Lab 2 SNoMaN for Spatial Social Network Visual Analytics

Lab 2 follows Lab 1, which introduced you to the SNoMaN tool. Here, we use the same flight network to walk you through these functionalities of SNoMaN. Tool Link: https://snoman.herokuapp.com/

Objectives:

-Use a scatterplot to find a correlation between network and geographic dimensions.
-Learn about the 'route factor' and find 'near strangers' and 'distant ties'.
-Use a modularity algorithm to see if groups in the network are also clustered on the map

List of Sections:

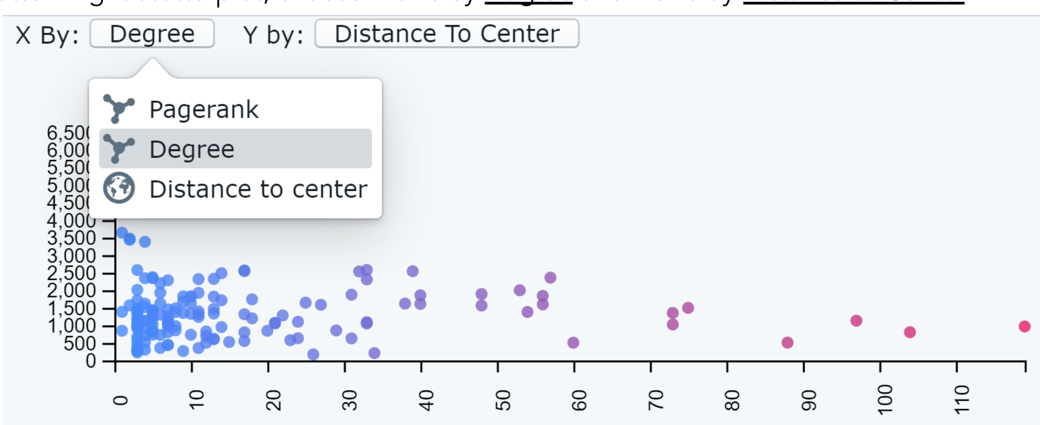
📗 Node-level Correlation Analysis: comparing node network centrality with location
📗 Path-level Correlation Analysis: comparing network and geographical distances
📗 Community-level Correlation Analysis

a. detecting network communities and mapping the results
b. visualizing community spatial expanse by convex hull
c. comparing network density with community spatial dispersion and community size

📗 Node-level Analysis: comparing node centrality with location

To shorten the overall travel distance between airports, let's find airports that *are close to the center of the network and have high degree.* The mean center of the network is the average longitude and latitude of all airports. The closer an airport is to the mean center, the more likely it can access most airports within a relatively short distance (in theory). Let's see if airports in the middle of the country are hubs.

- Clear all filters you applied before and clear your selections by clicking on blank areas.
- In the bottom right scatterplot, choose X-axis by Degree and Y-axis by Distance to Center.



Q1: What is the relationship between airport degree and its distance to center? (i.e., positive correlation, negative correlation, or no correlation?)

____
____.

- Brush the scatterplot to select and highlight airports of interest.

Q2: Are all high-degree airports close to the network center? What are the exceptions?

____

____.

📗Path-level Analysis: comparing network and geographical distances

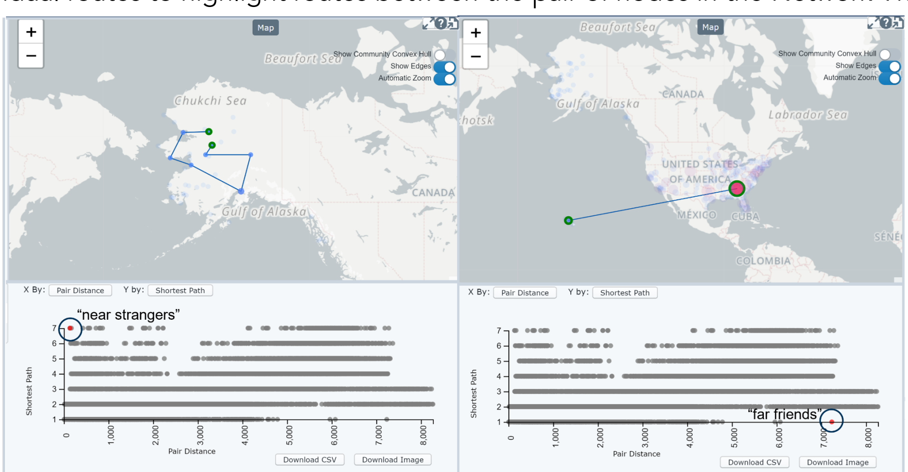
Distance has different meanings in network and geographic contexts. Network distance (aka geodesic distance or hops) measures the number of hops between two nodes. Euclidean distance measures the distance of a straight line between two nodes in the geographical space. The ratio of hops to Euclidean distance is called the **Route Factor**[7].

