Graph Theory and the Six Degrees of Separation

Iris Heikkinen

May 2021

Abstract

In this paper, we will introduce the basics of graph theory and learn how it is applied to networks through the study of random graphs, which links the subjects of graph theory and probability together. We will also explore and analyze the concept of the six degrees of separation and how random graphs can be applied to social networks.

1 Introduction to Graph Theory

Observe the image below, which represents the people who Diana knows from her son's middle school. Each line represents a social connection between two people.

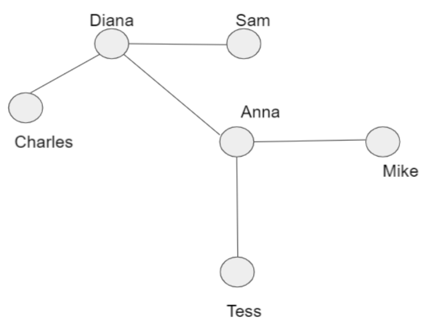

Figure 1: A network of mutual friends.

This figure is a classic example of a graph. A graph is defined as a mathematical figure compiled of connected nodes and edges.

Definition 1. A graph G is made from a set of vertices V(G) and edges E(G). A connection between two vertices can be labeled as (u,v). If e=(u,v) in the graph G, then the vertices u and v are said to be joined by edge e.

Example 1. *In this scenario, each person is represented by a node and each social connection is represented by an edge.*

Definition 2. A vertex v and edge e are said to be incident to each other if they are neighbors of each other. If two edges, p and q, are incident to a common vertex w, then p and q are considered to be adjacent edges.

Example 2. *In our model, edges (Charles, Diana) and (Diana, Anna) would be considered an example of adjacent edges.*

Definition 3. In graph G, the number of vertices is called the order of the graph while the number of edges is called the size. The order of any given graph must be at least 1.

Example 3. *In our model, the order of the graph is 6 and the size of the graph is 5.*

Definition 4. The degree of a graph G is the number of edges incident with a vertex v and is denoted by deg v or degGv. The set N(v) of neighbors of vertex v is called a neighborhood. From there, we can deduce that

$$N(v) = degv \tag{1}$$

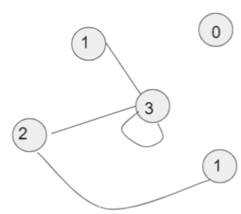

Figure 2: A graph with degrees labeled

Example 4. *As seen in the image above, each vertex of the graph is labeled with the name of a person. Connecting this to what we just learned, we can identify the degree of each person. If we were to look at Diana, we could say that deg(Diana)=3. Similarly N(Diana) would also be 3.*

2 Different Types of Graphs and Common Definitions

We've already introduced the concept of graph theory, so now lets explore the different types of graphs.

Earlier, we defined that the order of any given graph must be at least 1. What would happen in the scenario where we have a graph of order 1?

Definition 5. Whenever we have a graph of order 1, we call it a trivial graph. Similarly, any graph with an order ¿1 would be labeled non-trivial.

Definition 6. Whenever two graphs are complimentary, G and G share a vertex set, and for every distinct pair of vertices (u, v) of G, uv is an edge of G but not G (see example below).

Definition 7. A bipartite graph is a graph consisting of two disjoint sets where every vertex in one of the sets is connected to at least one in the other set. These sets are called partite sets.

Definition 8. A graph can be considered a k-partite graph when V(G) has k partite sets so that no two vertices from the same set are adjacent.

Definition 9. A complete bipartite graph is a bipartite graph where every vertex in the first set is connected to every vertex in the second set.

Definition 10. A walk is defined as a sequence of vertices and edges in a graph. An open walk is whenever the starting and ending vertices are different, and a closed walk is whenever the starting and ending vertices are the same.

Definition 11. A trail is defined as an open walk which does not repeat edges.

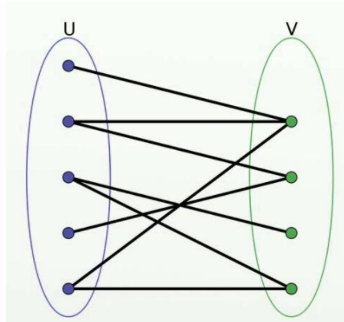

Figure 3: An example of a bipartite graph

Definition 12. A circuit is defined as a closed trail of length 3 or more. Circuits can repeat vertices, but cannot repeat edges.

Definition 13. A path is defined as a trail where neither circuits nor edges can be repeated.

Definition 14. A cycle is a circuit which repeats no vertex other than the first and the last.

3 A Brief Overview of Random Graphs

Random graphs are a term in graph theory which refer to probability distribution in graphs. In order to create a random graph, you begin with a set of n isolated vertices and add consecutive edges to them at random. The study of random graphs is focused around determining what stage a particular property is likely to appear. The most commonly studied graph model is known as the G(n, p) model. Other models exist, however the G(n, p) model will be the main focus of our paper since it is the easiest to analyze.

In the 1950's, mathematician Edgar Gilbert introduced the random graph model G(n, p) where n represents the order of the graph and p represents the probability. In this graph model, a graph with n vertices is constructed so that each pair of vertices (i, j) is connected by probability p. Mathematicians Edros and Renyi came to the conclusion that there are n(n-1)/2 possible edges for a graph of order n. For example, a graph of order 3 would have 3(3-1)/2=6/2=3 possible edges. This equation can be applied to any amount. A graph with order 250k would have it's maximum size be 32 billion. Most graphs have many fewer edge connections than the max.

3.1 The G(n, p) Model

In the G(n, p) model, p is commonly a function of n so that p=k/n for a constant k. For a small p, k¡1, and p=k/n, all connected components of the graph are small. As k increases, there is a much greater chance of there being a giant component. Whenever k¿1, there is a large component made of a constant fraction of the vertices in the graph.

