Information (Ora-GI) that allows the integration of geospatial data into the analysis of relational information. Both tools handle location-based spatial network relationships overlaid on a map without supporting network relationships among geographical units (e.g., countries and states). In addition, those tools focus on mathematical and statistical analysis, rather than visual data exploration.

Overall, the above work illustrates a trend to develop new visualization techniques to integrate social networks into a spatial framework, but current work has two major gaps: spatial-social visualization tools do not integrate network analysis methods to enable users to gain deep insight and no existing spatial-social visualization tool deals with network relationships among geographical units. This research aims to fill these gaps.

3. METHODS

GeoSocialApp incorporates network analysis and visualization in a geospatial framework. It is implemented using the GeoViz Toolkit (GVT), a component-based environment for developing multi-view, coordinated geovisualization applications [9].

The network component is based on the Java Universal Network/Graph (JUNG) Framework. JUNG is a JAVA API that provides a framework for modeling, analysis and visualization of relational data [15]. With JUNG, we implement two network views in GVT to facilitate visual exploration of patterns in social space and spatial space simultaneously, a Node-Link and a Dendrogram view (Figure 1). Below, we introduce the views, plus several other GVT components used in GeoSocialApp.

3.1 Node-Link View

The node-link view has two parts: the layout and individual-level network measures. The layout part supports the following layouts: KKL, FRL, Spring, Circle and Self-Organizing Map (SOM). Each layout is suitable for different network structures. The KKL layout implements the Kamada-Kawai algorithm [13], and the FRL implements the Fruchterman-Reingold force-directed algorithm [6]. The Spring layout models the edge as a spring and uses Hooke's law to determine the position of nodes in the visualization [10]. The KKL, FRL and Spring tend to place connected nodes together while unconnected components tend to be positioned far from each other. Thus, this kind of layout is suitable for a graph with strong communities and weak inter-community links. The circle layout organizes all graph nodes on a circle. It helps users find critical graph nodes in an intuitive way, by pointing out on first sight those nodes with many connections. The Circle layout visualization is suitable for social, computer, and other networks having critical nodes with high degree and complex among node connections. Finally the SOM implements a self-organizing map layout algorithm based on Meyer's self-organizing graph methods. SOM is closely related to force-based graph layout [14], but it is more efficient for large-scale graphs with an optimization algorithm.

Individual-level network measures include:

- Degree: the number of links connected directly by others.
- Betweenness: the extent to which a particular node lies between other nodes.
- Clustering coefficient: a measure of degree to which nodes in a graph tend to cluster together.

3.2 Dendrogram View

The dendrogram view implements the convergence of the iterated correlations (CONCOR) algorithm [3] to group actors with similar position in a single network or multiple social networks together. Actors in similar network positions are considered to have similar social influence in the relational network. The relation of each node with other nodes is represented as binary values (1=connected and 0=disconnected). Thus a node is associated with a binary string, which is treated as the sample data of that node. Based on these "sample data", CONCOR computes the Pearson product-moment correlation coefficients for every node pair; higher coefficients indicate a more similar position. Through iterative computation, these values will converge on either 1 or -1; the former indicates a strong similarity between positions of two nodes while the latter suggests a high difference. Nodes with a coefficient of 1 are classified into the same group, indicating their highly similar positions in the social network. Hierarchical structure can be achieved by running CONCOR on each subgroup.

The dendrogram view provides two layout methods to visualize the hierarchical structure of CONCOR results: a tree layout and a balloon layout (Figure 1). The tree layout positions child nodes under their common ancestors and organizes the graph in a hierarchical way. Also, an equivalent balloon view can be obtained from the tree by placing each node's children in its enclosing circle [4]. These two views are good at displaying hierarchical structure of the graph, and are especially adaptive for displaying clustering results as in our tool. There is a slider bar in the view to control the level of the dendrogram result.

