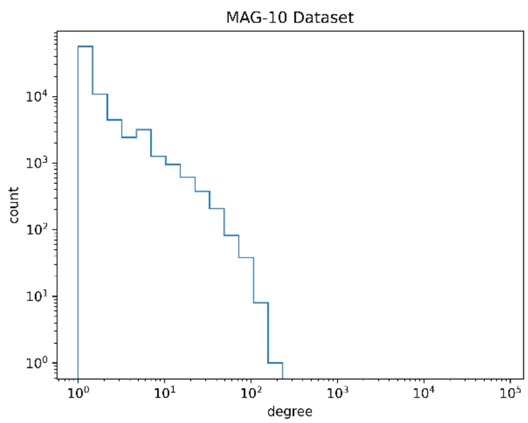
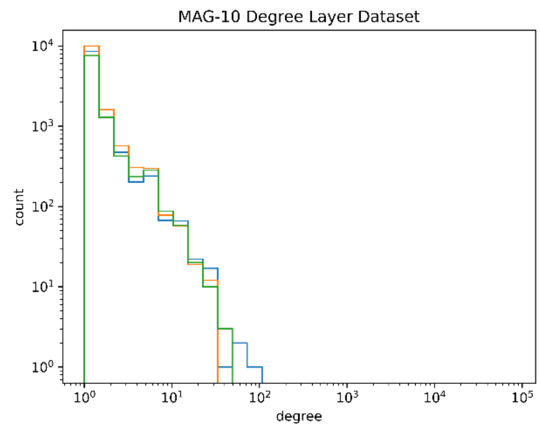
MAG-10 Dataset



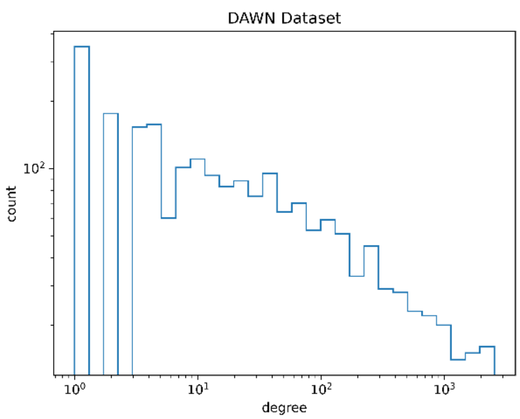


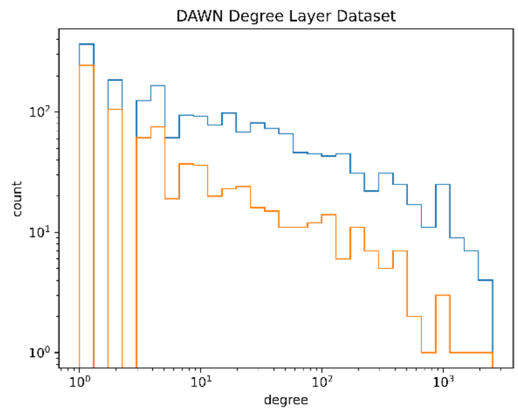
The MAG-10 Dataset is a collection of computer science authors who have published together in certain conferences. The x-axis represents degrees where each author has a certain number of contributors. The y-axis represents the count of each author. The graphs are scaled logarithmically to exhibit power law distribution as most nodes have few connections and a small number of nodes have hundreds or thousands. The power law distribution makes it easier to view the relationship and characteristics of the networks versus scaling them linearly.

The graph on the left depicts the degree distribution of authors in the whole dataset. The nodes of the hypergraph are specific authors while the hyperedges are publications from those corresponding authors. The graph has a steep drop on the left indicating that most authors have a low amount of contributions, most authors have 10 or fewer contributions. There are more than 1000 authors that have between 10-100 contributions. Beyond that there are very few in the context of the dataset that have more than 100 contributions.

The graph on the right depicts the degree distributions for different conference categories which are represented as layers. Similar, to the graph on the left, there is a steep drop on the left. The degree distributions of the different conferences depicted by the different colored lines are relatively similar. This indicates that the conferences have similar contribution patterns; a few conferences have authors with high densities of connections compared to other conferences. This trend could indicate that the same author is attending multiple events or authors in similar fields are populating these events.

