Mapping Networks: A New Method for Integrating Spatial and Network Data

Manish Nag
Princeton University

Abstract:

Spatial network mapping provides a new capability not addressed by current network graphing techniques and software. Through the use of systematic actor positioning and gradational tie coloring, spatial network mapping allows researchers to use network graphs to evaluate the spatial structure of social networks. Spatial network mapping is especially powerful in visualizing longitudinal network trends. The paper presents theoretical and empirical motivations for spatial network mapping, a detailed method, and an empirical component highlighting the method’s use for evaluating the spatial network structure of the world polity. The paper also introduces Sonoma, a software package that generates spatial network maps.

Send correspondence to: email: mnag@princeton.edu; mailing address: 106 Wallace Hall, Princeton University, Princeton, NJ 08544 USA; phone: 609-258-4530; fax: 609-258-2180.
Sociologists and geographers have long noted the power of space in determining the form and content of social relations (Simmel 1997, Gregory & Urry 1985, Soja 1985, Baldry 1999, Gieryn 2002, Taylor 2007). However, current network graphing techniques have not fully utilized visual space to depict the interplay between space and social relations. Current techniques in network graphing mostly bring closely tied actors into close spatial proximity. However, this technique prevents researchers from systematically positioning actors so that a graph’s spatial backdrop can itself serve as an analytical tool.

This paper provides a network graphing technique that allows for the systematic positioning of actors in geographic space, utilizing color as a means to express tie strength. In this manner, researchers can make multidimensional sense of actors, the spaces they inhabit, and the relationships that form across spaces. Highlighting the importance of spatial arrangement in social network visualizations (McGrath et al. 1997), and drawing on research that highlights the benefits of using GIS layers in analyzing social networks (Faust et al. 1999), this paper also introduces a new software tool to allow researchers to automate the creation of geospatial social network maps. The technique and tools described in this paper can provide immediate benefits to investigations of the interaction of space and place upon localized outcomes in education, health, crime (Papachristos 2009), and organizational ecology, as well for investigations into global phenomena such as international trade, global cities, and the world polity.

A representative example of a social network graph without geospatial context can be found in Figure 2 of Bellotti’s (2008) Social Networks article on the friendship communities of singles in Milan. In this figure, the friendship network of a particular single, named Mauricio, is given. In this diagram, a set of highly interconnected friends known as “the family” are located in one subnetwork, while Mauricio’s few other friends are spatially positioned outside the “family”. As with most sociograms, space is used in a limited capacity—in this case to denote whether actors are inside or outside a particular subnetwork. However, the spatial arrangement in this case is ad hoc, used to illustrate particular points about the network. By systematically laying out actors according to geographic coordinates, or coordinates in
other conceptual spaces, network graphs gain the ability to facilitate network discovery as much as post facto communication of network attributes.

Roadmap for this paper

After discussing a working definition of space, this paper will explore the theoretical motivations for spatial visualizations of social networks by highlighting the existing literature that examines the intersection of social networks and space. The paper then discusses the empirical motivation for a new visualization method by reviewing the limitations of existing visualization approaches. Following this, I present a brief introduction to Sonoma (the Social NetwOrk Mapping Application), a software package that allows for the creation of geospatial network visualizations, and I then provide a detailed method for creating spatial visualizations of social networks, with an empirical example that shows how the method can be used a visual means for testing network hypotheses. I conclude with a discussion of results and the implications of my findings.

Defining space

When referring to space, I refer to the abstraction of geographic space. In this usage, geographic space means the set of points along the surface of a spherical coordinate system that is represented by latitude and longitude identifiers. The spherical coordinate system aligns sufficiently to the size and shape of the earth, and therefore allows us to locate social phenomena against spatial locations. Using the descriptive convention of spherical geometry\(^b\), social phenomena can be mapped so that spatial distributions of social phenomena can be explored. Where social phenomena have unique spatial distributions, there is an opening for further investigation into the causes. In this model, space is not

\(^b\) Like mathematician Henri Poincare, I do not seek to make any judgments on the true structure of space. Like Casati and Varzi (1999), I seek to stand clear of the debate between Newton and Leibniz on the ontology of space. Rather, the convention of spherical geometry is used to describe the topology of the Earth’s surface.
itself the cause of social phenomena, rather an important lens through which to gauge their distribution. As Simmel notes in “Sociology of Space”, “it is not the form of spatial proximity or distance that creates the special phenomena of neighborliness or foreignness, no matter how irrefutable these might seem” (Simmel 1997).

This definition of space also distinguishes itself from place. As Tom Gieryn (2000) has noted, place is constituted by a unique location in the universe that possesses physicality, and is invested with meaning by people. The word space should not be confused with place, as geographic space in and of itself has no physicality. Since space is an abstraction wherein a location in geographic space corresponds to a physical location on the Earth, there can be a close correspondence between findings made in geographic space and in physical places. However, findings of social phenomena in geographic space do not necessarily indicate that place is an important causal factor for that spatial distribution. Only further research into the causes of spatial distribution of social phenomena can determine the causation of spatially distributed social phenomena.³

Conceiving of space as an abstraction also aligns with Downey’s work urging the use of GIS in sociology. As he writes, modeling space as a “continuous, unbounded surface” resolves longstanding spatial methodology problems, and opens the door for asking new questions and making new insights into physical space, spatial relationships, ecological context, and inequality (Downey 2006). Downey

³ I should also address the critique of abstract space that Henri Lefebvre puts forth in The Production of Space (1991). For Lefebvre, abstracting an “ideal” space from physical space parallels Descartes’ separation of cogito, the “mental self” from the practical, physical self. This abstraction separates the discourse of thought from the realities of power, and therefore reinforces a capitalist hegemony that uses this separation to continue a program of homogenization and social fragmentation in the name of economic production. I respond to this critique by first noting that our approach does not privilege “space” over “place”. It simply uses abstract space as an analytic tool that can provide valuable results for the further investigation of place. Used properly, mapping social networks across geographic space should allow sociologists to make more complex and meaningful connections between place and social phenomena. Lefebvre’s critique does highlight, however, the need for the investigatory vigilance required of a truly “reflexive” sociology (Bourdieu & Wacquant 1992).
notes that sociologists have long been tied to “discrete and bounded spatial units” that have facilitated the quantitative analysis of many social problems. However, many questions have been unsatisfactorily answered by such a discrete conception of space. Downey argues that people live in a world that is simultaneously discrete and continuous. Social analysis risks missing the distribution of particular phenomena if it takes as a given the administrative boundaries of current data sets, which assume certain spatial units such as census tracts. By taking a similar approach to space, the method in this paper allows investigators to freely move their analysis across varying levels of geographic scale, so that the distribution of social phenomena can be found across boundaries that have confined research to more national or regional boundaries.

