The Senators Problem: A Design Space of Node Placement Methods for Geospatial Network Visualization

Arnav Mardia
Georgia Institute of Technology, Atlanta, GA, USA

Sichen Jin
Georgia Institute of Technology, Atlanta, GA, USA

Kathleen M. Carley
Carnegie Mellon University, Pittsburgh, PA, USA

Yu-Ru Lin
University of Pittsburgh, PA, USA

Zachary P. Neal
Michigan State University, East Lansing, MI, USA

Patrick Park
Carnegie Mellon University, Pittsburgh, PA, USA

Clio Andris[1]
Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Geographic network visualizations often require assigning nodes to geographic coordinates, but this can be challenging when precise node locations are undefined. We explore this problem using U.S. senators as a case study. Each state has two senators, and thus it is difficult to assign clear individual locations. We devise eight different node placement strategies ranging from geometric approaches such as state centroids and longest axis midpoints to data-driven methods using population centers and home office locations. Through expert evaluation, we found that specific coordinates such as senators' office locations and state centroids are preferred strategies, while random placements and the longest axis method are least favored. The findings also highlight the importance of aligning node placement with research goals and avoiding potentially misleading encodings. This paper contributes to future advancements in geospatial network visualization software development and aims to facilitate more effective exploratory spatial data analysis.

2012 ACM Subject Classification Human-centered computing → Geographic visualization; Human-centered computing → Graph drawings

Keywords and phrases Spatial networks, Political networks, Social networks, Geovisualization, Node placement

Digital Object Identifier 10.4230/LIPIcs.COSIT.2024.19

Category Short Paper

Funding The following authors acknowledge funding for this project.
Clio Andris and Sichen Jin: National Science Foundation #2045271.
Kathleen M. Carley: Office of Naval Research (ONR) #N000142112229.
Yu-Ru Lin: AFOSR, ONR, Minerva, NSF #2318461, and Pitt Cyber Institute's PCAG awards. Any opinions, findings, and conclusions or recommendations expressed in this material do not necessarily reflect the views of the funding sources.

[1] Corresponding author: Clio Andris, School of City and Regional Planning, School of Interactive Computing, 245 Fourth St. NW, Georgia Institute of Technology, Atlanta GA 30332, clio@gatech.edu

1 Introduction

Collaboration and agreement among agents in a network is a topic of interest in the social and behavioral sciences. In political networks, for example, nodes may signify legislators and their edges may signify bill agreements, bill co-sponsorship, or committee co-membership. By looking at these networks, researchers can see who is at the center or periphery of the network, whether people are collaborating "across the aisle", how the networks change over time, and where power is centralized in the network.

Most network analyses of collaborating agents are performed without considering **geographic** space. Yet, when spatializing networks by geolocating nodes (and thus, the edges) on a map, new questions can be answered. In a network of political figures, these include questions such as which constituencies' representatives collaborate, whether rural or urban legislators collaborate, whether nearby legislators or legislators from adjacent districts collaborate, and whether intra-state legislators collaborate more frequently than those from different states.

To visualize and analyze a geographic social network, the nodes must be pinned to a specific longitude and latitude (i.e., geographic coordinates). Sometimes, assigning node to coordinates is straightforward, as perhaps the node may have a specific street address. Other times, when absolute location is unknown there are commonly accepted defaults. For instance, a county commissioner may be represented as a point at the centroid (geographic center) of their county. Similarly, in visualizations of flow data, the origin and destination points are often positioned by the centroid of polygonal units such as counties, cities, or countries, depending on the visualization granularity [7, 18, 11].

In this paper, we examine the problem of assigning nodes to represent polygons on a map using a special case of U.S. senators. There are 100 U.S. senators present in the U.S. Senate, and their term in office (one or more) lasts six years. Because there are two senators per state, it is not necessarily clear where to place their nodes on a map for geospatial network analysis. Even canonical methods such as placing nodes at the centroid of their administrative unit can present well-known problems. For example, the centroid of the polygon may be sparsely inhabited/uninhabited, it may be outside the polygon itself (in the latter case, nodes are often moved manually) or nodes may be placed atop each other.