DAWN Dataset



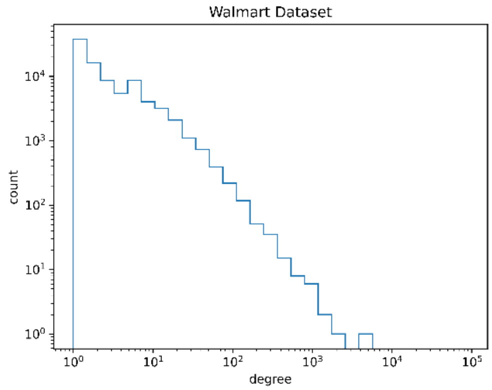


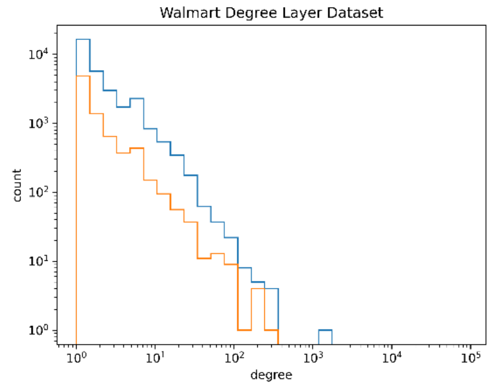
The DAWN Dataset is a collection of drugs people have ingested prior to emergency room admissions. The x-axis represents the degrees of each drug combination. The y-axis represents the count of how many drugs share that degree. The graphs are scaled logarithmically to exhibit power law distribution as most nodes have few connections and a small number of nodes have hundreds or thousands. The power law distribution makes it easier to view the relationship and characteristics of the networks versus scaling them linearly.

The left graph shows the degree distribution of the drug combinations in the entire dataset. The nodes of the dataset are unique drugs while the hyperedges are combinations of drugs taken by patients. There are a lot of drugs that have been detected few times (right) while there are also a couple drugs that have been detected hundreds of times. However, there appears to be a fluctuation in some drugs that have been detected as they appear often between 10-100 times suggesting that they could be used in different instances with other drugs.

The right graphs displays the degree distribution of edge categories, represented by patient response. These categories are: “sent home”, “surgery”, or “released to detox” and are displayed by the different colored lines on the graph. The individual outcomes from each category exhibit similar behaviors, but appear in different frequencies. Those with low frequencies could indicate more dramatic outcomes such as surgery where potentially lethal drugs were ingested by patients.

Walmart Dataset



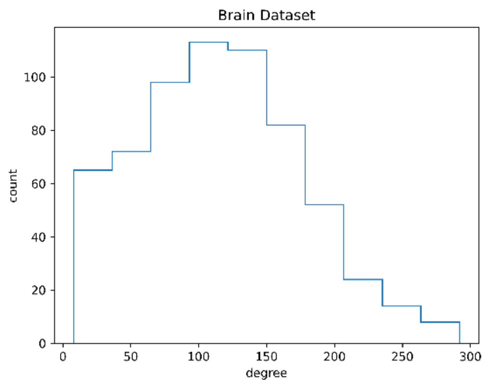


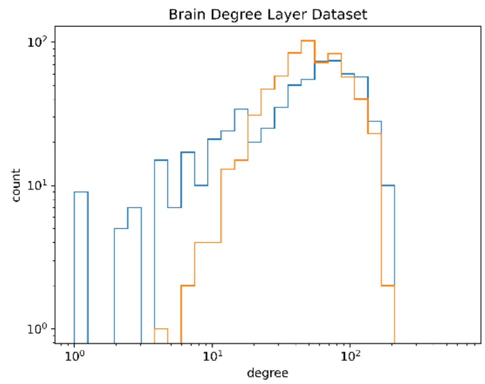
The Walmart Dataset is a collection of items in people's shopping carts. The x-axis represents the degrees of each item combination. The y-axis represents the count of how many items are shared that degree. The graphs are scaled logarithmically to exhibit power law distribution as most nodes have few connections and a small number of nodes have hundreds or thousands. The power law distribution makes it easier to view the relationship and characteristics of the networks versus scaling them linearly.

