1.1 Higher weights correspond to stronger relationships between characters in the book.

1.2 The dataset is set up in 5 different scv files that I found using the links below. They are each a csv file for one of the five books for Game of Thrones. For the chart that was seen in the notes it was using book3 from the dataset link below, a book called A Storm of Swords. They parsed the data from the eBook into that csv file and made 107 vertices, one verity representing each character in the book, and 353 edges, which represented a relationship between two characters in the book. These edges were weighted which helped to represent if a relationship was much more important or stronger then others.

https://www.kaggle.com/mmmarchetti/game-of-thrones-network-analysis/data

https://www.kaggle.com/mmmarchetti/game-of-thrones-dataset

2.1 The graph is cyclic because there can be a loop made in the graph. This means we can go from a node like 1 to 2 to 4 back to 1 and it does not break any of the rules of the graph.

2.2 Path 2->5->3->8

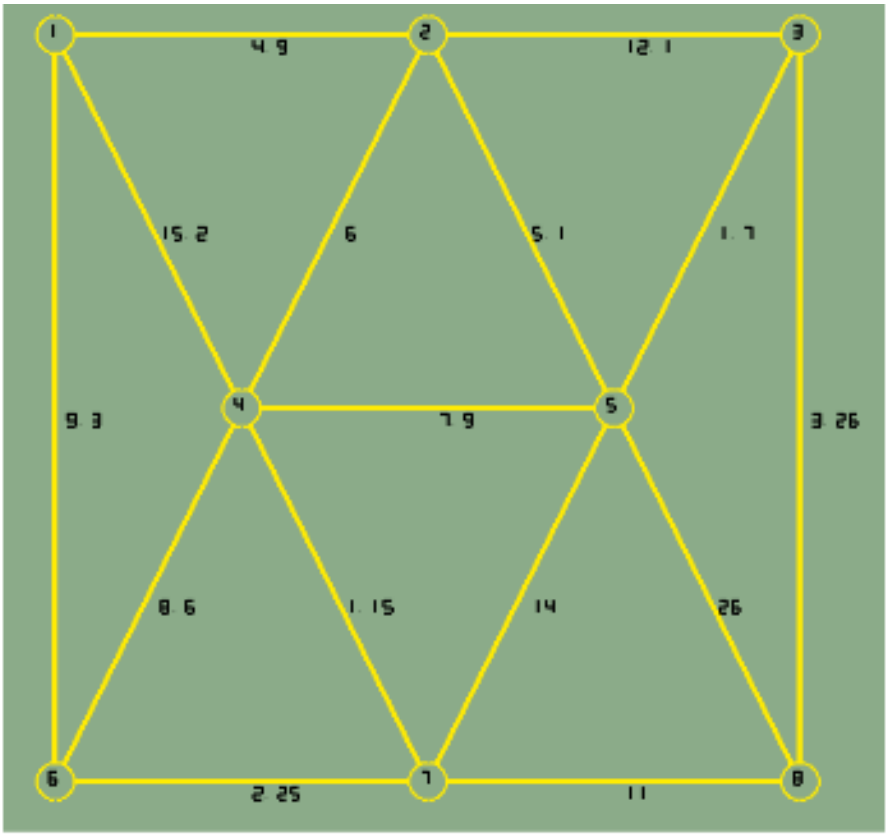
Weight 5.1+1.7+3.26 = 10.06

2.3 Path 5->3->8

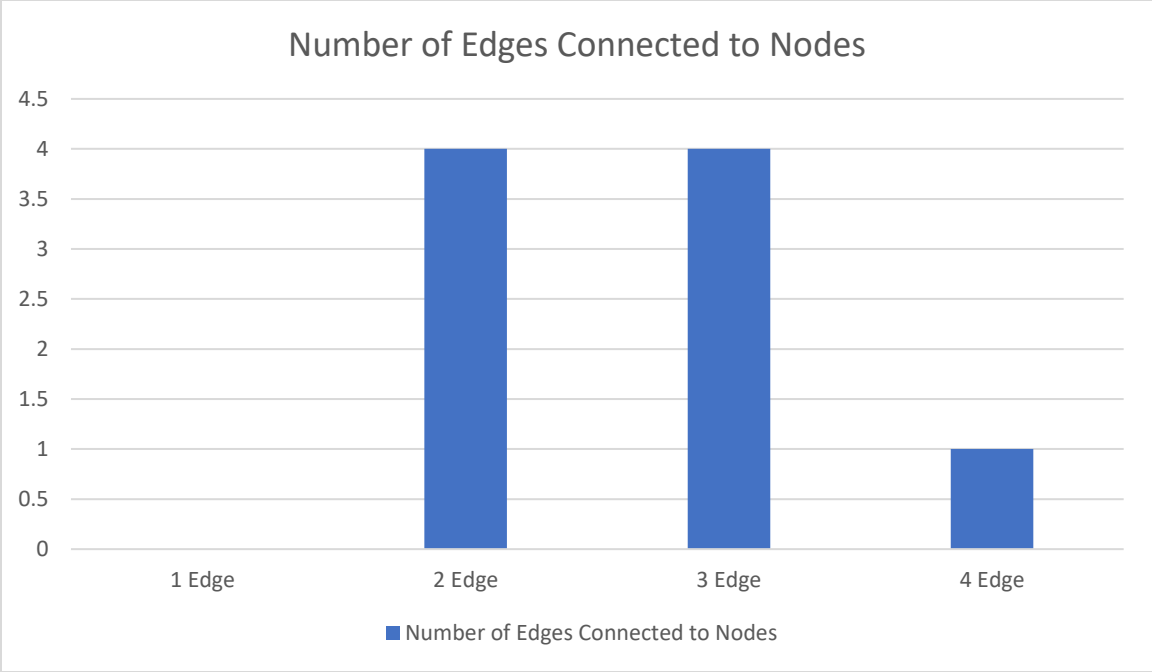
Weight 1.7+3.26 = 4.96

2.4 Path 5->4->7->8

Weight 7.9+1.15+11=20.05

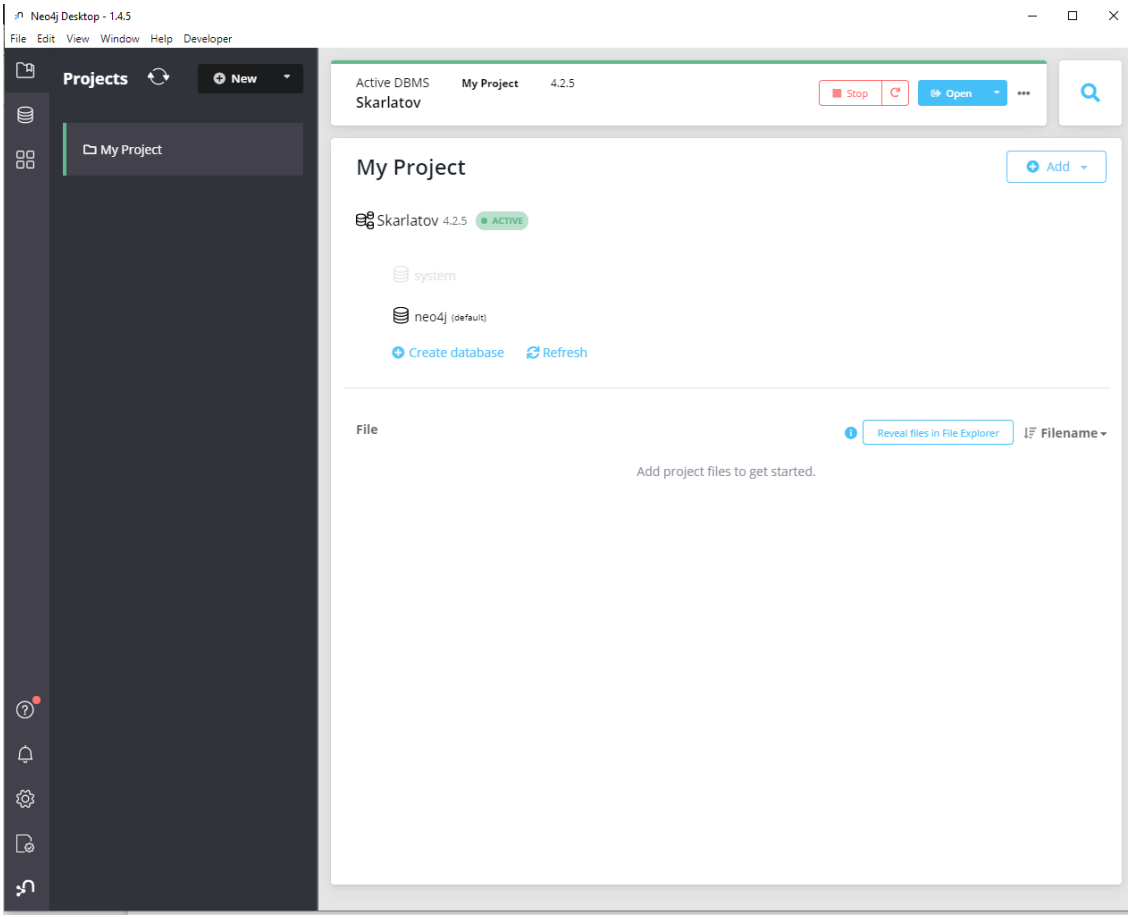


3.1



3.2 The power law distribution is related to common rule called the 80/20 rule in statistics. The 80/20 rule is important because it is thought of as being as one of the most common outcomes for most statistics. The 80/20 rule states that approximately 80% of the consequences come from 20% of the causes which means that for a majority of the information gathered would have been caused by 20% of the data. An example of this would be something like in senses, if 80% of the country's owned land was owned by 20% of the population this would be an expected outcome because of the 80/20 rule's common