**Theoretical motivations**

This paper follows in the path of existing sociological work that has investigated the intersection of social networks and space. The study of networks of friendship and social support has often noted the importance of space in forming networks, particularly within racial and class contexts, where access to a common space facilitates the random face-to-face connections that create social networks (Blau 1977, Wilson 1987, Fernandez & Harris 1992, Moody 2001). McPherson and Smith-Lovin, et al. (2001) have noted that “geographic propinquity” is among the factors that create contexts for the formation of homophilous social relations. Other research has explored the impact of space on the formation of social networks with regards to labor market outcomes. The “spatial mismatch” hypothesis has loomed large in the labor market literature. In their review of literature on space and the labor market, Fernandez and Su (2004) note that the spatial mismatch hypothesis argues that residential segregation and job decentralization combine to adversely affect the employment opportunities of minorities (Mouw 2000, Fernandez 2008). As Fernandez and Su (2004) also note, other major bodies of labor market literature focus on the expression of differing gender opportunities across space, while a last
body of literature focuses on spatial agglomeration, where economic opportunities are defined by the common spatial locations of firms in a given region (Saxenian 1994).

As the literatures on friendship and on the labor market demonstrate, interest exists in the current sociological literature on the intersection of space and social networks. As Fernandez and Su note, this interest is growing in the labor market literature. Other literatures also note the intersection of space and social networks. In economic geography, Storper (2004) has noted the vital role of face-to-face contact in urban economies, particularly in creative work. Recent literature reviews of medical sociology have begun to note the importance of social networks on health outcomes (Smith & Christakis 2008). Medical geography has also noted the impact of space on health outcomes, through examination of social networks and the “daily paths” and routines of groups like the homeless and HIV patients (Rowe 1990, Takahashi 2001). All these literatures share a common thread of face-to-face contacts’ impact on the formation of social networks, and that sharing a space plays an active role in forming and maintaining social networks of individuals.

As network analysis has scaled up to study the relations between larger groups, I propose that spatial mapping can also scale up. Sassen (2001) notes that agglomeration economies have also come to apply to “global cities”, where firms share a pool of expertise across a common business network. As specialized firms grow and become part of integrated production chain, city-to-city networks are formed that span national borders. Alderson and Beckfield (2004) use network analysis of cities to highlight the distribution of power and prestige in the “global urban hierarchy”. Snyder and Kick use network analysis to evaluate the applicability of world-system theory to studying varying rates of economic growth among nations (Snyder & Kick 1979). Clark and Beckfield further refine this analysis of the network of nations within a world-system framework, and find that network positioning has strong effects on outcomes of economic growth (Clark & Beckfield 2009). The visualization example in this paper will
demonstrate that spatial mapping applies to the networks of cities and countries as much as it applies to the networks of individuals.

As this review of existing literature has shown, interest is growing in the connection between space and social networks, exemplified in the literatures of sociology, economic geography, and medical geography. These new boundary-spanning disciplines no longer ignore the impact of space on the formation of social connections, nor do they limit space as the domain of geographers or urban studies scholars. As Downey has noted, many prior advances in quantitative sociology were based on the creation and analysis of data sets where geographic units were discrete in design. This focus on discrete spaces, ranging from census tracts to nations, facilitated new discoveries while inhibiting the observation of the interaction of continuous spaces with social outcomes. The advent of new GIS tools has allowed space to be considered more continuously, and in more depth than previously allowed. Researchers therefore have new empirical tools for describing space. With these tools in hand, connections between social life, geography, ecology and the built environment can be made available. The analysis of social life can therefore be grounded to the realities of physical space, whether natural or manmade. The geographic distribution of social phenomena such as social capital can also be made accessible.

When space has been considered in quantitative sociology, it has often focused on observing the interaction of geographic distance with social outcomes. This focus on geographic distance has limited the scope and power of spatial analysis in understanding social networks. This paper adds an additional analytical tool that broadens the way in which spatial connections can be made to social phenomena, so that the interplay of geographic distance and social proximity can be evaluated as distinct but connected forces. In the modern environment of globalization and network technologies,
geographic proximity does not directly translate to social proximity, and therefore visualization tools need to be able to simultaneously distinguish between these two forces.

**Empirical motivations**

Linton Freeman (2000) provides a thorough review of existing approaches for visualizing social network data. I use Freeman’s historical framework of visualizations to show how this paper builds upon and differs from existing social network visualization techniques.

Since Moreno’s pioneering work in the 1930’s, visualizations have held a place of privilege in social network analysis, both as a means for observing longitudinal network trends and for conveying network concepts. Two innovations made in the early decades of social network visualization are important for the method described in this paper. The first innovation, originally used by Moreno, is the use of color or line patterning to display different categorical differences in network ties. The second involves positioning actors based on locations they use in real life. A classic example from Moreno involves arranging a network of football players based on their starting position. Another visualization from rural ethnographers Leonard and Loomis (1941) charts the visiting patterns of the residents of El Cerrito, New Mexico by plotting households and their visiting patterns over a map of El Cerrito. Mapping actors based on geographic location is an example of what I will refer to as “absolute positioning”. In an absolute positioning scheme in a two-dimensional graph, the horizontal and vertical position given to an actor possesses meaning. In the case of the El Cerrito maps, the horizontal and vertical positioning corresponds to a household’s geographic position, conceptually represented by longitude and latitude.

Absolute positioning of actors is an important concept because it distinguishes this paper’s approach from the common network visualizations in use today. As Freeman (2005) notes, modern computational visualization packages rely mainly on multi-dimensional scaling (MDS) or singular value decomposition (SVD) approaches for placing actors in a network visualization. Both approaches seek to
create spatial distributions of actors that reflect patterns of social proximity in the data that is being represented. The relative positions of actors in these visualizations is important, but not the absolute positioning. Clusters of similar actors can be shifted on a graph without negatively affecting the usefulness of the visualization. In both MDS and SVD, social proximity is the driving force for the positioning of actors.

Current visualizations often utilize the concept of “spring embedding” layout algorithms. A network graph is generated through repeated iterations that place nodes randomly at first, and then subsequently reposition them to minimize the “energy” between nodes. Ties can be seen as “springs” that attract or repel nodes based on desired algorithm characteristics. For example, the Kamada-Kawai (1989) spring embedding algorithm seeks to minimize a graph’s total energy by minimizing the number of crossing ties while trying to maintain the distance between nodes, and the Fruchterman-Rheingold (1991) algorithm tries to maintain nodes in network cliques closely together and in separate areas on a graph.