In response, we developed eight different node placement strategies and show how changing node placement alters the network's layout and how the network may be perceived. We used an unweighted, undirected network of U.S. senators linked by higher than expected bill co-sponsorship activity for the 115th session of the U.S. Senate (Jan 3 2017 - Jan 3 2019) [13]. We found that maps that use random nodes and nodes that were placed on state edges were not helpful whereas using state centroids (with or without jittering) and putting nodes at home office locations created more helpful and intuitive maps. We also found that node placement should reflect the research questions at hand, i.e.g if a research question is about agricultural bills, perhaps nodes could be place alongside farmland.

Our primary objective is to examine a design space of different node placement strategies and evaluate the pros and cons of these strategies. This work can help social network analysts, spatial information theorists, network geographers, and geovisualization experts design better software design and refine cartographic techniques for geographic social networks where geolocation is not straightforward. These visualization and exploratory spatial data systems can be used to explore networks in the computational, social, and behavioral sciences in a way where the user can choose the geolocation strategy that best facilitates their network analyses.

2 Related Work

Geospatial network visualization includes the mapping of flow networks (e.g., mobility, immigration, commuting), telecommunication networks, location-based social networks (such as Yelp and Google check-in networks), and international import and export networks. Effective node placement strategies are crucial for reducing visual clutter, such as edge crossings and node overlap, and for representing uncertainty in geospatial network visualization [8]. For nodes with specific longitude and latitude coordinates, placing them on geographic maps is straightforward (e.g. [1, 16, 2]). However, when nodes lack specific coordinates, the recognized standard practice usually is positioning them at the centroid of geographical units [7, 18, 11].

Several approaches have been explored to address these challenges. Adjustments, such as jittering, can be applied to resolve node overlap [12]. Some studies have proposed using population centroids [6] or calculating weighted centroids [20] to mimic actual locations where migration occurs. Yang et al. [19] used bounding boxes, center distance, and line distance constraints to facilitate cross-free edges and avoid ambiguity. Otten et al. [15] proposed pseudo layouts that maintain relative node positions while relaxing the Euclidean distance constraint with alternative distances or similarities between end nodes, such as the frequency of travels between places. Flow bundling, tapering, and divergent-gradiation have also been suggested to help manage overlapping flows [9, 10].

3 Dataset and Methods

3.1 Dataset Specification and Mapping Methods

Our network contains 100 senators who were part of the 115th session of the U.S. Senate (Jan 3 2017 - Jan 3 2019). We used a bill co-sponsorship dataset [5] to establish edges between senators. We adopted a stochastic degree sequence model [13] and used the backbone R package [14] for edge assignment. This approach connects two senators if they (co-)sponsored statistically significantly more bills together than would be expected at the alpha = 0.05 level. Subsequently, we removed 106 edges longer than 3,000 km to avoid visual clutter, which resulted in 454 remaining edges in total.

We analyzed data and created sets of node locations in Python using the packages *geopandas* to clean and analyze data, *shapely* to analyze planar geometric objects, and *folium* to create preliminary visualizations. We created maps using QGIS. State outlines were sourced from the U.S. Census Bureau TIGER Line shapefiles for 2018 [3], and country outlines were obtained from the World Bank [17]. Maps were produced using the Albers equal-area conic projection (ESRI:102003). U.S. city data was supplied from Esri [4].

We computed state centroids as the center of each state's bounding box (which allowed for easier computation). To jitter the centroids, we applied a small random displacement of up to ± 0.25 decimal degrees for both the latitude and longitude. To plot the mid-points on the longest diagonal of each state, we found the diagonal from the bottom-left (southwest) to the top-right (northeast) of the bounding box and computed the two midpoints such that would split the line into thirds. To prevent nodes from exiting a state's polygon, we manually moved a few nodes (for example, Florida).

For the placement strategy based on population location, we placed the two nodes at the two cities with the largest population. The first senator alphabetically corresponds to the city with the largest population. We chose the third most populous city if the second was near the first and the fourth most populous city if both, the second and the third were very close (this only occured for Idaho). For a final design strategy on office location, we collected each senator's primary office address in their respective home state and geocoded the location using the U.S. Census Geocoder[1]. The design space was developed in conversation with a set of computer science students and professors, and will be expanded in the future. The methods are designed to take advantage of centroids, maximum distances, linear bisectors, and real-world node placements (e.g., highly populated cities and senators' home office locations) (Table 1, Figure 1).