occurrence in statistics. This rule is important to the power law because the power law relates to a relationship between two things that functionally correlate which means that if one part of the relationship is small the other part of the relationship will be functionally bigger then the other. When graphed out this creates something called a long tail where a majority of the data graphed is close to 0 point of the x axis. This again is only important because it is somewhat of an expected common outcome for most relations between two things.

4



5.1 93740 rows

59 columns

```
[28]: import pandas as pd
      df=pd.read_csv('input/owid-covid-data.csv')
      #df.head()
      #df.tail()
      len(df)
      #df.dtypes
```

```
[28]: 93740
```

```
[27]: len(df. columns)
```

```
[27]: 59
```

5.2 iso_code column: is the abbreviation of the country the location is in

continent column: is the continent that the location of the record is in

location column: is the actual location of the record

5.3 This dataset could be used to track the spread of cases of COVID-19 in specific countries. Since it has both the number of total cases for multiple dates in every country and a column for a number of new cases from the previous date cases were recorded then you could track how quickly COVID spread in a country by plotting the dates and number of cases on a chart. This could be useful to also determine when the peak of the infection rate was for a country.

The dataset also has a column for the number of deaths from covid and column for the number of people put into the hospital from covid. From these two columns we could see what the mortality rate is for people who end up in the hospital with COVID. You can also use the age groups of the people affected by COVID from the aged_# column to determine if a certain age group was more affected by COVID then other age groups.