The number of hops between two nodes will often rise with physical separation. Comparing network and geographical distances can point out whether this is a regularity with a particular dataset and where any anomalies occur. For instance, nearby neighbors that require many hops to reach each other in the network may suggest a missing link that can be added to the network.

- Run force-directed layout on the network and pause the layout.

- Under the Statistics Panel, click on Run Shortest Path. It takes a few seconds for the result to be updated in the bottom right scatterplot.

- Hover over individual routes to highlight routes between the pair of nodes in the Network View and the Map View



Q3: Give examples of where two airports are geographically close to each other but distant in network distance (i.e., "near strangers")_________________________________;

Give examples of where two airports are geographically distant but are directly connected (i.e., "far friends")____.

Q4: What is the network diameter (i.e., the longest shortest path in the network)? ____________.
Give examples of node pairs separated by the network diameter:
____.

What suggestions could you give to decrease the network diameter effectively? Hint: it involves adding strategic edges.

____

____

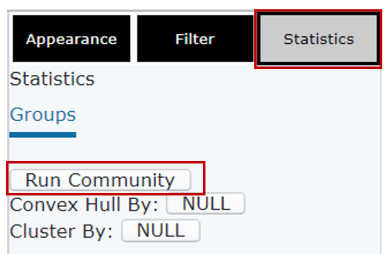
____

____.

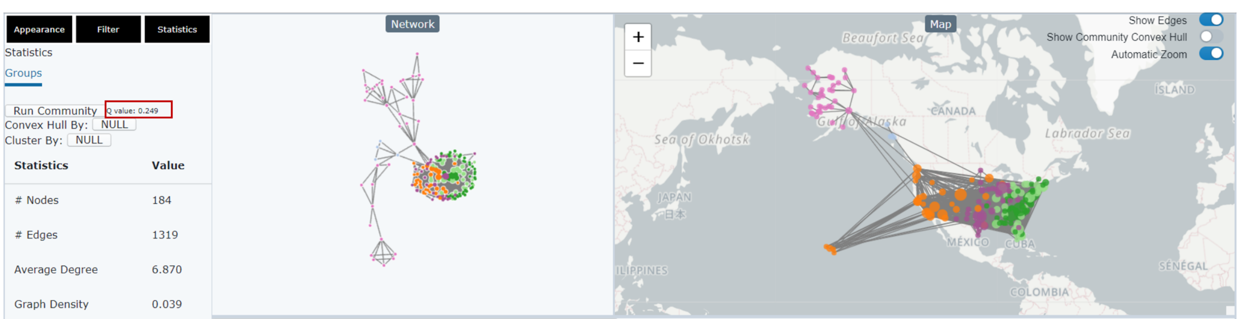
Community-level Analysis

Network community detection algorithms are used to partition networks into subgraphs wherein the subgraphs contain nodes that are more likely to connect within the subgraph than with nodes in another subgraph. It is widely used to detect clusters in a network.

- SNoMaN implements the Louvain algorithm[4] to identify network communities. Under the Statistics Panel, click on Run Community. It takes a few seconds for the algorithm to run.
.



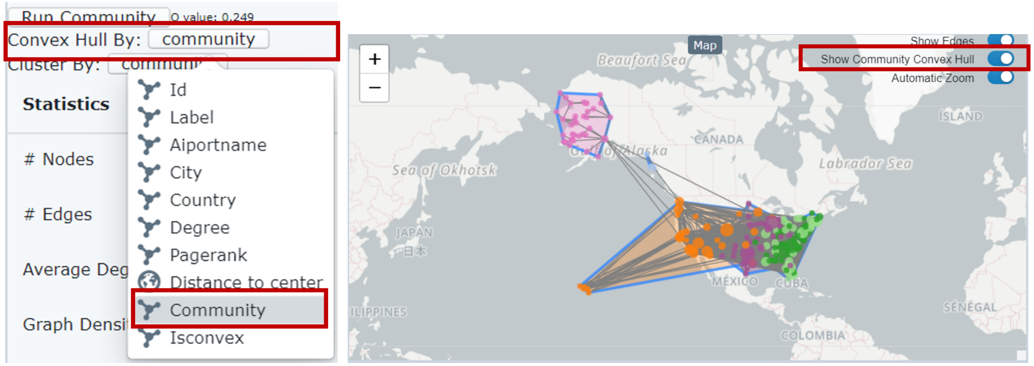
- The output of the algorithm is a Q-value that represents the strength of the partition, ranging from -1 to +1; labels for nodes with information on which community (i.e., subgraph) they are assigned to. The Q-value will be shown next to the button. The node community label will be color coded in other views.