3.2 Evaluation

Four experts in the field of political network science and organizational systems evaluated the different designs through conversation. The map making team met with each individually over the video conferencing software Zoom, and presented the maps in the same order as in Table 1 using a powerpoint presentation. The experts were encouraged to talk through their thoughts and to describe the pros and cons of the designs. Through a casual conversation they were prompted to give examples such as: "I would use map 1 if...", "Map 5 seems [adjective]...", or "I would probably not use map 4 because...". Zoom meetings were recorded and the three non-evaluating authors reported the results from their notes and by replaying the Zoom meetings. The reports of this commentary were confirmed by the three non-evaluating authors.

4 Results: Responding to the Design Space of Geolocation Methods

4.1 Preferred methods and strategies

In general, the experts did not express a clear consensus on a singular design strategy, although there was some consensus in the group. Both the maps that use jittered centroids (#3) and the locations of the home offices (#8) were named by three out of the four experts as their favorite choices. The centroids, whether jittered or not, were also popular. One said: "I would prefer the jitter over the non-jittered because you see the different connections." The overlapping (i.e., non-jittered) centroids (#2) were attractive to a few experts. The experts emphasized that if the perspective is on the behavior of the state then this method is helpful. Pragmatically, an expert said that "it would be great if there was only one dot per state." Another liked the centroids because they thought it was intuitive to the user, saying there was "no risk that the location of the dot would get misinterpreted." We note that the centroid method (#2) also yields fewer visible edges and may appear cleaner and less cluttered than the other maps (Figure 1).

One expert said they would not use centroids because of the overlap, while another said that the overlap may be a problem, but this would not prevent them from using the strategy. One said of the jittered version: "I don't think it's much different from the [non-jittered] centroids." Another expert preferred that the jittered nodes slightly overlap, like a Venn Diagram, and another suggested that they do not overlap at all.

One expert had a proclivity toward the trisected midpoints of the state's longest North or South axis (#5) or the state's longest diagonal line (#6). They said that they "saw a pattern" and that these maps were "easier to look at." The expert said, "I also like the fact that many of these nodes are not touching the state borders." The property of nodes not

[1] U.S. Census Geocoder: `https://geocoding.geo.census.gov/geocoder/`.

Table 1 A set of eight different node placement strategies for U.S. senators within the United States. Highly recommended means that at least two experts liked the strategy, recommended means at least one expert liked the strategy, and not recommended means no experts liked the strategy.

	Name	Description	Evaluation
1	Two Random In-State Locations	Two random sets of coordinates are generated within a state's bounding box, manually ensuring that the point is plotted within the geographical limits of a state.	Not recommended
2	State Centroids	Nodes are placed at the geographic center (bounding box centroid) of their respective states, manually ensuring the node is within the polygon.	Highly Recommended
3	Jittered State Centroids	Random noise is added to a state centroid's coordinates to offset the nodes' locations.	Recommended
4	State Centroid and Most Distant In-state Point	One node is represented as a centroid and another is placed at the most distant in-state location from the centroid.	Not Recommended
5	Mid-points on Longest Axis	Nodes are trisector points along a state's longest central meridian or parallel (i.e., the N-S line or the E-W line), depending on which line is longer.	Recommended
6	Mid-points on the Longest Diagonal	Nodes are trisector points along a a line that connects the southwest and northeast coordinates of a state's bounding box. The south-to-north progression was chosen to reflect the shape of the U.S.	Recommended
7	Two Most Populated Within-State Cities	Nodes are placed at the largest population centers in each state. If the two population centers induce node overlap, the city with the 3rd highest population is used.	Recommended
8	Senator's In-State Office Location	Nodes are placed at the centroid of the city listed as part of the senator's home office address (within the senator's home state).	Highly Recommended

touching boundaries may help users understand which nodes belong to which states. The expert noted issues switching between directions: "In some cases, the nodes are separated horizontally, and then in some cases, they're vertical, like Illinois or Indiana; There is this mental shift I need to make." This horizontal vs. vertical distinction is a result of this node placement method where nodes were spread according to the longest N-S or E-W axis, respectively (Table 1 #5), but per the expert, may confuse some viewers.

