Classification and Features of Graphs

Emma Slack, Advisor: Emily J. King

Colorado State University

Introduction

In 2011, there were 721 million people Facebook. The median number of friends users had was 99 friends.

Yet despite the massive number of people on the platform, the average distance between two different users on Facebook was 4.7 people [1], meaning that you could connect any two people on Facebook through the friends between them in usually 5 friends or less.

We can observe this phenomenon and study it further using mathematical objects known as *graphs*.

What are Graphs?

Graphs are the visual representation of relationships via vertices and edges. In the case of Facebook users, each user is a vertex. An edge connects two vertices (aka users) if they are connected as friends on Facebook.

Graphs exist in a multitude of settings in addition to social media platforms. Some examples include:

- The relationships of actors on the Internet Movie Database (IMDb) based on shared films/TV shows.
- The connections between different areas of the brain.
- The interactions between computers on the Internet.

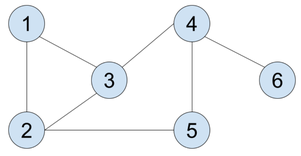


Figure: Example of a graph with 6 vertices and 7 edges.

Two key measurements for graphs are *average path length* and *clustering coefficient*.

- **Average Path Length:** The average of the shortest distances between any two vertices on a graph.

  For Facebook, this is the average number of friends between two users.

- **Clustering Coefficient:** The likelihood that any two connected vertices are connected to a third vertex.

  For Facebook, this is the probability that two friends have a mutual friend.

Low average path length is usually preferred because it means data can travel faster, while high clustering coefficient is preferred because it ensures information can still travel when an edge is removed. For Facebook, this means only a few friends are needed to connect two users and friends most likely have mutual friends.

Mathematicians Duncan J. Watts Steven H. Strogatz defined a new type of graph in 1998 which has both of these preferences.[4]

This type of graph is called a *small-world graph*.

Small-World Graphs

The defining properties of small-world graphs are

- Short Average Path Length
- High Clustering Coefficient

Small-World Graphs *cont.*

Small-world graphs are common in real-world data sets. For instance, the Facebook and IMDb examples both generate small-world graphs.

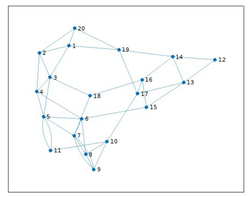


Figure: Example of Small World Graph on 20 Vertices with Average Degree 4.

Yet the definition for "short" and "high" in regards to path length and clustering coefficient are vague at best. We must understand other types of graphs which we can compare to small world graphs, namely *random graphs* and *circulant graphs*.

Random Graphs

Random graphs are graphs in which edges are randomly assigned to vertices.

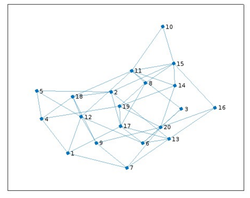


Figure: Example of a Random Graph on Twenty Vertices and of Average Degree Four

A couple interesting properties of random graphs are:

- Shorter Average Path Length
- Smaller Clustering Coefficient

Circulant Graphs

Circulant graphs are a type of lattice which creates a highly structured ring and whose adjacency matrix is constructed by shifting the first row to the right by one column for each new row.

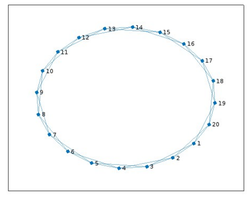


Figure: Lattice (Specifically Circulant) Graph of Degree Four on Twenty Vertices

Circulant graphs have the following characteristics:

- Higher Average Path Length
- Larger Clustering Coefficient

A Range of Graphs

We notice that as a circulant graph becomes more random, its clustering coefficient and average path length decrease. We can see this graphically:

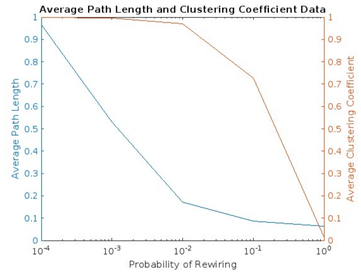


Figure: Average Path Length & Clustering Coefficient vs. Probability of Rewiring.

