Tree Diversity in Berkeley

Final Paper Draft

Kamal Bains

Omeed Tavasoli

Rosetta Wang

Table of Contents

[**Introduction**](#_cuk0vav2i41f) **3**

[Purpose](#_jeuquj1oq6ld) 3

[Nodes](#_bb0vuxa57mrf) 3

[Links](#_odlvawauhvr6) 3

[**Related Paper**](#_7pprtz3lzk5f) **4**

[**Locality**](#_fpvjaugfg6pc) **5**

[**Methods**](#_t344y7dnh1a7) **5**

[**Analysis**](#_dcv4tqvm1ic7) **8**

[**Future Work and Conclusions**](#_f2i8i3kg22j5) **11**

[**Works Cited**](#_2vfwo4gi85sp) **12**

# 

# Purpose

Our purpose is to explore the tree diversity of the urban area of Berkeley.

# Related Paper

An article within the Journal of Urban Ecology investigated the social-ecological and demographic dynamics of urban foraging in the USA through analysis of user-generated data from the community-based monitoring application Falling Fruit (fallingfruit.org). It identifies urban foraging as a potentially critical element of food security because of its ability to support social cohesion, food justice, and ecosystem restoration. Therefore due to growing urban populations and an increasing need for sustainable food systems in cities and surrounding areas, urban foraging is a potentially critical element of food security. Their main results indicate high ethnic diversity in gathering communities and that the demographic variability of gathering communities implies a range of drivers motivating urban foraging—from sustenance to recreation.

We chose to use this same Falling Fruit network to source data on flora, particularly trees, in the Berkeley area. Through the Falling Fruit network we found a specific dataset from the City of Berkeley that when cleaned provided the genus, species, latitude, and longitude of all of the trees in our Berkeley area of interest.

# Methods

In order to assess the ecological diversity of tree genuses in Berkeley, we created a novel technique that utilizes the network properties we learned throughout the course.

First, after analyzing the data, we realized that, due to computational limits, we would need to significantly reduce the scope of our project. This was due to the fact that, as a part of our analysis, we would have to consider every combination of size 2 of our nodes in order to create our links, meaning as we include more nodes, the compute time would increase factorially. As a result, we began by selecting a 1km by 1km square of Northern Berkeley to use for our analysis. The process by which we chose our location of Berkeley was somewhat arbitrary. We ended up choosing bounding streets of Shattuck Ave, Eunice St, Euclid Ave, and Vine St. The bounding box we chose also included the Berkeley Rose Garden and Live Oak Park. Figure 2 is a visualization of our nodes geographically mapped according to latitude and longitude positions.

After finalizing the scope of our analysis, we were then tasked with creating a network out of our data. Because we wanted to create a measure of ecological diversity, we decided that our links should represent some genetic difference between the nodes (trees) and that the weight should reflect the degree to which the two nodes differ. We finally settled on the following schema in our link creation.

Figure 3: Weight Schema

|  | Below inner threshold | Between inner and outer thresholds |
| --- | --- | --- |
| Different genus and species | 4 | 3 |
| Same genus, different species | 2 | 1 |

The above schema reflects our two-part desire: to reflect a taxonomic difference between nodes and to reward an idea of closeness between nodes. Note, this schema also penalizes common planning tactics that lead to genetically homogenous city streets (i.e. planting many of the same species of tree next to each other). As for how we decided on the thresholds, we wanted thresholds large enough that plants along the same block could create connections, but not so large such that trees across the entire bounded area could create connections. After cleaning our data and creating all of our connections, we were left with a **graph of 800 nodes and 5187 links.**

After constructing our network, we now needed to evaluate the genetic relationships between our nodes. To do this, we considered 2 different properties of our nodes: the weighted degree by genus as well as the closeness centrality of our nodes. Here are our results for the weighted degree by genus analysis. Our resulting network has small world properties since our calculations revealed a very low average shortest path and high clustering coefficient.

## Nodes

Each of our nodes represents a single tree at its respective location.

