Lab 3 Spatial Social Network Metrics in SNoMaN

Tool Link: https://snoman.herokuapp.com/

Objectives: In this lab, you will use SSN metrics and algorithms of SNoMaN tool for spatial social network (SSN) analysis. These include:

List of Sections:

📗 K-fulfillment: A metric measuring local connection/disconnection.
📗 Flattening Ratio: Assessing the overall spatial tightness of the network.
📗 Data Assortativity Function: Comparing the degree correlation between connected nodes.
📗 Average Nearest Neighbor Plots: Analyzing the spatial clustering of communities.

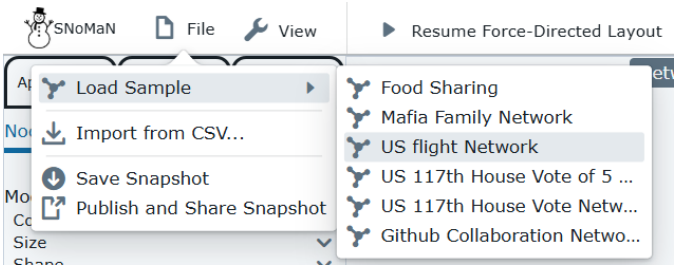
📗 K-fulfillment

K-fulfillment is a node-level metric to describe local (dis)connections. It is defined as the percentage of a node's K-nearest neighbors (in Euclidean space) that it is connected (i.e., connected K-nearest neighbors divided by total K-nearest neighbors). Here, K is equal to the node's degree. Nodes that are exclusively connected to their nearest neighbors will have a K-fulfillment value of 1. K-fulfillment assumes that the target SSN is an unweighted, undirected network.

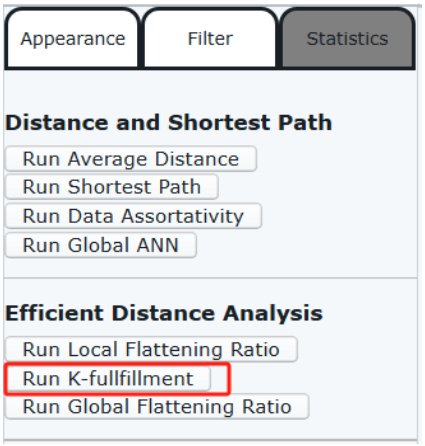
Here are some example research questions that can be answered by the K-fulfillment metric:

(Mafia members SSN) Which mafia members have strong local connections (i.e., a higher percentage of K-nearest neighbors connected)?
(Restaurant POI visits SSN) Which restaurant tends to serve residents from nearby census block groups?
(Food sharing SSN) Which organization is highly connected to other local organizations?

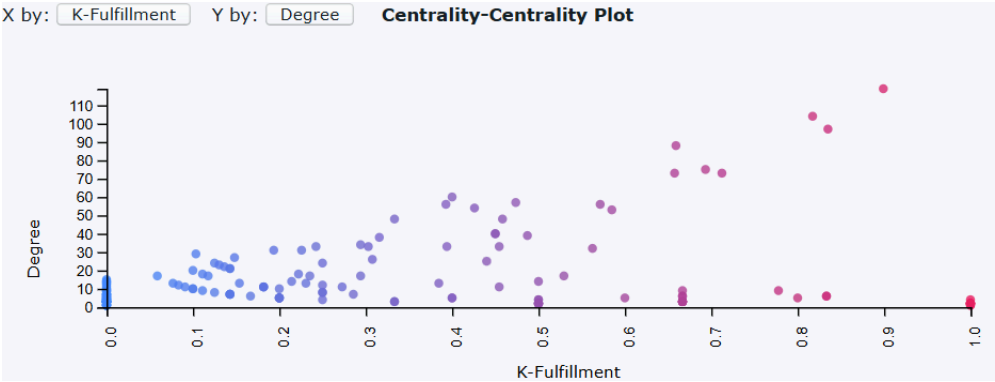
- Load the flight network in SNoMaN



- Click on Run K-fulfillment under the Statistics Panel



Q1: Which node's connections include all of its nearest neighbors? _____
Is there a correlation between nodes' degree and its K-fulfillment score?_____
_____.