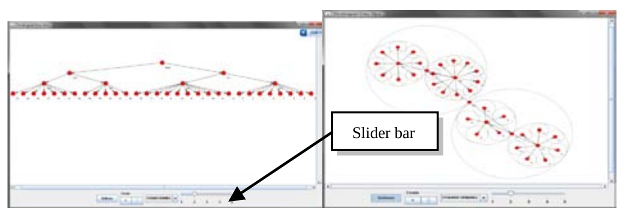

Figure 1. Dendrogram View

3.3 Components in GVT

Existing components in GVT used in this study include the choropleth map view, the histogram view, and the star plot view. The choropleth map view is used here for visual exploration in geographical space. The histogram depicts the distribution of univariate data. The star plot view is used to represent multiple variables (Figure 2). The number of variables corresponds to the number of rays emanating from the center of the star plot, and the length of each ray is proportional to the value it represents. For this research, basic network node measures calculated in the node-link view are imported into the histogram and star plot views, in order to explore the distribution of individual measures and the relationship among multiple measures.

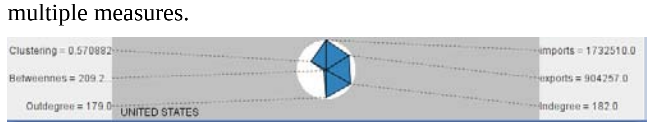

Figure 2. Star Plot View

3.4 Network-GVT View Coordination

The GeoSocialApp allows users to explore spatial-social networks in multiple spaces by supporting coordination among network views and GVT views, with each indicating different perspectives on the same datasets. The dendrogram view and nodelink view allow users to explore networks from the community-level (clusters) to the individual-level (node measures). The choropleth map view gives users an impression about the geographical positions. In addition, each node in the network views represents a geographical unit (i.e., states, countries) in the choropleth map; thus, the linked network views and geographical views allow the explicit interaction between geographical space and social space. Other components in GVT such as the histogram and the star plot allow users to explore the distribution of individual measures and the relationships among multiple measures. Therefore, coordination among network views and other GVT views allows users to explore spatial-social network data in network space, geographical space and attribute space, facilitating the insight gaining process.

4. WORLD TRADE NETWORK ANALYSIS: A CASE STUDY

We apply GeoSocialApp to world trade data analysis to demonstrate the potential of the methods to support economic research and policy-making. When using a network approach to study international trade, each country is typically considered to be a node of the network. International trade is usually measured by the monetary value of exports and imports between countries, so trading relationships are analogous to links in a network of country nodes. The focus here is on network structure, thus relationships, rather than the monetary values.

4.1 Data Preprocessing

We use the import and export data in current U.S. dollars among 192 countries in 2005 as a case study test of this tool. These data were extracted from the CorrelatesOfWar (COW) Database [2]. Countries are the nodes of the network and a link between two countries represents a trading relationship. We organize the data in matrix form with columns as exporting countries and rows as importing countries. As an illustration, Table 1 is the binary matrix for the first 10 countries in our data; “1” represents trade between countries, “0” represents no trade.

Table 1. Partial Binary Matrix for 0% Threshold in 2005

	AFGHANIST/	ALBANIA	ALGERIA	AMERICAI	ANDORRA	ANGOLA	ANTIGUA	ARGENTIN	ARMENIA	AUSTRALI
AFGHANISTAN	0	0	0	0	0	0	0	1	0	1
ALBANIA	0	0	1	0	0	0	0	1	1	1
ALGERIA	1	0	0	0	0	1	0	1	0	1
AMERICAN SAM	0	0	0	0	0	0	0	1	0	1
ANDORRA	0	0	0	0	0	0	0	0	0	0
ANGOLA	0	0	1	0	0	0	0	1	0	1
ANTIGUA AND E	0	0	0	0	0	0	0	1	0	1
ARGENTINA	1	1	1	0	0	0	0	0	0	1
ARMENIA	0	0	0	0	0	0	0	1	0	1
AUSTRALIA	1	1	1	1	0	1	1	1	1	0