The left graph represents the degree distribution of Walmart shopping carts by items where nodes are individual items and hyperedges are groups of items that were purchased together. This graph shows power law distribution as only a few products have a high frequency being bought together, specifically around 1000-10000 times sometimes even exceeding 10000 times. On the other end of the graph, the right side, items were bought in combination 10 times or less. These items could represent products that are less popular while the high frequency items could represent popular items such as milk, bread, or eggs.

The right graph is a degree distribution of Walmart shopping trips. The different colored lines are different trip types, items are bought in combination through a certain type of trip which could be delivery, in-person shopping, pick-up, etc. While the overall trend is consistent between some trip types the frequencies are varied. Some trip types have a wider range of degrees or higher frequency in degree counts. This could show that only certain items are bought in combination through certain trip methods.

Brain Dataset



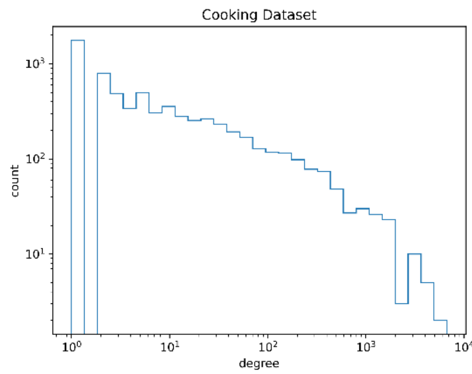


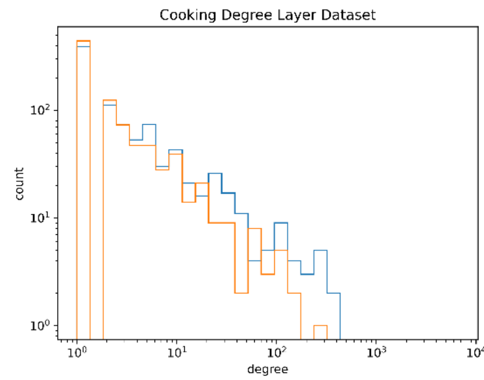
The Brain Dataset is a collection of items in people’s shopping carts. The x-axis represents the degrees of each brain region from an MRI scan. The y-axis represents the count of how many items are shared that degree.

The left graph represents the degree distribution of brain regions. This dataset is different from others that have been analyzed. There are almost no regions of the brain that aren’t active. However, certain areas are more active than others indicating that these areas are necessary for daily life. For example, they could be regions where background processes like breathing could be taking place. Some other regions that aren’t as active could be where some hormones are being processed, ones that aren’t necessary for everyday activity like adrenaline.

The right graph represents the degree distribution of brain regions by connection categories. The two edge categories are one for connecting regions with high fMRI correlation and the other for connection regions with similar activation patterns. There is vast variation between the two categories. However a similarity between the two is the area of concentration that appears between degree 10-100. These two categories could show how more unique brain activity could attribute to specialized activity in the brain. The blue line is more consistent with less variation possibly showing broader ranges of activity within the brain that occur often during the day.

Cooking Dataset





The Cooking Dataset is a collection of ingredients used together in combinations in recipes. The x-axis represents the degrees of each ingredient used. The y-axis represents the frequency of ingredient use in relation to the ingredient combination or recipe. The graphs are scaled logarithmically to exhibit power law distribution as most nodes have few connections and a small number of nodes have hundreds or thousands. The power law distribution makes it easier to view the relationship and characteristics of the networks versus scaling them linearly.