Rather than using random graphing iterations where node positioning is ad hoc, this paper argues that social proximity can be reflected via the combination of network ties and color intensity to reflect tie strength. With social proximity reflected in the visualization of ties, the absolute positioning of actors can be used to segregate actors based on attributes other than social proximity. In the case of geographic mapping, actors can be placed according to the latitude and longitude of their geographic location. In this manner, patterns can be observed based on geographic attributes in combination with social proximity attributes. I argue that this approach efficiently uses visual space for the purposes of social network visualization. As noted by information designer Edward Tufte, visual information must work to “maximize the data to ink ratio within reason”. In other words, the amount of useful meaning that can be packed into a visualization should be maximized to whatever extent possible (Tufte 1983)
There are other benefits of utilizing network visualizations that show gradational strength of ties, and absolute positioning of actors. First, mapping a social network over geographic space allows a viewer to interact with a graph at simultaneously varying levels of scale. For example, a viewer can easily shift at a vertical scale between microscopic and macroscopic perspectives on a single region, while simultaneously shifting perspective at a horizontal scale, where disparate geographic units can be simultaneously compared against one another. Secondly, utilizing color to display the strength of different links allows the eye to discern patterns within a network structure. As the eye is sensitive to very subtle gradations in color, rich connections can be drawn from a network where ties are given different color values based on the strength of ties. This color sensitivity spans not only the full color spectrum, but also the grayscale spectrum, thereby allowing spatial visualization to appear in traditional black and white print formats, as well as full color online versions. Using commonalities in color, connections on a micro and macro scale can be simultaneously understood. Third, the use of absolute positioning and gradational tie coloring allows a viewer to make important longitudinal observations. When comparing network maps across time, one can discern the presence or absence over time of actors and links in certain regions. Longitudinal patterns in network strength can also be deduced by observing shifts in color intensity over time. Since actors are mapped to stable locations on a map, one can also easily compare the structure of networks around given actors. As James Moody (2005) notes, fixing the positions of actors enables the creation of “static flip books” to show network changes over time. Moody argues that static flip books make sense only on graphs where network relations are sparse. As the visualizations in this paper will show, even highly dense network visualizations of the world polity can hold up to the technique of flip books that are visualized across geographic space. Lastly, one should not ignore the intuitive, aesthetic appeal of an effective spatial visualization of a social network. A well-designed geographic map of social network can be appreciated by both sociologists and
laypeople. In dense networks, a well-designed map provides convincing evidence of network structures where many spring-embedded algorithms become unintelligible.

A practical example of spatial mapping

Having discussed the theoretical and empirical motivations for this paper, I present a method for creating geographic social network visualizations. By explaining the method in detail, I hope to show that these types of visualizations are within reach of social network researchers. A step by step approach will be used to detail the method. After each step, reference will be made to how the step was executed in constructing an actual project that investigated the networks of intergovernmental organizations (IGOs) from 1940 to 2000. Before explaining the method, it is important to give some context on the research project that will be used as an example for the method. The following draws on work by Beckfield (2008, 2009).

Previous studies of the world polity have noted the importance of international governmental organizations (IGOs) on world affairs. Paul Ingram and his colleagues have explored the impact of IGOs on international trade (Ingram et al. 2005). John Boli and George Thomas have noted how membership in IGOs seems to embody the very definition of being a “world actor” (Boli & Thomas 1999). John Meyer, Boli, et al. have forwarded the “world polity” theory, where many of the features of nation states are driven by a world culture, where no clear leader among nation states exists, but where global consensus around universal concepts of rational progress drive nation states to legitimize themselves through alignment to world culture. Given that the world polity is “stateless” as Meyer describes it, IGOs play a key role in ensuring alignment to world culture (Meyer et al 1997).

The purpose of the research example is to continue analysis of the social structure of the world polity and advance the methods of analysis. The research seeks to analyze the social network of the world polity at more of a disaggregate level than previous studies. Continuing research started by Jason
Beckfield on the structure and shape of the world polity (Beckfield 2008), this paper constructs visualizations of international networks to evaluate against theories of the world polity. Given that the world polity perspective sees wide-ranging impact of its theory on the social structure of the world polity, these impacts should be verifiable through the methods of social network analysis (Beckfield 2008). The method of spatial network mapping provides a means for testing hypotheses about the structure and shape of the world polity implied by any number of theories of international relations.

If we assume that the diffusion of world culture occurs in a global manner, one implication of world polity theory is that the network of nation-states is growing increasingly decentralized, as no one nation controls the universal concepts of world culture. Another implication is that the world is growing increasingly connected as world cultural models diffuse across the network of states. If a network is growing decentralized, then this fact should appear visually in the form of a network where stronger ties are not seen to be converging at certain epicenters in the network. Viewed longitudinally, the centrality of any epicenters should diminish visually, implying that epicenters should be less visible across time. Interconnectedness should appear on spatial network maps through an increase in the spread of ties, implying a reduction over time in areas of the map where ties are predominantly absent. Lastly, if world cultural models diffuse across the network of states, we should see a reduction in network regionalization, as states embrace world culture. The existence of regionalization will be verifiable visually through inspection of the networks for subnetworks that stand out within regions. The gradational coloring of ties allows for regional networks to be visualized against a large number of other ties in a global network visualization.

The question this research example examines is the spread of IGOs based on the geographic focus of their missions statements. Therefore, the network of IGOs is partitioned into three subnetworks: regional IGOs whose name and mission focus on a contiguous geographic region, inter-
regional IGOs, whose mission focus on inter-regional connections, and global IGOs. The data set we will use is the Correlates of War IGO data set v 2.1. The data set stores the set of IGO memberships from 1840 to 2000, with data stored as a matrix of countries and igo/year combinations. For this study, networks will be constructed where actors are individual countries. Network ties will be created between any countries that share a common IGO membership. To measure the strength of ties between actors, I measure tie strength by the number of IGO’s that two countries hold in common. Network visualizations will be constructed such that vertices are positioned at the capital city for a country, and ties are drawn between capital cities. Network ties will be drawn such that line widths and line color will be gradational based on tie strength.

A method and toolkit for creating spatial social network visualizations

Below I present a method for creating spatial mappings of social networks, with each step of the method followed by a discussion of its practical application on the IGO research project. The steps below do not require use of the Sonoma software, but the software provides a guided user interface that should increase the speed with which researchers can develop spatial visualizations. The software was constructed after investigations of NetDraw (Borgatti 2002) and Pajek (Batagelj & Mrvar 2002). Both packages provide network visualizations that can be used in conjunction with geographic images generated from GIS software. However, both packages did not support the ability to render tie strengths using color in a gradational manner, although NetDraw does allow for the variation of tie width by strength of tie. In addition, neither software package provided a simple way to easily position actors in geographic space. For both packages, positioning of actors requires translating spherical actor positions into two dimensional Cartesian space. As a result, researchers must use their own libraries to project spherical coordinates into a manner that NetDraw and Pajek understand. Sonoma makes the work of spherical projection simple. Researchers can provide the simple spherical positioning of actors, and the
software can automatically project actors onto numerous map projections. The ability to customize map projections is also a feature not available in other network visualization packages examined.

The Sonoma software allows researchers to design the visual aspects of a geographic map, as well as the visual attributes of the network graph that are overlaid on the map. Visual aspects that can be customized include map projections, the edges of the map, physical terrain features, the colors for land and sea, and other options described in the Appendix. Once data files are uploaded for the unimodal matrix of network ties and the geographic coordinates of actors, network maps can be previewed instantaneously. Sonoma allows for the iterative creation and revision of spatial network visualizations. A tool for defining color schemes also allows users to define a custom color scheme for representing gradational variation in tie strength. This tool allows users to choose a base color for a tie, and then vary the coloring of ties by hue (a color in the horizontal ROYGBIV spectrum), saturation (the presence or absence of color) or intensity (the lightness or darkness within a particular color) (Foley et. al 1997). Screenshots and further details on using Sonoma can be found in the Appendix.