4.2 Spatial-Social Network Analysis and Visualization

After converting international trade into the network format, we use the GeoSocialApp to explore it. Initially, we use the dendrogram view to divide the network data into two groups (Figure 3). After highlighting one group (yellow nodes in network views and blue outlines in map view), we find most countries in the highlighted group are economic periphery countries (i.e., most countries in Central America and Africa) and most countries in the other group are core and semi-periphery countries (i.e., North America and European Union). The bivariate choropleth map uses two variables: total values of exports and total values of imports for each country. The bivariate colors reinforce this classification: economically less-important countries are indicated by light green, whereas other, more important countries are indicated by dark green. The node-link view in Figure 3 also indicates the core and periphery relationship with the highlighted group on the periphery of the international trade network, because the KKL layout tends to put connected nodes together and separate unconnected components. From the exploration of the dendrogram, bivariate choropleth, and node-link views, we find that global trade is hierarchical with a core-periphery structure at higher levels of trade; this result matches previous research [12].

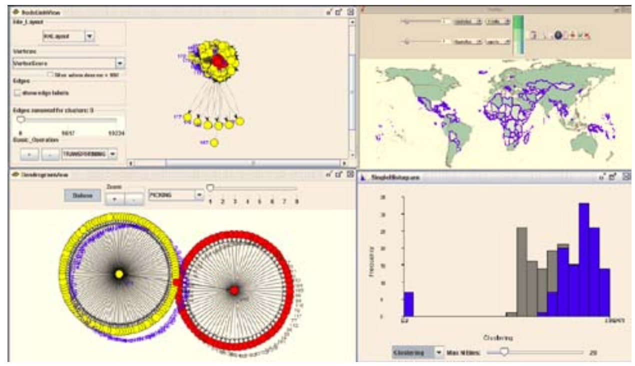

Figure 3. Screenshot with Node-Link, Dendrogram, Bivariate Choropleth Map and Histrogram Views. Large Pictures Available at: http://www.geovista.psu.edu/GeoSocialApp

Second, we are interested in the relationship between basic network node measures for countries and their positions in the world trade network. From the histogram view in Figure 3, we see that periphery countries (highlighted in blue) have a high cluster coefficient. This indicates that periphery countries tend to have international trade among each other, whereas a low coefficient indicates that more developed countries tend to have more trade partners than less developed countries. A few periphery country outliers do not follow the distribution of most other periphery countries. We continue to explore the underlying reason behind those outliers. In Figure 4, we highlight those outliers that have only one trade partner (one export link and one import link). One trade partner results in cluster coefficient of zero. Therefore, those countries are outliers that do not violate the result we find in Figure 3. The result is also supported by the star plot view (Figure 5). In this view, we represent six variables: imports, exports, in-degree, out-degree, betweenness, and clustering coefficient. All 192 countries are ranked from highest to lowest according to the first variable: import. It is obvious that most countries with a high clustering coefficient (long ray from the center of the stars) have a low value with other five variables. The negative relationship implies that rich countries may benefit more from more diversified trade partners and small economies may benefit more from concentrated trade partners. The result is also supported by Kali et al. [11]: the number of trading partners and the concentration of trade are both positively correlated with growth across all countries, but the former is greater for rich countries and the latter is concentrated in poor countries.

Results found using the spatial-social network visualization tool correspond to previous research [11-12]. Most past research has used traditional network measures and computational models, but their results are not easy for non-experts to understand. Our tool can give direct visual representation of patterns to a broader audience, facilitating insight development.

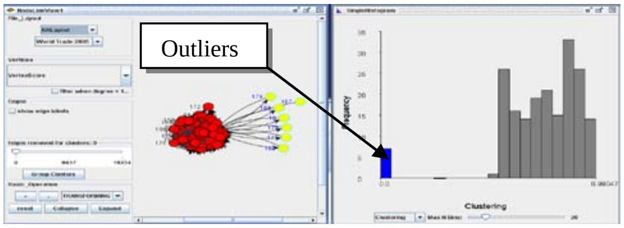

Figure 4.