An expert liked the strategy that used the highest population cities (#7) because it emphasized the state's demographic context. Yet, they also cautioned that it "might create some other confusion...if senator A was put in a city that actually doesn't support this senator." They suggested that only the most populous city be used for both senators. Finally, placing nodes in office locations was cited as providing the "most value added" and "the best solution." Some positive aspects included that the method gave the network a deeper meaning, that it showed the senator's political base, and it reduced confusion.

4.2 Critiques of methods and strategies

All four experts saw little benefit in randomizing node locations (#1) and stated that it would potentially confuse readers. One expert said, "the most meaningless location is the one where it is purposely randomized." Another expert said that users may be "looking for a reason, like 'why is this point here versus over there?'"

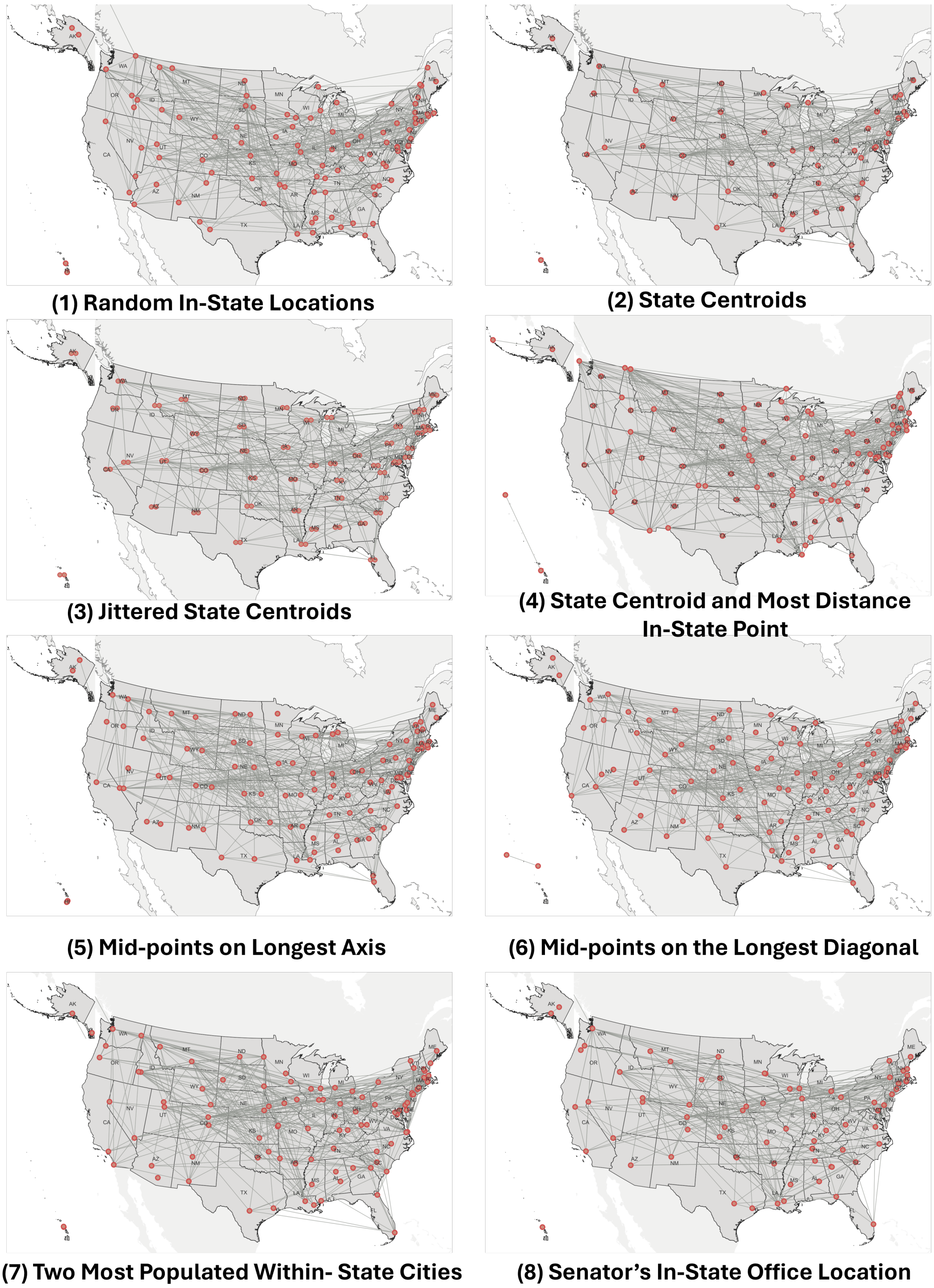

Figure 1 Maps showing the locations of senators using eight node placement methods. Some manual changes to node placement were made to prevent nodes from exiting their respective states.

For similar reasons, two experts did not like the strategy of placing nodes at the centroid and farthest point from the centroid (#4), as well as mid-points on the longest axis and on the states' diagonal axis due to the unclear rationale behind these methods. They noted that users may "accidentally think that there's something special about where the node is" when "the spatial location doesn't really tell you anything about the senators themselves." They said this method did not make sense and that the method placed emphasis on the state's borders over the senators.

Furthermore, an expert noted that the farthest points from the state centroids were located at the state borders or intersections of several states. They pointed out that "if (a node is) sitting on a state border, it would require additional attention to determine whether it belongs to state A or state B."

All four experts saw problems with the strategies that used the two most populated cities for node placement (one caveat was aforementioned in section 4.1). One said that the city "does not necessarily have anything to do with the senators or how they were voting." They mentioned, "we don't know if that city voted for them. Cities tend to be more democratic or more liberal so it would be jarring to see the senators there." Two experts added that unless the network pattern shown relates to the most populated cities or the population information is visualized in the basemap, these placements may appear arbitrary.

4.3 General takeaways