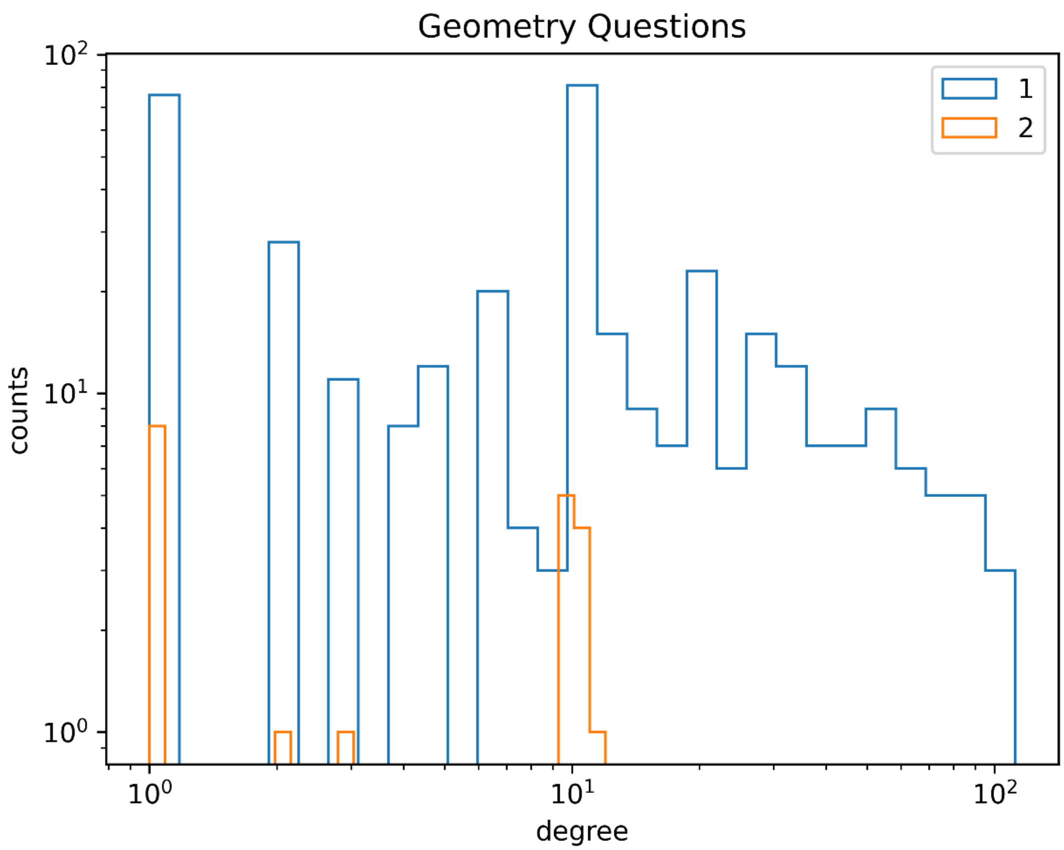
The left graph represents the degree distribution of certain ingredients used together in recipes. The nodes are represented by induvial ingredients while hyperedges represent are the combinations of ingredients in recipes. There a certain ingredient used in high frequencies across many recipes, essentially cooking cornerstones that are necessities across dishes. There are also many ingredients that aren’t used a much maybe 10 times or less, perhaps for certain cuisines or dishes.

The right graph represents the degree distribution of certain ingredients used together in recipes across different cuisines. Here represented are only two types of cuisines, the distribution is very similar, with little variation between the two. This could indicate that the cuisines could be rooted in the same regions of the world, using similar ingredients. Culture impact cuisine and food greatly, it’s possible both cuisines represented could have similar cultures where certain dishes appear in both cultures.