Figure 5. Star Plot View

5. CONCLUSIONS

We introduce the GeoSocialApp that supports exploration of spatial-social networks among multiple spaces: geographical, network, and attribute space. It also supports the exploration of network attributes from community-level (positional clustering) to individual-level (network node measures). The brief case study presented illustrates that this tool can facilitate an insight gaining process about spatial-social networks underlying international trade. GeoSocialApp has been designed as general ones suited to any data aggregated to enumeration units. Thus the system is applicable not just to country-level data as shown here but can be used with data for individuals (e.g., data about social or other connections between individuals aggregated to census blocks in a city). Two possible extensions of this work in terms of application of the tools to trade analysis can be explored in the future: (a) increasing threshold values from zero monetary units enables us to understand the sensitivity of various topological characteristics of the network to differing trade magnitudes; (b) exploring multiple years of data can enable us to understand international trade over time.

6. ACKNOWLEDGMENTS

This material is based, in part, upon work supported by the U.S. Department of Homeland Security under Award #: 2009-ST-061-CI0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security; a grant from the Gates Foundation also provided partial support.

REFERENCES

[1] Andrienko, G., Andrienko, N., Keim, D., MacEachren, A. M. and Wrobel, S. 2011. Challenging Problems of Geospatial Visual Analytics (editorial introduction). *Journal of Visual Languages & Computing*, 22, 4, 251-256.

[2] Barbieri, K., Keshk, O. M. G. and Pollins, B. 2008. Correlates of war project trade data set codebook, Version 2.0. .

[3] Breiger, R., Boorman, S. and Arabie, P. 1975. An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12, 3, 328-383.

[4] Carriere, J. and Kazman, R. 1995. *Research report: Interacting with huge hierarchies: beyond cone trees*. In *Proceedings of the Proc. IEEE Information Visualization' 95* (1995).

[5] Davis, G., Olson, J. and Carley, K. 2008. *OraGIS and Loom: Spatial and temporal extensions to the ORA Analysis Platform*. Carnegie Mellon University.

[6] Fruchterman, T. and Reingold, E. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21, 11, 1129-1164.

[7] Goodchild, M. and Janelle, D. 2010. Toward critical spatial thinking in the social sciences and humanities. *GeoJournal*, 75, 1, 3-13.

[8] Guo, D. 2009. Flow mapping and multivariate visualization of large spatial interaction data. *Visualization and Computer Graphics, IEEE Transactions on*, 15, 6, 1041-1048.

[9] Hardisty, F. and Robinson, A. 2010. The geoviz toolkit: using component-oriented coordination methods for geographic visualization and analysis. *International Journal of Geographical Information Science*, 25, 191-210.

[10] Herman, I., Melançon, G. and Marshall, M. 2000. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6, 1, 24-43.

[11] Kali, R., Méndez, F. and Reyes, J. 2007. Trade structure and economic growth. *Journal of International Trade Economic Development*, 16, 2, 245-269.

[12] Kali, R. and Reyes, J. 2007. The architecture of globalization: a network approach to international economic integration. *Journal of International Business Studies*, 38, 4, 595-620.

[13] Kamada, T. and Kawai, S. 1989. An algorithm for drawing general undirected graphs. *Information processing letters*, 31, 12, 7-15.

[14] Meyer, B. 1998. *Self-organizing graphs—a neural network perspective of graph layout*. In *Proceedings of the Graph Drawing. 6th International Symposium, GD' 98. Proceedings* (Berlin, Germany, 1998). Springer-Verlag.

[15] O'Madadhain, J., Fisher, D., Smyth, P., White, S. and Boey, Y. 2005. Analysis and visualization of network data using JUNG. *Journal of Statistical Software*, 10, 1-35.

[16] Perer, A. and Shneiderman, B. 2008. *Integrating statistics and visualization: case studies of gaining clarity during exploratory data analysis*. In *Proceedings of the Proc. SIGCHI conference on Human factors in computing systems* (2008). ACM.

[17] Tobler, W. R. 1987. Experiments in migration mapping by computer. *Cartography and Geographic Information Science*, 14, 2, 155-163.

[18] Valente, T. 2010. *Social Networks and Health: Models, Methods, and Applications*. Oxford Univ Pr.

[19] Wood, J., Dykes, J. and Slingsby, A. 2010. Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal*, 47, 2, 117-129.