All experts emphasized that the selection of map strategies *depends on the research goals* and the specific information intended to be conveyed. The experts recommended avoiding "meaningless" location encodings or location encodings that could lead to misinterpretations. Geometrically, node overlap was not cited as a major problem, and placing nodes away from the boundaries was largely viewed as a good idea. An expert noted that partisanship strongly drives the network and thus, the geography of the network was not a particularly important feature.

In general, the experts commented more about the nodes and relatively little about how the node placement affected the edges. However, the discussion drew new ideas for edges such as "drawing connections from the coastal senators the other way around the globe", so that, for instance, a senator from Oregon would connect with New England senators without the edge traversing through the central U.S. Another suggested "folding" the map sideways so the lines are on a different plane than the map itself. One expert suggested applying spring-like constraints (e.g., a forced-directed algorithm) to force nodes apart but to constrain them within their respective state.

Node symbology was also a point of discussion. An expert suggested coloring a state's (single) node as a pie chart with red and blue halves for senators from two different parties, or scaling nodes' circle sizes by how many votes they received. Self-nodes were also mentioned, as one expert said, "I would also use curved lines, within state, so that even if they're on top of each other, there's a little loop that goes back and forth as opposed to trying to go straight between."

5 Conclusion

In this manuscript, we outlined geovisualization strategies for node placement in a spatial social network. We used a case study of U.S. senators, which presents a unique problem, as there are two senators assigned for each state and it is unclear where to place nodes to represent each senator's in-state locale. We created eight different strategies for placing nodes. Using ancillary data, such as city size and office location of the senator, was valuable because it provided more information to the reader. We recommend that, when appropriate for the research question, developers consider this strategy instead of defaulting to centroids or another purely-geometric computation.

The ultimate goal of this work is to contribute to software development design spaces for geospatial network analysis and to facilitate exploratory spatial data analysis for such datasets. Future work will include a wider exploration of the design space (i.e., adding more options and refining unpopular options), more experimentation with political party using color, and junior vs. senior senator distinction. We also hope to conduct a more in-depth evaluation via a user study. Future work will hopefully involve improved node placement options for software that allows for spatial network visualization such as ORA Lite, Gephi, and ArcGIS/QGIS.