Small Datasets

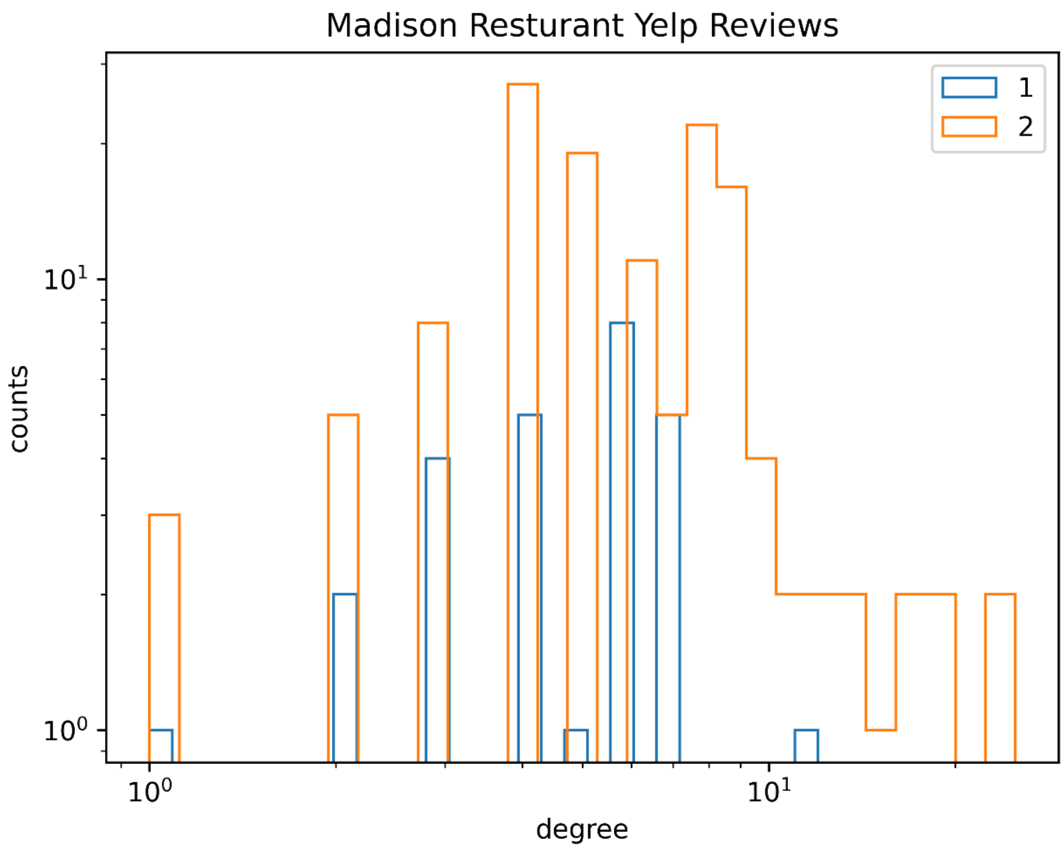
Geometry Questions Dataset

Layers 1 and 2 represent a group of students that answered a particular question within a period of a month. The degree represents the group of students answering questions. Meanwhile, the count represents how many times the group of students answered the specific question.