## Links

For our links we considered what the weight of each link represents. Since we are interested in measuring diversity, the weight will be higher with geographically closer nodes and with nodes of trees that are more taxonomically different. In other words, the weight is negatively correlated with physical distance and positively correlated with taxonomic difference. Some interesting metrics we will be able to obtain from our work will be what the average diversity is across the entire network as well as understand how important individual genuses are to the ecological diversity of the urban area. We did not want to have every node linked with every other node, so we filtered based on the geographic distance and taxonomic distance i.e. if it is too far away and not different enough of a plant, it is not linked. The runtime for procuring the edges was roughly the same, even when different amounts of links were created , but only needed to be run once and then saved as a csv for future use.

## Locality

We isolated data that covered 1 square kilometer that corresponds with the area shown in the map in figure 1. The runtime for creating all the links was a limiting factor on the area of the locality. This 1 kilometer plot gives a runtime of about 12 minutes given the number of

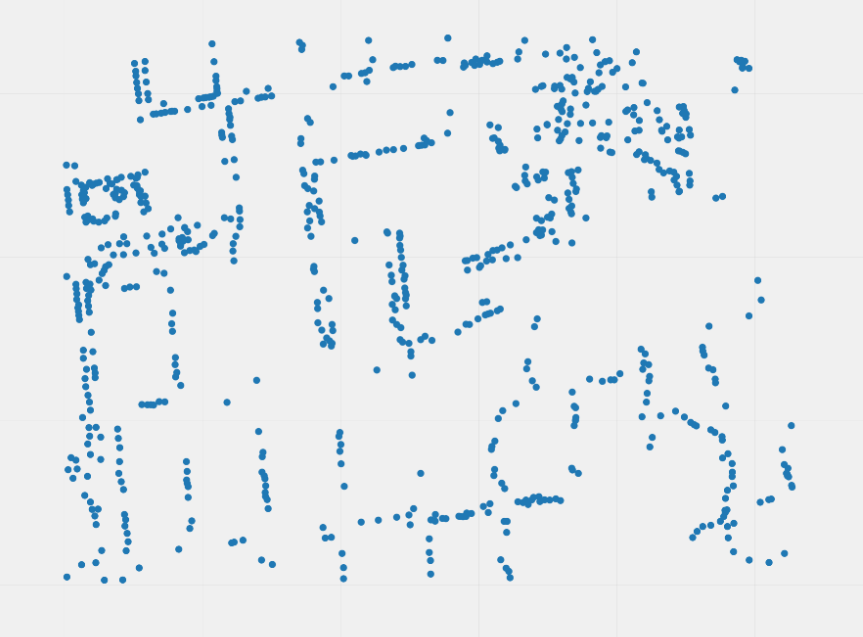


Figure 1 & 2 , Location (Google)

The nodes in figure 1 are mostly located on streets or parks because the dataset maps trees in public spaces.

## Thresholds

We chose the thresholds 50 and 25 (meters). The outer rim will have bias because there will be less nodes to connect to. Too high of a threshold would put a large bias on a high proportion of the area we are inspecting. Too low of a threshold would leave many nodes unconnected. Also, consider the width of a street block since the nodes would mostly occur on streets.