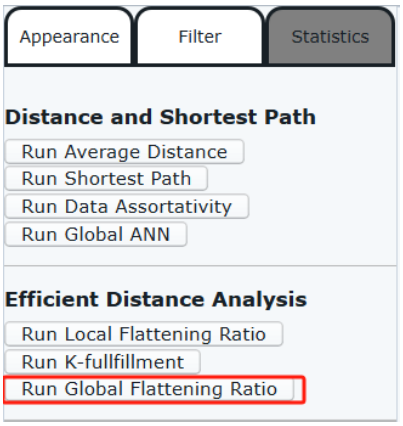


📗 Network Flattening Ratio [2]

Global Flattening Ratio is a network-level metric to measure the spatial tightness of a network [2]. To define the flattening ratio, we first create a degree-constrained nearest neighbor network G' from the given social network G by reconfiguration, such that each node i in G' with degree K connects to its nearest K neighbors in Euclidean space. As such, the (global) flattening ratio is the ratio of the sum of the Euclidean distance of edges in G' where all nodes are connected to their K-nearest neighbors versus the sum of the Euclidean distance of actual edges in G. As many G' can be possible for one G, we iterate several times to calculate the average sum distance of many G'.

Example research questions that the Global Flattening Ratio metric can answer include: Is the SSN spatially efficient? In other words, do nodes prefer to connect to their nearest neighbors or far friends?

- Navigate to the Statistic Panel and click on 'Run Global Flattening Ratio'

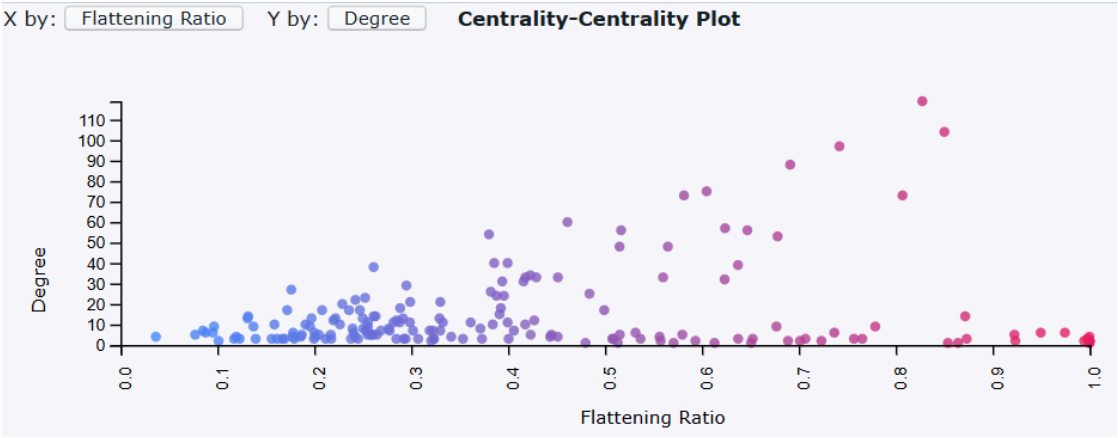


Q2: What is the global flattening ratio of the flight network? _____

Q3: Calculate the global flattening ratios of the food-sharing network and the Mafia network. Which network exhibits the most spatial tightness among the three?_____.

Local Flattening Ratio is a node-level metric, adapted from the Global Flattening Ratio definition. It is defined as the ratio of a node's minimized distance (d_opt) needed to connect to any k nearest neighbors to the total actual distance (d_act) of its connections. Nodes with low values prioritize distant connections.

- Navigate to the Statistic Panel, click on 'Run Local Flattening Ratio', and answer the following question

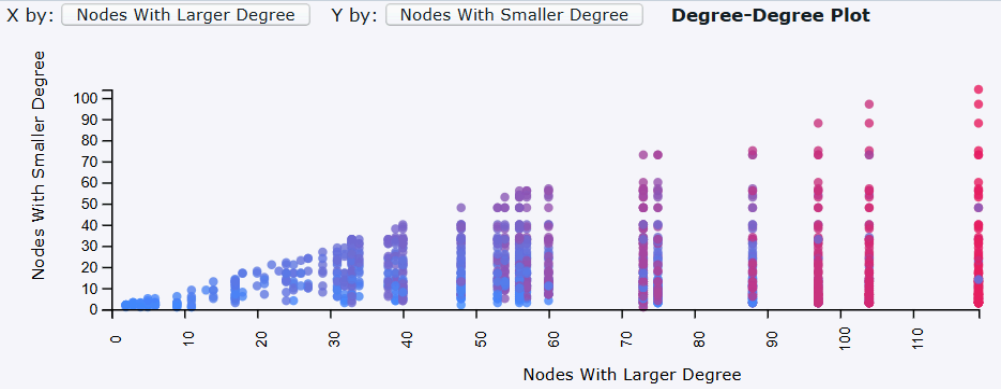


Q4: Name example SSN nodes that tend to connect to their neighbors (or have the most spatially tight social connections): _____
and nodes that tend to have most of their connections spatially distant:
_____.