**Step 1: Decide on which geographic projection to use for a spatial backdrop.**

Given a study’s requirements, researchers should choose an appropriate geographic projection that transforms the spherical map of the globe to a rectangular projection suitable for publishing via print or the web. If actors can be located on a map that spans an area smaller than a continent, the choice of a map projection can be reasonably flexible. However, when a map must be global to fit all actors, projections should be chosen consciously, as to avoid cartographic projections that reflect historical biases. For example, the often used Mercator projection shrinks the size of Africa and grows the size of Europe simply as function of the mathematical projection, which tends to grow land masses that are farther away from the equator. A useful resource for evaluating the applicability of geographic projections to particular research problems can be found in the USGS manual on map projections.
(Snyder, 1987), as well as the USGS online map projections poster (http://egsc.usgs.gov/isb/pubs/MapProjections/projections.html).

In the IGO research project, since the actors span the globe, a global projection was sought that minimized historical bias by representing the sizes of continents with an accurate sense of the relative sizes of continents. The Gall-Peters equal area cylindric projection was chosen due to its ability to properly represent the relative sizes of all countries.

**Step 2: Determine if other visual information layers are needed to give the visualization meaning**

Based on the scope of research, researchers should determine if additional layers of visual information would provide readers with the appropriate spatial context. Additional layers of information include administrative boundaries (cities, counties, states, countries, etc) as well as streets, buildings, bodies of water, or any other items that provide meaningful context for the research at hand. In addition, if the visualizations will be longitudinal, it is important to assess if any of the additional visual layers will change over time. For example, country boundaries evolve over time (Wimmer & Min 2006), as do built landscapes and bodies of water. In the case where visual layers change over time, separate visual layers will need to be constructed for each point in time that is being analyzed. These visual layers should be created in a GIS program, so that all visual artifacts can be georeferenced. ArcGIS is the recommended tool for creating additional visual layers for maps.

For the research example, an additional visual layer was created to depict the changing political boundaries of nations for each decade between 1940 and 2000. This spurred a research sub-project, as no GIS data existed that held the required data on historical boundaries. A method was devised using ArcGIS to trace historical political boundaries based on a set of historical map images from the past century. These “tracings” of maps were then georeferenced in ArcGIS. On ArcGIS, see the ESRI bibliography section (http://training.esri.com/campus/library/index.cfm) as well as GIS.com. Examples
of the visual layers for the changing political boundaries can be found by comparing Map 1 and Map 2 in the Appendix, where new boundaries can be seen in locations such as Germany, West Africa, and South Asia between 1940 and 1970.

Step 3: Devise color, line, and symbol schemes for spatial backdrop

Researchers should utilize color, line, and symbol schemes that enhance meaning for the visualization without distracting the viewer from the central focus of the map. In the case of network visualizations, the central focus should be the actors and ties. Citing Tufte, the use of color is an important tool that should not be overused. In addition, the color, line, shape, and size of objects that are not actors or ties should draw the viewer’s secondary attention and not serve as a distraction. This recommendation applies to any map boundaries, symbols, and text that form the backdrop for a map. An effective spatial backdrop should make important distinctions that describe the space, but the backdrop should not draw excessive attention.

For the research example, color choices were constrained to shades of gray for print publication. Within these constraints, colors were required that distinguished land masses from water. Given that the majority of the globe is water, and the fact that areas of water would not contain vertices for actors, the color white was chosen for water. This allowed any links that spanned oceans to be clearly visible. To distinguish land from water, gray was chosen for land. A shade of gray was chosen for land that was lighter than any potential ties. In this way, ties would be clearly visible across land masses. White was chosen for political boundaries, in order to be clearly visible against landmasses without drawing excessive attention from ties. As the visualization design only focused on countries and ties at a global scale, no additional symbols were required for the spatial backdrop. In other cases, map symbols can serve as important reference points for a visualization. Map artifacts such as buildings, streets, rivers can help give users the necessary context for making sense of a visualization.
Step 4: Devise symbolic, line and color schemes for representing actors, ties and their associated attributes.

After the design of the spatial backdrop, the symbolic, line, and color schemes should be devised for representing the foreground of the visualization: actors and ties. For both actors and ties, attributes that are both categorical and gradational can be represented visually. For actors, different colors and fill styles can differentiate categorical attributes, while color intensity and vertex size can be used to represent gradational attributes. For ties, line color and line patterns can be applied for categorical differentiation, while line width and color schemes can be used for gradational differentiation by hue, saturation, or intensity of color.

For the research example, it was decided not to represent vertices at all, as actors would be implied by the fact that all ties would begin and end at vertices located at the national capitals. Due to the high density of ties, it was decided that circles at national capitals would take up visual space that could be better used for ties. The primary purpose of vertices is to distinguish actors from one another. In these maps, countries distinguish themselves by their political boundaries and the unique latitude and longitude of their state capitals. Therefore, no symbols were used to represent vertices.

For ties, lines of the lowest width possible were created in order to maximize the number of ties visible on maps. A color scheme for tie strength was chosen where the strongest ties would show as black, and the weakest ties would appear as a shade of light gray that was slightly darker that the shade of grey used for land masses. Given that the maps would be longitudinal, the color representations of tie strengths remained constant across points in time. Therefore, the darkest shade of gray was reserved for the highest possible tie strength across the study’s time span, and the lightest shade was kept for the weakest tie strength possible across time. Varying color intensity was calculated by keeping hue and saturation constant at 0, while varying intensity from 1 to 255. The resulting RGB values were tuples of the form r/g/b where r=g=b and the value of r ranged from 1 to 255. For non grayscale color schemes
that utilized variation by intensity, hue and saturation would be set to constant values above 0, where
hue would not necessarily equal saturation. Intensity would then vary between 1 and 255, and be
translated into RGB color values using the formula in Figure 13.33 of Foley et al. (1997).

Step 5: Build a list of unique actors, and map them to spatial dimension attributes

After the primary visual design for the maps is completed, data on actors and their spatial
locations should be compiled. While actors can be given any number of spatial attributes in a non-
geographic spatial visualization (eg. Blau spaces), in a geographic visualization, spatial locations would
be represented by latitude and longitude pairs.

For the research example, a list of unique countries was derived from the COW IGO dataset v
2.1. Countries were given latitude and longitude values based on data in the CIA World Factbook. The
small minority of countries that had capital cities not found in the CIA World Factbook was applied
latitude and longitude values found using data from the Google Maps website.

Step 6: Devise tie attribute schemes that enable partitioning of ties into separate visualizations

For highly dense maps, or if research requirements demand partitioning visualizations, it
becomes important to assign criteria to ties that enable partitioning a map into sub-maps. Partitionable
attributes do not necessarily need to apply directly to ties. Attributes of actors can also be used for
partitioning ties. Let us consider the example of a nondirected bimodal network with actors with modes
A and B, where links can be represented by the set of vertices \{A1,B1\} and \{B1, A2\}. Unimodal
visualizations of this network can be made, by creating unimodal links such as \{A1,A2\}, where vertex B1
is simply a tie attribute describing the link between A1 and A2. Attributes of vertex B1, and all other
mode B actors can then be used to partition the ties that appear on sub-maps.
For the research project, the highly dense network of the world polity demanded the partitioning of links. As the network was bimodal where country-to-country links were intermediated by IGO’s, links were partitioned based on an attribute that applied to IGOs. The IGO attribute was then absorbed as a tie attribute for country-to-country links. IGOs were assigned the attribute of “Mission Type” based on mission data kept in COW IGO v. 2 Codebook that described the regional focus of an IGO. Using the unique value types of the “Mission Type” attribute, multiple sub-maps were generated for each decade, in this case for the Mission Types of “Single Region”, “Inter-Region”, and “Global”.