In some scenario's, we might take interest in finding equations that can be used to determine the probability of the G(n, p) graph model being connected. When taking such scenarios into consideration, researchers focus on asymptotic behaviors of graphs. This ties into Percolation theory, which describes the behaviors of networks when vertices or edges are added.

Say we have a random graph with n vertices and an average degree of k.

Let us remove a fraction 1-p of vertices so that only a fraction p remains. A threshold for the Percolation theory exists;

Above Pc, our network will become fragmented, while below Pc, a giant component exists.

$$p_c = \frac{1}{\langle k \rangle}$$

Figure 4: The critical percolation threshold

Theorem 1. *A graph G(n, p) will almost surely be connected if p¿((1+e)ln(n))/n. Similarly, a graph G(n, p) will almost surely be disconnected if p¡((1-e)ln(n))/n.*

Relating our graph back to the social networks, we can measure the probability of friendships and social connections. Say two people randomly meet and become acquainted. The probability of two people forming a friendship can be related back to our p=k/n model. Now, k represents the number of people the person knows. Assuming that each friendship in our model is independent from all other friendships, we can confirm that a giant component will be present in every scenario where a person has more than one friend.

3.2 Phase Transitions

Many random graphs also undergo phase transitions. Phase transitions occur whenever the edge probability passes through a thresh-hold value. Thresh-holds exist when an abrupt transition from not having a property to having a property takes place. A graph is considered a threshold graph if it can be built from the empty graph by adding either an isolated vertex or a dominating vertex repeatedly. Analysis of Edros-Renyi graphs are traditionally preformed with the limit as n as it approaches infinity, and the value p serves as a function of n.

Definition 15.

$$\overset{n\to\infty}{\to} \; 0 \; if \; \frac{p(n)}{t(n)} \overset{n\to\infty}{\to} 0$$

Figure 5: For a given property P, t(n) is a threshold function

Definition 16. For a graph G with property P, a graph with P is monotone if any graph with the same set of vertices containing G as a subgraph also has P. Every non-trivial monotone increasing property of subsets has a threshold function.

3.3 Connectivity

Once we understand threshold functions and how they work within random graphs, we can explore the connectivity of these graphs. Say n represents the number of vertices in a graph, e represents the maximum number of edges between vertices, G(n) represents a random graph, and P represents the property of the graph. Once we define e(n), we are able to find the threshold function for that property (denoted as ep). Graph G(n) has property P if e is greater than ep and does not have P if e is less than ep.

We are interested in finding a good n and e such that all vertices are connected and doesn't conceal topological features.

Definition 17.

$$\varepsilon(n) = \left(\frac{\log(n)}{n}\right)^{\frac{1}{2}}$$

Figure 6: The threshold for points in the plane. In this model, G(n, e) is connected.

4 Graph Theory Applied to Social Networks

Stanley Milgram conducted an experiment in 1967 where he asked people to pass along a letter to someone they didn't know in Boston through other people. Similar to the graph we observed in figure 1, this network was then represented through nodes and edges (nodes represented individual people and edges represented social connections.)

This theory has been since developed and applied to many different scenarios, whether it finds links between actors who have co-starred in a movie or social media profiles.

Perhaps the most famous example of this is the Kevin Bacon Graph, which analyses how many degrees it takes to connect any given actor to Kevin Bacon ("Bacon number" is used to represent the minimum amount of edges between said actor and Bacon).

However, flaws have been found in Milgram's experiment. It was shown that random graphs typically followed a bell-curve when examining the numbers of edges per node, however, on a scale-free network, the number of nodes logarithmically declined as the number of edges increased. Real people do not randomly connect, instead their choices directly influence the people they interact with.

In the early 2000's, Duncan Watts attempted to replicate Milgram's experiment by randomly assigning participant one of 18 people from 13 various countries and telling them to locate them. The participant were then asked about the links used in order to forward messages. Researchers found that the majority (67 percent) of participants used their friends in order to pass the message along. The average degree of separation was found to be 4 degrees. Now that the internet assists us in reaching more people, studies have shown that the average degree of separation between people will continue to shrink.

Random graphs and the six degrees of separation can be combined into one algorithm. Watts and Strogatz proved that the average path between two nodes in a random network is equal to log (n)/log (k). In this model, n is equivalent to the total number of nodes and k represents acquaintances per node. If we set n equal to 300,000, which is 90 percent of the US population, and we set k equal to 30, then the degrees of separation will come out to be 5.7. If we use the world population in this example and make N=90 percent of the world population (6,000,000,000) and keep K=30, then the degrees of separation will be about 6.6 (assuming that the other 10 percent of the population is too young to participate).

5 Conclusion

Throughout this paper we have learned some of the basic definitions used in graph theory and have explored the theory of random graphs. Towards the end of this paper, we discussed how these random graphs have been applied to real life scenarios through social networks. We introduced multiple theorems and definitions that tie into the use of random graphs in higher mathematics.In total, this paper serves as a deeper dive introduction to properties of graphs and I am excited to continue future explorations.

6 References

Frieze, Alan; Karonski, Michal (2015). Introduction to Random Graphs. Cambridge University Press

Newman, M. E. J. (2010). Networks: An Introduction. Oxford.

Etrata, Daniel, "The connectivity threshold on random graphs", "http://mcl.math.uic.edu/wp-content/uploads/2015/11/S16-TAV-poster2.pdf"

Bernick, Lee, "Modeling Human Networks Using Random Graphs",

"https://math.mit.edu/ apost/courses/18.204$_2$018/$Lee_Bernick_paper.pdf$"