a. visualize community spatial expanse by convex hull

- To see the spatial expanse of each community, run Convex Hull by Community under the Statistics Panel. The convex hull algorithm will outline community geographical boundaries by the minimal convex set containing all group members on the map. You can choose to see the convex hull or not by switching on and off the toggle in the Map

View. You can also choose to hide or show edges on the map with the "Show Edges" switch.



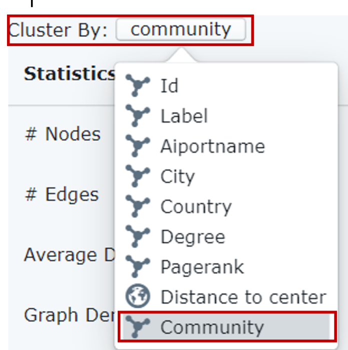
Q5: Based on the network community mapping results and convex hulls, give evidence of where network communities are divided by geographic separation.

____.

b. compare network density with community spatial dispersion and community size

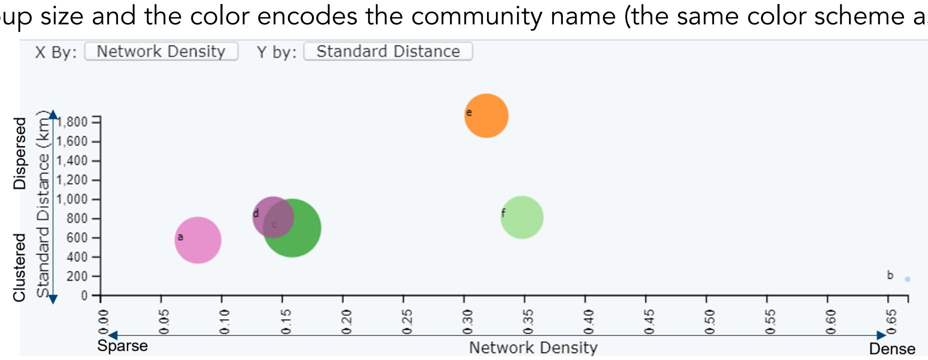
The community size and the spatial distribution of community members can impact their level of connectivity. For example, nearness can increase network density by minimizing travel costs between members. SNoMaN implements the Cluster-Cluster plot to explore the relationship between network density and spatial dispersion of communities and detect any unusual cases.

- Under the Statistics Panel, you can choose Cluster By either a-priori group labels or the community detection result to group nodes and draw the Cluster-Cluster plot.



- The Y-axis of the Cluster-Cluster plot encodes the standard distance of communities, which measures how spatially dispersed a group of nodes is. A high value of standard distance indicates the group members are dispersed in geographic space while a low value of it indicates the group members are clustered in geographic space.

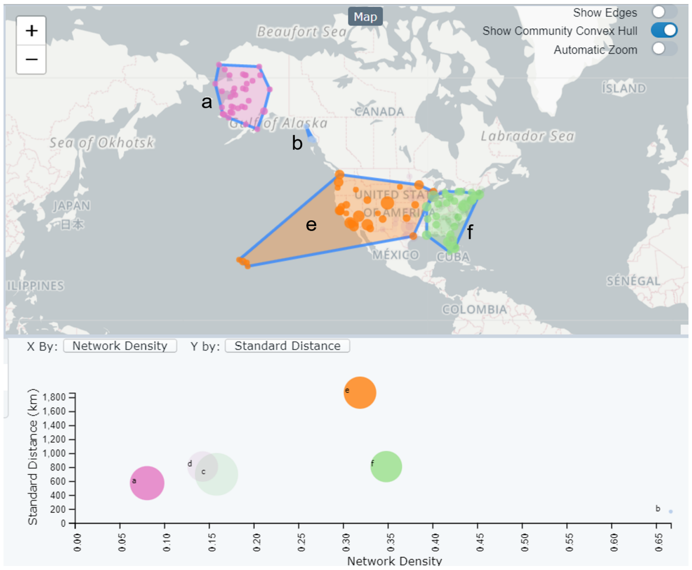
- The X-axis encodes the group network density, measuring how well the group members are connected. The dot size encodes the group size and the color encodes the community name (the same color scheme as other views).