We notice that in this process, there exists a graph which has a relatively high clustering coefficient and low average path length. This means that small-world graphs reside between circulant and random graphs, and we can create small-world graphs by randomly rewiring the edges of a circulant graph with some probability.

In our tests shown above, we find that small-world graphs exist when the probability of rewiring is roughly 0.01.

Classification of Small-World Graphs

To identify small-world graphs, Humphries and their associates created a new measure for graphs [3]:

$$\sigma = \frac{C/C_{Rand}}{L/L_{Rand}}$$

where C is clustering coefficient, L is average path length, and C_{Rand} and L_{Rand} are the clustering coefficient and average path length for a random graph of the same size. A graph is considered small-world if

$$\sigma > 1.$$

However, Telesford and their colleagues [3] created a new measure, ω, which tells us if a small-world graph is more circulant or random:

$$\omega = \frac{L_{Rand}}{L} - \frac{C}{C_{Latt}}$$

where C_{Latt} is the clustering coefficient of a circulant graph of the same size. A graph is considered small-world if

$$\omega \approx 0,$$

and is more circulant if $\omega < 0$ and more randomized if $\omega > 0$.

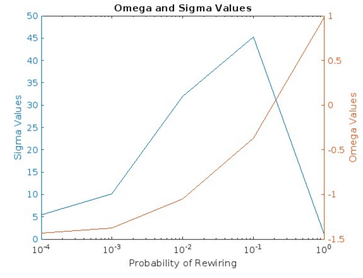


Figure: Sigma Values and Omega Values as Probability of Rewiring Increases.

Classification of Small-World Graphs *cont.*

From our tests, it appears sigma and omega classify small-world graphs when the probability of rewiring is about 0.1. This suggests there may be a range of rewiring probabilities in which small-world graphs exist.

Identifying Isomorphisms

In addition to studying small-world graphs, I worked on understanding graphs on a small scale (i.e. less than 5 vertices). I found most graphs with the same number of edges and vertices have the same structure.

This is what is known as a *graph isomorphism*.

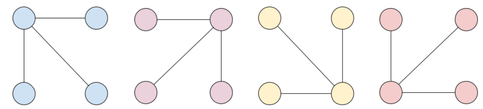


Figure: These four graphs are all isomorphic to each other.

To identify isomorphisms, I used the following metrics:

- **Average Path Length**
- **Clustering Coefficients**
- **Eigenvalues of its Adjacency Matrix:** The adjacency matrix of a graph is the matrix whose columns and rows are numbered according to the numbering of the vertices on the graph (i.e. the first row and column correspond to the vertex labeled 1, the second row and column for the vertex 2, and so on).

  If there is an edge between the i-th and j-th vertex, then a 1 is put in the (i, j) entry of the matrix. Otherwise, the entry is a 0.

  An eigenvalue of the adjacency matrix M is a scalar α which satisfies the equation $M\vec{x} = \alpha\vec{x}$ for some eigenvector $\vec{x}$.

- **Maximum and Average Degrees:** The degree of a vertex represents the number of edges it has.
- **Number of Components:** A component is an isolated part of a graph which has no edges connecting it outside the component.

I found that there are only 18 graphs structures (accounting for isomorphisms) for graphs with less than five vertices.

References

[1] Ugander, J., Karrer, B., Backstrom, L., Marlow, C. (2011). The Anatomy of the Facebook Social Graph. doi: 10.48550/arXiv.1111.4503

[2] Spielman, D.A. (2019). *Spectral and Algebraic Graph Theory*. Retrieved from http://cs-www.cs.yale.edu/homes/spielman/sagt/sagt.pdf.

[3] Telesford, Q.K., Joyce, K.E., Hayasaka, S., Burdette J.H., Laurienti, P.J. (2011). The Ubiquity of Small-World Networks. *Brain Connectivity*, *1*(5), 367-375. doi: 10.1089/brain.2011.0038

[4] Watts, D.J. Strogatz, S.H. (1998). Collectice Dynamics of 'Small-World' Networks. *Nature*, *393*, 440-442. doi: 10.1038/30918

Colorado State University

Joint Mathematics Meeting Undergraduate Poster Session
Hynes Convention Center, January 6, 2023