📗 Data Assortativity Function

- The Data Assortativity Function measures the tendency of nodes in a network to be connected to other nodes with similar attributes. It is commonly used in network analysis to quantify the level of homophily (similarity-based connectivity) in a network. SNoMaN plots degree assortativity by comparing the degrees of the two end nodes of each edge on the x-axis and y-axis.

- Navigate to the Statistic Panel, click on 'Run Data Assortativity', and answer the following question.



Q5: Is there a tendency for nodes to connect with others of similar degree? If not, what patterns of connectivity can be observed?

_____

_____

Q6: What other spatial or social attributes do you think could relate to connections in a social network? For instance, previous studies have shown that in Facebook's friendship network, people with similar income and education levels are more likely to connect. [3]

_____

_____.

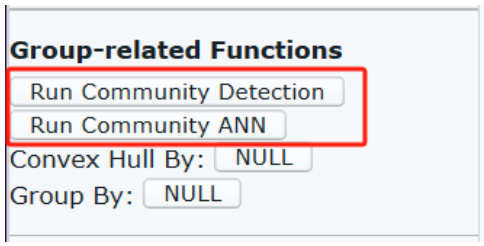
📗 Average Nearest Neighbor Plots:

An average nearest neighbor (ANN) analysis measures the average distance from each point in the study area to its nearest point. An **Average Nearest Neighbor (ANN) plots** the ANN values for different order neighbors, that is for the first closest point, then the second closest point, and so forth. It helps determine whether the distribution of nodes is clustered, random, or dispersed by calculating the average distance between each point and its nearest neighbor.

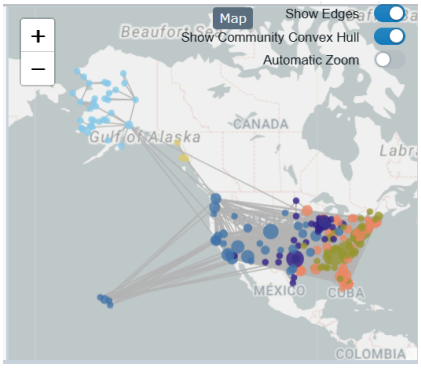
SNoMaN plots the ANN for each social group (the social group division results from the aspatial community detection SNA algorithm) with different colors. The dashed line represents the randomly generated groups by randomly selecting nodes from the network, with the size of the groups matching the largest social group. By comparing the distribution of social groups with the

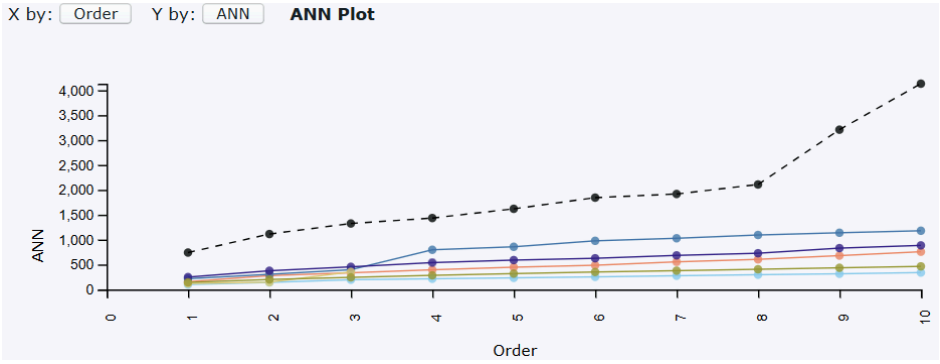
dashed line, which shows the expected clustering tendency in a spatially random network, we can determine if the social groups in the network are more clustered or dispersed than expected.

- Under Group-related Functions, first run community detection on the network and then run the community ANN function.



- Look at the results and answer the following questions:





Q9: Do different flight communities in the network tend to be spatially clustered more than expected? _____

Q10: Among the social groups, which community is the most spatially clustered and which is the most dispersed compared to the other social groups?

_____

_____

Q11: What is the problem with using the dashed line for the expectation of spatially random distribution when the network is worldwide, with nodes located on different continents?_____

_____

_____

References

[1] Kelly, Jaimie, Dipto Sarkar, and Clio Andris. "Locality, Personal Ties, and Efficiency in a Food Security Network." Annals of the American Association of Geographers (2024): 1-12.
[2] Sarkar, Dipto, et al. "Metrics for characterizing network structure and node importance in Spatial Social Networks." International Journal of Geographical Information Science 33.5 (2019): 1017-1039.
[3] Chetty, Raj, et al. "Social capital and economic mobility." Opportunity Insights. https://opportunityinsights. org/wp-content/uploads/2022/07/socialcapital_nontech. pdf (2022).