Q6: Do spatially dispersed communities have lower network density than spatially clustered communities? ______________________________. Are small communities more likely to be spatially clustered and well-connected within the community? ____.

- Click on each node to highlight its group members and spatial expanse on the map.

Q7: Take screenshots of the potential "service area" of each community. Do these "service areas" overlap or not with each other? ____. Paste your screen captures below as evidence.



Appendix: Vocabulary Table

The following table presents the metrics and representations employed in spatial social network analysis (SSNA) and their implementation in the SNoMaN software. The metrics and representations are categorized as either Network (in orange) or Geographic (in green) techniques.

SSNA Levels	SSNA Vocabularies	Definition and Calculation Methods
Network-level	force-directed layout	A force-directed layout is widely used for network visualization. It is calculated by assigning spring-like forces among the set of edges and the set of nodes, where connected nodes are attracted to one another while disconnected nodes are repelled. This layout tends to position high-degree nodes at the central region of the drawing. SNoMaN uses the D3 force-directed layout [1] for implementation.
	node degree distribution	A node degree distribution is the probability distribution of node degrees over the whole network. The degree of a node is the number of connections it has to other nodes [2].
	edge distance distribution	An edge distance distribution is a histogram of the Euclidean distance between all pairs of connected nodes.
	network density	The network density represents the proportion of possible connections in the network that are actually present. The value ranges from 0 to 1, with the lower limit corresponding to networks with no relationships and the upper limit representing all nodes are directly connected[2]. SNoMaN currently takes imported networks as undirected when calculating network density.
	network diameter	A network diameter is the length of the longest shortest path between two nodes in the network [2].
	connected component	A component is a connected subgraph that is not part of any larger connected subgraph. All the nodes in a connected component can reach one another through the connections [2]. SNoMaN currently takes imported networks as undirected when calculating connected components.
	clustering coefficient	A clustering coefficient measures the degree to which nodes in a graph tend to cluster together. SNoMaN uses the average clustering coefficient from NetworkX for implementation [3].
Node-level	degree	The degree of a node in a network is the number of connections it has to other nodes [2].
	distance to center	The distance to center metric measures a node's Euclidean distance to the network Mean Center (i.e., a conceptual center calculated by the average longitude and latitude).

SSNA Levels	SSNA Vocabularies	Definition and Calculation Methods
Path-level	shortest path	A shortest path, or geodesic path, between two nodes in a network is a path between them with the minimum number of edges [2].
	pair distance	The term pair distance used in SNoMaN refers to the Euclidean distance between a pair of nodes.
Community-level	network community/ modularity	Network communities are subgraphs wherein nodes are densely connected internally. There are many mathematical ways to partition the network into communities. SNoMaN implements the Louvain algorithm [4]. The output Q-value measures the strength of partition, ranging from -1 to 1.
	convex hull	Convex hulls are used to visualize network communities' spatial expanse, which outlines the geographical boundaries by the minimal convex set containing all group members on the map [5]. SNoMaN uses Z-score to eliminate outliers when calculating convex hulls.
	standard distance	The standard distance[6] measures feature distribution around their center. SNoMaN uses standard distance to measure how dispersed community members are from its community mean center.

References

[1] https://github.com/d3/d3-force
[2] C. Prell. Social network analysis: History, theory, and methodology. Social Network Analysis, pp. 1–272, 2011.
[3] Hagberg, Aric, and Drew Conway. "Networkx: Network analysis with Python." URL: https://networkx. github. Io (2020).
[4] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10): P10008, 2008.
[5] F. P. Preparata and M. I. Shamos. Computational Geometry: an Introduction. Springer Science & Business Media, 2012
[6] B. K. Flury and H. Riedwyl. Standard distance in univariate and multivariate analysis. The American Statistician, 40(3):249–251, 1986
[7] D. Sarkar, C. Andris, C. A. Chapman, and R. Sengupta. Metrics for characterizing network structure and node importance in spatial social networks. International Journal of Geographical Information Science, 33(5):1017–1039, 2019.