# 

# 

# Analysis

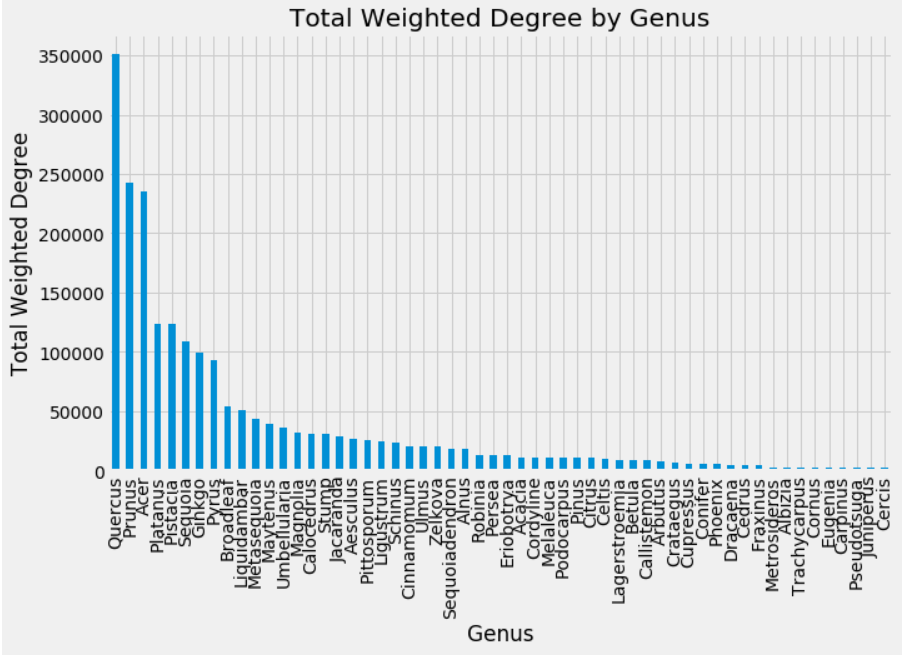


Figure 4: Total Weighted Degree by Genus

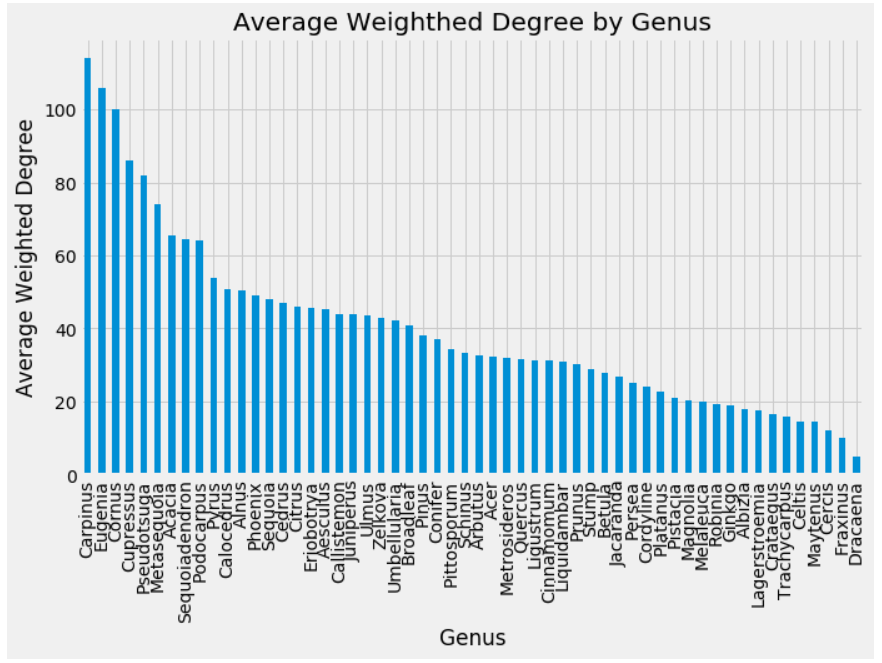
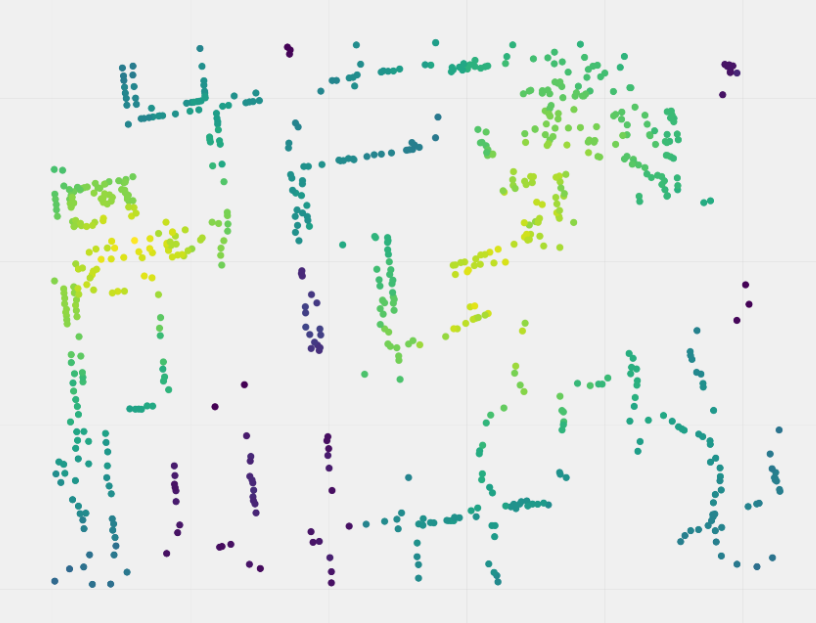