References

1 Christian Abizaid, Oliver T Coomes, Yoshito Takasaki, and J Pablo Arroyo-Mora. Rural social networks along Amazonian Rivers: Seeds, labor and soccer among communities on the Napo River, Peru. *Geographical Review*, 108(1):92–119, 2018.

2 P Jeffrey Brantingham, George E Tita, Martin B Short, and Shannon E Reid. The ecology of gang territorial boundaries. *Criminology*, 50(3):851–885, 2012.

3 US Cenus Bureau. Cartographic boundary files-shapefile. *Updated April,. Accessed August,. p://p. census. gov/geo/tiger/GENZ/shp*, 2016.

4 Esri. USA Major Cities, 2023. URL: `https://hub.arcgis.com/datasets/esri::usa-major-cities/explore?location=21.084198%2C-7996.679932%2C4.00`.

5 James H Fowler. Legislative cosponsorship networks in the US House and Senate. *Social Networks*, 28(4):454–465, 2006.

6 Diansheng Guo. Flow mapping and multivariate visualization of large spatial interaction data. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1041–1048, 2009.

7 Su Yeon Han, Keith C Clarke, and Ming-Hsiang Tsou. Animated flow maps for visualizing human movement: Two demonstrations with air traffic and twitter data. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Analytics for Local Events and News*, pages 1–10, 2017.

8 Derek L Hansen, Dana Rotman, Elizabeth Bonsignore, Nataa Milic-Frayling, Eduarda Mendes Rodrigues, Marc Smith, and Ben Shneiderman. Do you know the way to SNA?: A process model for analyzing and visualizing social media network data. In *2012 International Conference on Social Informatics*, pages 304–313. IEEE, 2012.

9 Caglar Koylu and Diansheng Guo. Design and evaluation of line symbolizations for origin–destination flow maps. *Information Visualization*, 16(4):309–331, 2017.

10 Caglar Koylu, Geng Tian, and Mary Windsor. Flowmapper. org: a web-based framework for designing origin–destination flow maps. *Journal of Maps*, 19(1):1996479, 2023.

11 Kalev Leetaru, Shaowen Wang, Guofeng Cao, Anand Padmanabhan, and Eric Shook. Mapping the global Twitter heartbeat: The geography of Twitter. *First Monday*, 18(5-6), 2013.

12 Till Nagel, Erik Duval, and Andrew Vande Moere. Interactive exploration of geospatial network visualization. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 557–572. ACM, 2012.

13 Zachary P Neal. Backbone: An R package to extract network backbones. *PloS one*, 17(5):e0269137, 2022.

14 Zachary P Neal. Constructing legislative networks in R using incidentally and backbone. *Connections*, 42(1):1–9, 2022.

15 Heike Otten, Lennart Hildebrand, Till Nagel, Marian Dörk, and Boris Müller. Shifted maps: Revealing spatio-temporal topologies in movement data. In *2018 IEEE VIS Arts Program (VISAP)*, pages 1–10. IEEE, 2018.

16 Yuri P Springer, Michael C Samuel, and Gail Bolan. Socioeconomic gradients in sexually transmitted diseases: A geographic information system–based analysis of poverty, race/ethnicity, and gonorrhea rates in California, 2004–2006. *American Journal of Public Health*, 100(6):1060–1067, 2010.

17 The World Bank. World Bank Official Boundaries. https://datacatalog.worldbank.org/search/dataset/0038272/World-Bank-Official-Boundaries, 2023. Accessed: March 20, 2023.

18 Yang Xing-zhu and Wang Qun. Exploratory space–time analysis of inbound tourism flows to China cities. *International Journal of Tourism Research*, 16(3):303–312, 2014.

19 Yalong Yang, Tim Dwyer, Sarah Goodwin, and Kim Marriott. Many-to-many geographically-embedded flow visualisation: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):411–420, 2016.

20 Xuesong Yu, Kun Qin, Tao Jia, Yang Zhou, and Xieqing Gao. Modeling the interactive patterns of international migration network through a reverse gravity approach. *Sustainability*, 16(6):2502, 2024.