*Step 7: Utilizing automation, calculate ties, tie weights and tie colors for the visualization. Segregate data based on link attributes where applicable.*

Wherever possible, the calculation of ties, tie weights and tie colors should utilize automation in order to enable data processing and reprocessing when new information comes to light that requires changes to the calculation logic. In addition, automation greatly reduces the time to bring the visualizations of large data sets to light. With this paper’s introduction of the Spatial Mapping Toolkit, a simple user interface exists to create spatial network visualizations.

For the research example, a relational database system was created for the purpose of automated data processing, given that nearly a half million country-to-IGO dyads existed for the COW dataset between 1940 and 2000. The Perl programming language was used to implement an automation system that could execute data processing with little manual intervention. Use of relational databases requires understanding the SQL query language. For researchers comfortable with the R programming environment, SQL should not pose a challenge. Perl is a powerful yet easy-to-learn programming language that also does not require an extensive software background. Both relational databases and Perl run in PC, Mac, and Linux/UNIX environments, and can therefore integrate easily with the majority of research computing environments. Please note that Perl is not required for this method, and most data analysis can be facilitated using tools like Microsoft Excel and UCINET.
To start data processing, the COW IGO membership data was loaded as-is into the relational database. The COW data was presented as a bimodal matrix of countries and IGO’s for a given year. A set of relational database operations was executed to transform this bimodal matrix into a set of unimodal matrices. In total, 21 matrices were created through the automation for the seven decades from 1940 to 2000, and the three IGO Types of “Regional”, “Inter-Regional”, and “Global” that we visualized. Tie weights represented the number of IGO’s a pair of countries held in common.

**Step 8: Prepare data for consumption by mapping program**

After the data representing actors, ties, and visual attributes is calculated, the data should be prepared for consumption by the visualization package of choice. Huisman and van Duijn (2005) provide an overview of the network visualization packages most suitable for social network analysis. If Pajek is used, an additional step is needed to convert latitude and longitude values into positions on a rectangular coordinate system that Pajek expects. The requirements for absolute positioning in Pajek may not be the same for other network packages. (For a formula that roughly translates latitude and longitude values to Pajek’s expected rectangular coordinates, please contact the author). Once this translation is performed, x and y coordinate can be fed to Pajek as additional columns of data in a .net file’s “vertices” section. After a visualization is generated as an image in .eps format, image manipulation tools like Photoshop or GIMP can be used to overlay a network visualization over a pre-existing map image. For further instructions on map overlay, seek out the documentation of your respective visualization package.

For an excellent integrated set of tools for generating maps, the open-source Generic Mapping Tools (GMT) is recommended. This software combines the features of GIS software with a powerful and simple interface for generating network diagrams over map projections. GMT also possesses data-fed control over tie colors, a feature that was not found in a review of the software in Huisman and van
Duijn. GMT also plots all artifacts using longitude and latitude in both decimal degrees and degree-minutes-seconds format, and plots great arcs between locations, meaning that the shortest global path is always charted between locations.

The Sonoma software package uses GMT as its underlying map generation engine. By providing an organized workflow for designing maps and network graphs, the software allows for the rapid creation of spatial visualizations from standard unimodal network matrices as those used in UCINET. For more information on the use of the toolkit, see the Appendix.

Although ArcGIS is the standard software for GIS visualizations, a review of ArcGIS showed that a straightforward interface is not available for creating social network graphs over a map. For this reason I recommend the use of GMT. GMT runs as a simple command-line application, which facilitates automating the application for use in a multistep workflow. To feed arcs to GMT requires the preparation of a text input file. To generate an arc from one point to another requires the following instructions in the text file:

```
>> -W0.01p,208
13:14E,8:50S
13:20E,52:31N
```

The first parameter after “-W” refers to width of the arc. In this case I type .01p to tell the program to set the width of lines to .01 pixels. This ensures that lines take up the minimum space possible on maps. The second number after the comma on the first line refers to the color of the line. The next line of instruction provides the longitude and latitude for the starting point. The following line provides the longitude and latitude for the endpoint of the arc. That is all that is needed to generate a tie with variable colors using GMT. The commands above need to be repeated to create a set of multiple arcs.
from one text file. The application has a very mature set of features beyond the scope of this paper, and should be explored further by interested readers. In addition, GIS layers created in ArcGIS can be utilized using the shp2gmt utility, which can convert ArcGIS shapefiles where the base coordinate system is in WGS84 format. This can provide very helpful for integrating additional visual layers such as political boundaries or street maps.

For the research application, GMT was used for visualizations. The text file described above was generated from the Perl automation suite based on the data created in Step 7. A separate text file was created for every year and Mission Type combination. Within each file, all country-to-country dyads for the respective year and Mission Type were selected. For the year 1940, three unique files were created that store the arcs, 1940_Single.xy, 1940_Regional.xy and 1940_Other.xy. The naming convention for files was only used for internal purposes. GMT is agnostic with regards to file names. In addition, the shp2gmt utility was used to convert the seven political boundary files for the decades from 1940 to 2000 from ArcGIS format to a format understood by GMT. Converted border files were saved with the naming convention exemplified by the file name 1940_Borders.gmt.

**Step 9: Generate maps for each link attribute class, overlay any additional visual information layers**

With all visualization data prepared, maps are ready to be generated. Using GMT, multiple passes are run to generate different layers of the visualizations. The generation of the spatial backdrop occurs first, followed by the generation of arcs, and then any additional text. Files generated in GMT are postscript files. To be viewed on screen, these postscript files must be converted to PDF or another image format using tools like Adobe Distiller or the open source Ghostscript.

For the research example, this is the sequence of steps that were run from Window Command Line. The step by step explanation for the steps can be found in the lines starting with “#” that precede
each command. The steps below would need to be rerun for all years and tie attribute combinations that your research requires. The software programs (pscoast, psxy, and pstext) invoked in this sequence are all part of the GMT suite of tools. Pscoast plots maps, psxy plots arcs and symbols, and pstext plots text. Used in tandem, these tools generate the spatial backdrop, network graph, and captions for the maps.

Step 10: Review and Revise

The creation of visualizations should be viewed as an iterative process. Each step in the process provides new information with which to review previous decisions. Each step should be used as an opportunity to make revisions that improve the overall utility and impact of a visualization. In the area of data processing, visualizations can highlight data errors that preexisted, or were created during subsequent processing (Faust et al. 1997). In the area of visual presentation, the visual attributes of actors and ties can be changed as well as the spatial backdrop. For a full listing of the types of visualization parameters that can be adjusted, please see Appendix Figures 1, 2, and 3 in the description of Sonoma. Geographic parameters, as well as network graph parameters that can be adjusted are discussed in the Appendix.