Figure 5: Average Weighted Degree by Genus

To get figure 4, Total Weighted Degree by Genus, we simply summed the weighted degree of nodes by genus whereas for figure 5, Average Weighted Degree by Genus, we took the average. From this analysis, one can clearly see a trend in the urban ecology of Berkeley where a few types (genuses) of trees dominate the landscape, but because of their homogeneity with regards to species, they are not impactful to the overall genetic diversity. For example, the genus Quercus (oak) has, by far, the highest total sum of the weighted degree of its nodes, but because each individual oak tree does not greatly increase diversity, the average weighted degree for Quercus is quite low and is not even in the top 50% of genuses. Similarly, the genus Carpinus has one of the lowest total weighted degree measurements, but has the highest average, indicating that each node that is a part of the Carpinus genus has a very high level of impact on the overall genetic diversity of Berkeley. This is due to the fact that the weighted degree of any node tells us two things: the number of connections between that node and other nodes and how significant (i.e. genetically diverse) that connection is.

Figure 6: Closeness Centrality

We also calculated the closeness centrality of each of our nodes. Our logic for the use of this metric was that if a node has very small shortest paths between other nodes (large closeness centrality), then the node is most likely genetically distinct in comparison to its neighbors (meaning connections can be made easier). Thus, closeness centrality gives us a measure of how important an individual node is to the overall diversity of our network, a distinct metric from our weighted degree analysis which was in the context of genuses.

From figure 6, it is clear that there are certain hotspots of nodes that have much higher closeness measures than others. This is most likely explained by the fact that many major streets in cities are lined with the same few species of trees. Moreover, the hotspots we see align geographically with both Live Oak Park and the Berkeley Rose Garden, which suggests that because of urban planning choices in these areas, we see a high degree of genetic diversity.

Overall, our efforts of transforming this plant data into a genetic network reveal that Berkeley, like many other urban areas, has a largely homogenous plant ecology, where few genuses of trees dominate and where genetic diversity is limited to carefully planned areas such as parks or gardens.

# Future Work and Conclusions

We determined that Berkeley’s tree community is not very diverse considering:

1. The difference between the total and average weighted degree was significant
2. The closeness centrality map is non-uniform with distinct hotspots centered around the parks and gardens

Therefore the biological connections built do not seem of high quality and most diverse plants are the result of planned recreational areas whose visual aesthetic driven diversity is not a representation of the city wide general oak and maple tree homogeneity.

The related paper related fruit tree diversity to social justice and social disparities. For future analysis, a potential avenue is to use census data in conjunction to plant data to reveal more interesting trends.

# 

# Works Cited

Google Maps. (n.d.). San Francisco, California. Map data ©2020 Google. Retrieved April 29, 2020, from https://www.google.com/maps/@37.8801863,-122.2676369,15.78z. Screenshot by author.

Dataset: https://data.cityofberkeley.info/Natural-Resources/City-Trees/9t35-jmin