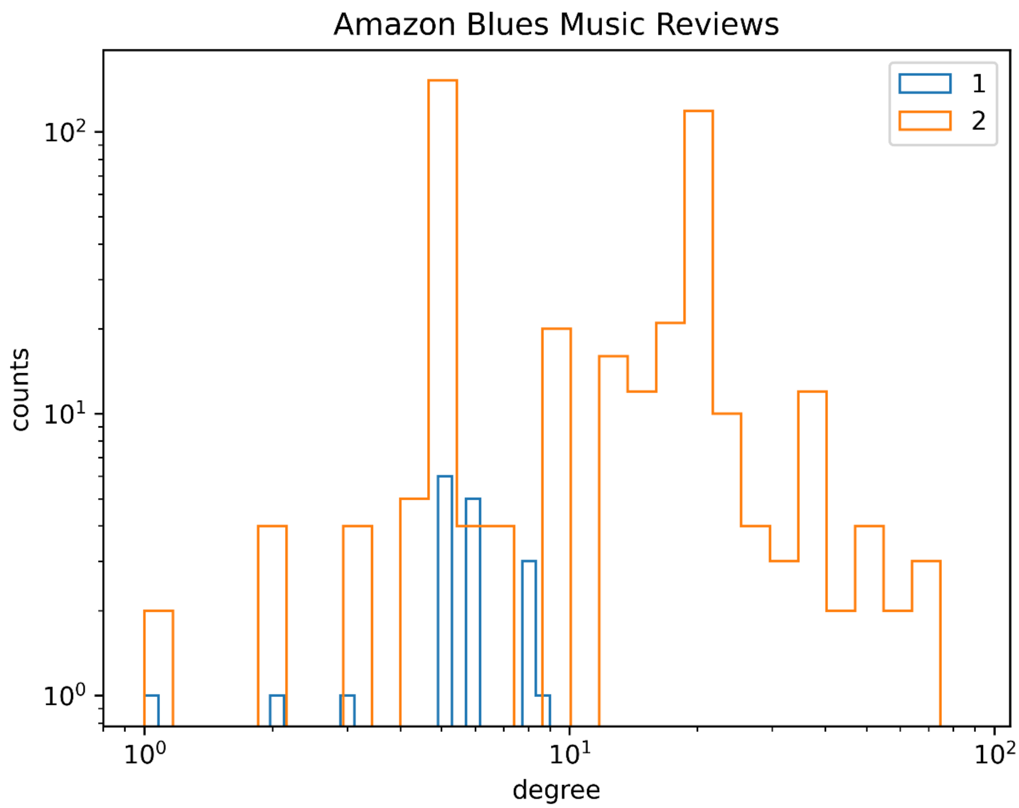
Madison Restaurant Reviews Dataset

This dataset covers yelp reviews of all different restaurants in Madison, WI. The degrees represent a certain establishment and their reviews within a month. The counts represent the number of times the restaurant has been reviewed.



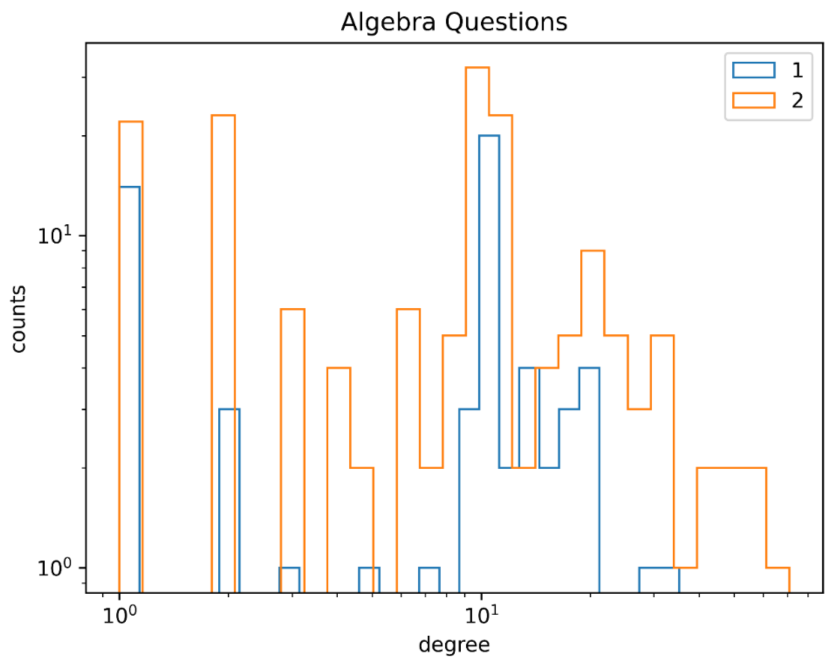
<u>Amazon Blues Music Dataset</u>

This graph shows Amazon reviews of blues music and their reviews within a month. The degrees are the lists of reviews of a certain blues product. The counts are how often the product was reviewed.



Algebra Questions Dataset

This data set analyzed groups of students who answered a particular algebra question within a month. The degree represents the groups of students who answered one question, and the counts is how many times they did.



Analysis

None of my graphs adhere to power law, there was almost no correlation between the layers in the graphs. In some graphs like the geometry data set, there seemed to be not enough data to form anything of significance. The only graph there was some minor correlation between the layers was the algebra questions dataset.

Errors

The Vegas Bars output an error of “max() iterable argument is empty”, I looked through the txt file, and I did not see anything that had a 2, but I am not sure if I was analyzing the file correctly. Though, I made a graph of the first and third layer. There still was not any correlation, or any power law distribution. The graph on the right is the first layer by itself.