For the research example, map output from initial revisions inspired subsequent changes. When it was hard to visually differentiate tie colors among darker shades of gray, the color range for network ties was changed to a lighter set of grays. The proliferation of ties in maps for the year 2000 also obscured any sight of political borders underneath, an issue which I resolved by moving political borders into the visual layer of network ties just below the top decile of tie weights. Lastly, highly weighted network ties were found not to be distinguishable against the proliferation of other lower weighted ties. Therefore, the formula for mapping tie weight to tie width was adjusted to subtly widen higher weight ties without obscuring any lesser weighted ties.
Results

The 21 maps used for the research example have been packaged into 3 PDF booklets, one for each IGO group. These booklets are ordered in chronological order and then analyzed on screen using the Acrobat Reader to easily switch back and forth across decades. Although sample versions of some of the maps are provided in the Appendix, researchers should interact with these maps online, where they can move across decades and regions to verify the trends observed in this article. The PDF files serve as an implementation of Moody’s idea of “static flip books”. PDF booklets are available online at http://furnaceblast.com/igos/booklets/.

The resulting visualizations have been evaluated against the empirical motivations for the use of absolute positioning of actors, gradational coloring of ties, and a deliberately designed spatial backdrop. The maps permit viewing at simultaneous multiple scales. When moving through vertical levels of scale, the 1940 “Global IGO” map (Appendix C Map 7, Mission_Global_Sequence.pdf page 1 online) highlights that at the national level South Africa has numerous links with Europe, while at the continental level the rest of Africa holds few links, with exceptions only in Egypt, Ethiopia, and Liberia (a visual representation of the amount of Africa under imperial control in 1940). In addition, vertical scale can be changed by zooming in on the PDF up to 400%, where the complex of links in Europe or South America can be evaluated at close range. Evaluating the map horizontal levels of scale, one can compare Africa’s sparseness of ties with the similar sparseness in Asia. Asia and Africa primarily link to Europe, although Asia-Europe ties are stronger that Africa-Europe ties. Gradational colors of ties also serve as useful indicators for understanding the strength of regional ties. The countries of South America show significant linkages to Europe while showing weaker links amongst each other. When viewing the maps longitudinally, absolute positioning and tie color serve as the frame of reference as a visualization moves through time. This process can be viewed online by scrolling through the pages of PDF file. On the
Global IGO map sequence, an epicenter of connection can be observed between Europe and the rest of the world. This epicenter has increased in strength while the rest of the network has also increased in strength, leaving links to the epicenter well visible in the crowded network map of 2000 (Appendix C Map 9). However, at the same time the global IGO network has grown more interconnected. With each passing decade, the global IGO map has grown consistently darker overall, and regions like Asia, Africa and the Pacific receive new interconnecting links. In addition, on the Single Region map sequence (Appendix C Map 1-3, Mission_SingleRegion_Sequence.pdf), a subnetwork in Western and Southern Africa emerges in 1960 that continuously grows wider and stronger through 2000, while similar trends occur within Europe and the North African/Middle Eastern region.

Based on the results obtained from spatial network mapping, we do find evidence of growing interconnection especially in the network of IGO’s with a global mission. The global spread of network ties is apparent in the lack of spatial gaps found in the 2000 global map. However, our evidence does seem to show that the centrality of IGO epicenters in the Global North is still quite prominent. In addition, regional networks have not declined, but have instead grown significantly the network of single region IGO’s in areas like Western Africa and the North African/Middle East region. These findings suggest that a simple story of the uniform global diffusion of world culture does not bear up to evidence. The visually demonstrated persistence of epicenters in the Global North and the growth of regionalism in areas of the Global South demonstrate that spatial network mapping provides compelling evidence to judge current theories of globalization. In the case of persistent epicenters, we find that the spatial accumulation of advantage in the Global North seems to carry forward from 1940 to 2000. In the case of growing regionalization, we find that spatial proximity still seems to matter in the network of the world polity. Spatial network mapping therefore provides a tool to meaningfully evaluate the interaction between space and the formation, continuation, and change of social network structures.
Conclusion

Spatial mappings show the complex interplay between geographic space and social networks. Spatial network maps are differentiated from common computer-generated visualizations by the use of absolute positioning of actors, and continuous color coding of ties. Tools now exist to create spatial network maps using off-the-shelf technologies for databases and visualization that are well within the grasp of social scientists. The use of software automation will prove very important for social researchers, who will need to be able to process large data sets efficiently, and respond rapidly to new findings in the field. Also, as the step-by-step method illustrates, a combination of technical and visual design skills are required to create successful visualizations, but these skills are within reach for social network analysts with the advent of tools like Sonoma.

There is a distinctly intuitive accessibility to pictorial representations of information, and this has been a subject of interest in cognitive science, psychology, as well as aesthetics (Gibson 1978, Rollins 1999). The ability to spot network structures visually between regions and across time on a spatial network map is accessible to researchers and non-researchers alike. The visual, intuitive impact of spatial visualizations can serve as an instrument to help bridge the space between quantitative and qualitative sociology, as well as the four-part division of labor within sociology that Michael Burawoy highlights in his 2004 ASR presidential address, while also serving as a powerful instrument for Burawoy’s “public sociologies” that seek to connect the academic discipline to numerous publics (Burawoy 2005).

Looking forward, one of the first next steps will be to integrate spatial mapping analyses within existing quantitative frameworks (Wasserman and Faust 1994, Carrington et al 2005). For example, the subnetworks identified by spatial network mapping can tell us that certain subnetworks stand out. If ties appear stronger in the subnetwork, studies should be able to gauge how much stronger these ties are by...
calculating average strength scores for the subnetwork compared to the rest of the network. If ties and actors are absent in a region, studies should be able to quantify the absence by studying network density that includes actors who are not part of the network. Studies should also be able to quantify aspects of the subnetwork’s structure, such as centrality or average path length to understand the connectivity of subnetworks. If maps show changes occurring over time, quantitative tools should be able to tell us about the pace and amount of change.

The ability to gauge the effectiveness of geospatial mapping to convey specific network concepts should also be evaluated in the future. Using methods similar to McGrath, Blythe, and Krackhardt (1997), users should be given different visual variations of network graphs, with geospatial variations included, and then be asked to comment on their observations. Using this experimental technique which borrows from social psychology, the situations where geospatial mappings are most effective can be ascertained. Using web-based experiments, such investigations can now be accomplished using unprecedented numbers of research subjects (Skitka & Sargis 2006, Salganik et al. 2006).

To assist research, geographic mapping should be more closely integrated with network analysis packages like UCINET and Pajek. Far more interesting research results will be achieved when GIS data can be integrated directly within a network analysis package’s interface. In this way, vertices and ties could be manipulated in a package like UCINET while taking advantage of the software’s pre-existing quantitative measurement tools. Although it might seem that this would violate the notion of “absolute positioning” for spatial mapping, subway route maps serve as an example where geographic correctness is often violated in favor of clearly showing the relations between transit stops and rail lines. Such “pseudo-spatial” mappings could also be an interesting direction to take spatial network visualizations, especially when the density of actors and ties makes visualizing difficult. Such “pseudo-spatial” mapping
algorithms could even make their way into the existing network drawing algorithms. Also, network tools should adopt the ability to continuously color code ties between actors. Enabling this feature in conjunction with absolute positioning of actors will allow for the spatial mapping not just in geographic space, but other conceptual spaces such as “Blau space” (McPherson 2004).

Spatial mapping provides a new capability not addressed by current network graphing techniques. By positioning actors systematically in space, and by utilizing color to display tie strengths, spatial network mapping allows researchers to make multidimensional sense of the interaction between space and social networks. Spatial network mapping is equally suited to mapping international networks and networks of individuals. Spatial network mapping is particularly suited when visualizing longitudinal data. Researchers examining local and global outcomes can benefit from this analytical tool when researching outcomes as diverse as health, education, crime, the world polity, international trade, and global cities. In an environment where the interaction of social life and space changes rapidly, spatial network mapping provides an additional tool for analyzing and presenting network data.
References


Central Intelligence Agency. The World Factbook. Available at:

Clark, R., Beckfield, J.. 2009. A New Trichotomous Measure of World-system Position Using the

Downey, L., 2006. Using geographic information systems to reconceptualize spatial relationships and

ESRI, Inc. ESRI GIS and Mapping Software. Available at:

ESRI, Inc. GIS.com. Available at: http://www.gis.com

ESRI, Inc. GIS Training and Education Library. Available at:
http://training.esri.com/campus/library/index.cfm

Faust, K., et al., 1999. Spatial arrangement of social and economic networks among villages in Nang
Rong District, Thailand. Social Networks 21, 311-337.

P.J., et al. (Eds.), Models and Methods in Social Network Analysis. Cambridge University Press, New York,
NY, pp. 117-147.

569.

Social Science Research 37, 953-975.


Freeman, L.C., 2005. Graphic techniques for exploring social network data. In: Carrington PJ, et al. (Eds.),
Models and Methods in Social Network Analysis. Cambridge University Press, New York, NY, pp. 248-
269.

Fruchterman, T., Reingold, E., 1991. Graph drawing by force-directed placement. Software Practice and
Experience 21, 1129-1164.


Harary, F., 1969. Graph Theory. Addison Wesley, Reading, MA.


Tonkiss, F., 2005. Space, the City and Social Theory. Polity Press, Malden, MA.


Appendix

Appendix A: Instructions for Using Sonoma

Figure 1. Design Map Screen

This is the initial screen that appears when users startup Sonoma. From this screen, users may generate a new map by selecting File -> New from the main menu. To open an existing map project, users may select File -> Open.
Figure 2. Design Map Screen

This screen of Sonoma allows for the designing of the cartographic layer of a visualization (also known as the “base map”). From this screen, the type of geographic projection can be chosen. Once the projection is chosen, projection specific attributes can be entered, such as the central longitude of focus, and any relevant latitudes for the projection. Fields required for a given projection automatically have their labels set in boldface type.

Once projection-level attributes are chosen, a user can specify the region of the globe to depict using the “Map Edge” parameters to specify the latitude and longitude of the map’s edges. The colors for filling in land masses and areas of water can then be specified using a color chooser tool. The type of rivers to depict (i.e. major or minor rivers) can then be chosen and depicted using the color of water chosen previously. The type of political boundary to display (eq. national or state) can then be chosen, along with the width and color of political boundary lines. An option to show or hide the display coastlines as lines is also provided, along with the option to choose a color and width for coastlines. A field to specify the minimum area of map features is also provided so that map features such as mountains and/or lakes can be shown or hidden as desired. Coordinate gridlines can also be chosen for display, along with options for setting the frequency of gridlines, and optional labels that specify the latitude and longitude of gridlines. Lastly, additional graphical layers can be overlaid on the map using the “additional visual layer fields”. For these fields, commands from Generic Mapping Tools (GMT) can be executed and overlaid upon the base map. Please see the GMT manual (http://gmt.soest.hawaii.edu/) for more information.
Figure 3. Load Network Data Screen

Using this screen, users can load data files for their networks. The “tie file” should be a matrix file in the form of comma-delimited text, or a dyad file in the form of comma delimited text.

For matrix files, the first line of the tie file should be a comma separated list of actors, where the first column is blank. Subsequent lines of data in the tie file each start with an actor name followed by a comma, and then a comma—separated list of tie values when a pair of actors possess a network tie (see Figure 5 below). Tie weights must be represented as integer values.

For dyad format files, each dyad should appear on a single row with three columns of data. The first two columns hold the two actors involved in the dyad, while the third column holds the tie weight in integer form.

The actor file is a comma delimited list of actors, latitudes, and longitudes. The first line of the file should be the text “Actor, Latitude, Longitude”. Each subsequent line represents the actor name, and corresponding latitude and longitude for a single actor. Actor names in the actor file must match the names provide along the first row and column of the tie file.
Figure 4. Design Network Graph Screen

This screen of Sonoma allows for the designing of the network graph layer. A user can specify the shape of vertices, along with fill color and diameter for all vertices. For network ties, the option to specify a tie color scheme is provided (see the Define Color Scheme screen). The color scheme allows for the gradational coloring of ties based on tie weight. In addition, the type of scaling for mapping (e.g. linear or cubic) tie weights to line colors can be set. To calculate color scaling, the input range (i.e. the minimum and maximum) values of tie weights must be specified. Similarly, gradational thickness of ties can be set based on tie weight. Scaling type and input range can be set for tie widths similarly to tie colors. An output range (in pixels) is also provided for tie weights to provide a minimum and maximum value for the possible widths of ties based on the tie weights found in the user’s data. Output range is not provided as a text field for color scaling because the “color scheme” tool automatically provides the output range for coloring ties.
This screen of Sonoma allows for the designing of the color scheme for coloring ties based on the tie weights found in the user’s network data. The abstract space of colors can be described in three dimensions—hue, saturation, and intensity. The “select base color” square represents the matrix of hue and saturation values across the horizontal and vertical axes respectively. Below the color square, a slider allows for setting the third dimension of intensity on the Select Base Color square. By clicking on a point in the square, the user can choose the base color for the color scheme. While rolling over the color square, text below the square will display the current hue, saturation, and intensity values being navigated over. Once a base color is chosen, the “color scheme” bar on the right will be filled. By default, the base color will be shown with varying levels of intensity. The type of color variation can be reset to either hue, saturation, or intensity by using the “Vary Color By” dropdown. Based on the choice of this dropdown, the color scheme bar on the right side will change accordingly. Sliders on the top and bottom of the color scheme box allow for the setting of colors for the minimum and maximum tie weights. Since both sliders can be moved to either end of the color scheme box, users can specify that colors depict weights in either ascending or descending order of variation (i.e. whether colors go from light to dark in intensity, or dark to light in intensity) The numeric fields that represent the color scheme selected will be displayed below the color scheme box. Users can press the “Ok” button to return to the Design Network Graph screen.

Figure 5. Design Color Scheme Screen
Figure 6. Tie File

The file above represents a sample tie file in matrix format. The file is a unimodal version of the network data provided by Katherine Faust in her chapter on affiliation networks in *Models and Methods in Social Network Analysis* (2005). Sample data can be found as part of the “samples” folder included with the SNMT installation.

Figure 7. Actor File

The file above is an example of an actor file that provides actor names (country names) along with the latitude and longitude of each country’s capital city.
Figure 8. Map View Screen

This screen appears when a user selects the “View Map” button at the bottom of any of the main screens. If the user has not selected an actor or tie file, the preview window will only show an image of the map layer of the visualization. If both files are provided, and all map and network attributes entered, the map and network graph layers appear in the window. Files are presented in PNG format, and can be saved and repurposed from this window by right clicking on the image. The image above represents the data from the Faust example mentioned in Figure 5. An orthographic projection of the globe is used where the longitude 0 value is set to -90 degrees, and where ties vary in color by intensity in a linear fashion, where darker colors represent stronger country-to-country ties.
Appendix B: Transformation of the COW data set

The COW data is in matrix format, where the list of countries for each year sits on the horizontal axis and the list of IGO-Year combinations sits on the vertical axis. This format is most space efficient to track country memberships in an IGO. Each row in the IGO dataset can be represented with the set of tuples (associated sets of name-value pairs) named C, composed of tuples \( C_{1..n} \) where \( C_i \) is of the form \{Year, Igo, Afghanistan, Algeria, Albania,...,Zambia, Zanzibar, Zimbabwe\}. For each country attribute in \( C \), an integer value was stored that reflected the “membership status” of the country for that IGO in the given year. For relational database processing, the matrix was transformed so that a separate row of data was generated for each country-to-IGO dyad in a particular year. In order to implement this transformation, the list of COW countries was extracted manually from COW data files, creating list D, composed of elements \( D_{1..n} \) where the elements are part of the list of COW countries. List D was then fed into a Perl program that queried the tuples in set C, finding the subset E composed of tuples \( E_{1..n} \) where \( E_i \) is of the form \{Year, Igo, CountryD\}, where CountryD is named \( D\_i \), and the value \( E_i\{\text{CountryD}\}=1 \), where 1 represents the COW coding for full membership in an IGO. For each tuple in set E, a new row was inserted into a new relational table called “COW_Memberships” in the form of tuple \{Year, Igo, Country\}. In this manner, the original COW matrix of Countries and IGO-Years, was transformed into normalized relational form, where a unique row existed for each valid combination of COW, Year, and IGO value. The dataset in relational table COW_Memberships at this point is represented by the set Q, composed of tuples \( Q_{1..n} \) where \( Q_i \) is of the form \{Year, Country, Igo\}.

Once the matrix of IGO memberships was transformed it was time to supplement set Q with data on "Mission Type". A separate relational table was created for the unique list of IGOs, and for each IGO, a “Mission Type” value was stored. This set of IGO codings is represented as set I, where \( I_{1..n} \) are tuples of the form \{Igo, MissionType\}. In order to overlay Mission Type data into set Q, a Cartesian product was taken, represented by \( Q \times I \), where \( Q_i\{Igo\}=I_j\{Igo\} \). In relational database terms, the
The preceding step is known as a relational join. The resulting set is of the form \( R \), composed of tuples \( R_{1..n} \) where tuple \( R_x \) is of the form \{Year, Country, Igo, MissionType\}.

Once this transformation was completed, it was time to transform country-to-IGO links into country-to-country links. To create country-to-country dyads out of the set of country-to-IGO dyads, a cross-product was taken between the set of country-to-IGO dyads and itself. This cross product is represented by \( R \times R' \), where \( R = R' \), and where \( R_x\{\text{Year}\} = R'_y\{\text{Year}\} \), \( R_x\{\text{Igo}\} = R'_y\{\text{Igo}\} \) and \( R_x\{\text{Country}\} < R'_y\{\text{Country}\} \). This resulted in a dataset \( S \), composed of tuples \( S_{1..n} \) where \( S_x \) is of the form \{MissionType, Year, CountryR, Igo, CountryR'\}. The preceding steps created a dataset where countries were only connected once when they shared a common IGO membership in the same year. The data set was then collapsed so that country-to-country pairs are retained, with a count kept of the number of common IGOs held in common between two countries for a given year and common mission type. This is represented by the creation of set \( T \), where \( T \) is a transformation of \( S \) where a single tuple \( T_x \) is created for all tuples in \( S \) where \( S_x\{\text{MissionType}\} = S_y\{\text{MissionType}\} \), \( S_x\{\text{Year}\} = S_y\{\text{Year}\} \), \( S_x\{\text{CountryR}\} = S_y\{\text{CountryR}\} \) and \( S_x\{\text{CountryR}'\} = S_y\{\text{CountryR}'\} \). The resulting set is composed of tuples \( T_{1..n} \) where \( T_x \) is of the form \{Year, MissionType, CountryR, CountryR', TieStrength\}, where TieStrength is the count of all IGO’s of a given MissionType that two countries hold in common during a given year. The ease of relational technology was especially powerful here, because collapsing data involved simply using a “group by” clause in SQL that is a standard part of relational query grammar.

After the set of country-to-country dyads was constructed, I overlaid this data with the set of spatial attributes for actors that was constructed in Step 5. The set of unique countries and their spatial attributes is represented by the set \( B \), composed of tuples \( B_{1..n} \) where \( B_x \) is of the form \{Country, Latitude, Longitude\}. A double cross product is taken where set \( T' \) is formed from \( T \times B \times B' \), where \( B = B' \). A subset is taken of \( T' \) where \( T'_{xw}\{\text{Country-R}\} = B_x\{\text{Country}\} \), and \( T'_{y'}\{\text{Country-R}'\} = B'_y\{\text{Country}\} \).
resulting set formed is represented by U, composed of tuples U₁..ₙ where Uᵢ is of the form \{Year, MissionType, CountryR, CountryR’, TieStrength, CountryRLatitude, CountryRLongitude, CountryR’Latitude, CountryR’Longitude\}.

Lastly, colors for tie strengths were calculated based on an analysis of set U. Colors are represented as RGB color tuples where each of three segments holds a numeric value between 0 and 255, where 0 represents the absence of color (darkness) and 255 represents the full intensity of the colors red, green, and blue. Values of gray are represented by one integer between 0 and 255, where Rᵢ=Gᵢ=Bᵢ and 0 indicates black and 255 indicates white. Therefore, only one number was calculated to obtain color values for ties. Given the work done in step 4, the range of color values for tie strengths should fit between the predefined shades of gray that represent maximum and minimum tie strengths. These respective color values are represented as ColorMax and ColorMin. The maximum and minimum values for TieStrength were obtained from the full dataset for the years 1940 through 2000, represented by TieStrengthMax and TieStrengthMin. Then, a new attribute TieColor was added to the tuples in set U where for all tuples Uᵢ{TieColor} = int(ColorMin – ((Uᵢ{TieStrength} - TieStrengthMin)*((ColorMin-ColorMax)/(TieStrengthMax-TieStrengthMin)))).

The final set of tuples U was saved in a new relational table called “COW_Ties”. The set of tuples V in this relational table was composed of tuples V₁..ₙ, where Vᵢ was of form \{Year, MissionType, CountryR, CountryR’, TieStrength, TieColor, CountryRLatitude, CountryRLongitude, CountryR’Longitude, CountryR’Latitude\}. This table is now ready to support transmitting the data on actors and ties to the visualization software.
Appendix C: Research example output

Map 1. 1940 Single Region

Map 2. 1970 Single Region
Map 3. 2000 Single Region

Map 4. 1940 Inter-Region
Map 5. 1970 Inter-Region

Map 6. 2000 Inter-Region
Map 7. 1940 Global

1940: Global

Map 8. 1970 Global

1970: Global

Map 9. 2000 Global