1

From Königsberg to Connectomes

Section 1.1 is a broad introduction to graph theory and network science. Sections 1.2 and 1.3 introduce the nuts and bolts of the subject. Section 1.2 describes when two graphs may be considered the same. Section 1.3 describes ways of obtaining new graphs from old ones and ends with a detailed discussion on minors.

1.1 Introduction

Our story begins in eighteenth century Königsberg, now known as Kaliningrad, a Russian city between Poland and Lithuania. The citizens of Königsberg enjoyed taking walks across the seven bridges over the river Pregel. Some had noticed that they could not cross all the bridges exactly once and return home. Swiss mathematician Leonhard Euler heard about this puzzle and wrote a paper on it called "Solution of a problem relating to the geometry of position" (Euler, 1736). His paper had the simple schematic representation of a map shown in Figure 1.1, where the land masses are labeled A, B, C, D. This drawing subsequently became the point and line drawing that is called a graph. The points are called vertices or nodes and the lines are called edges or links.[1]

Euler explained that if there was a way of walking along the edges of the graph, crossing every edge exactly once, and returning to the starting vertex, then one would have to enter and leave a vertex the same number of times. Therefore every

1 An English translation of Euler (1736) appears in Biggs et al. (1976). In Hopkins and Wilson (2004), the authors note that the diagram on the right in Figure 1.1 made its first appearance in 1892 in W.W. Rouse Ball's book *Mathematical Recreations and Problems of Past and Present Times* (Ball, 1892). According to Parks (2012), the first appearance of the word "graph" is in a note by J.J. Sylvester published in *Nature* (Sylvester, 1878a) followed by a long paper on the same topic in the journal he founded, *American Journal of Mathematics* (Sylvester, 1878b).

Graphs and Networks, First Edition. S. R. Kingan.

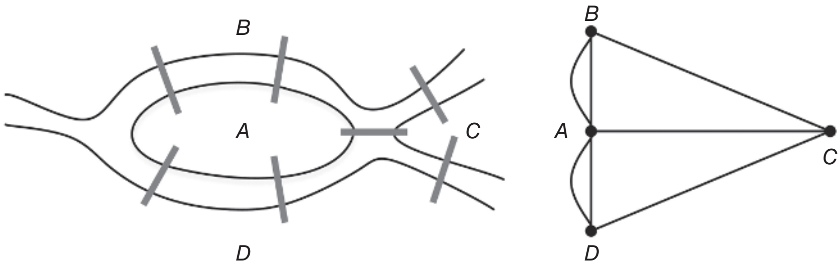


Figure 1.1 The bridges of Königsberg.

vertex must have an even number of edges incident to it. Further, he observed that this necessary condition was also sufficient. Euler did not prove sufficiency. A young German mathematician by the name of Carl Hierholzer gave the first complete proof of this result (Hierholzer, 1873).

Let us change the question slightly and ask "Is there a way of walking along the edges of the graph so as to cross every vertex exactly once, and return to the starting vertex?" This question, posed in the mid-nineteenth century independently by British mathematicians Thomas Kirkman and William Hamilton (Kirkman, 1856; Hamilton, 1858), is still unresolved. There are many theorems giving necessary conditions and many giving sufficient conditions. However, a theorem giving a necessary and sufficient condition, if such a theorem exists, remains elusive.

A *graph* is defined as a finite non-empty set of objects called *vertices* or *nodes* together with a set of unordered pairs of distinct vertices called *edges* or *links*. Typically a graph is denoted by the letter G, the vertex set by $V(G)$ or just V when the context is clear, and the edge set by $E(G)$ or E. The number of vertices is called the *order* of G and the number of edges is called the *size* of G. Throughout the book n stands for the number of vertices and m stands for the number of edges. All this information is summarized by writing $G = (V, E)$.[2]

The graph in Figure 1.2 is called the prism graph since it looks like a flattened rectangular prism viewed from the top. It has vertex set

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6\},$$

2 The terminology comes from set theory. The number of elements in a set A is denoted by $|A|$. If x is an element of set A, we write $x \in A$, otherwise we write $x \notin A$. If B is a subset of set A, we write $B \subseteq A$, otherwise we write $B \nsubseteq A$. The empty set ϕ and A itself are called improper subsets of A. All other subsets are called proper subsets. The union of sets A and B, denoted by $A \cup B$, is the set of all elements that are either in A or in B. The intersection of sets A and B, denoted by $A \cap B$, is the set of all elements that are in both A and B. The set of elements in A and not in B is denoted by $A - B$. We assume a set A is a subset of a universal set U. The complement of A, denoted by $\overline{A}$, is the set $(U - A)$. The symmetric difference between two sets A and B is $A \triangle B = (A - B) \cup (B - A) = A \cup B - A \cap B$.

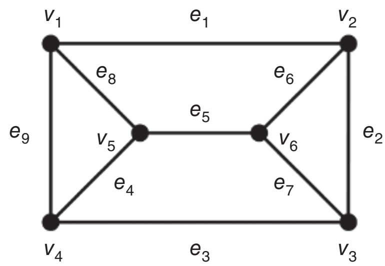


Figure 1.2 The prism graph.

and edge set

$$\{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\},$$

where $e_1 = (v_1, v_2)$, $e_2 = (v_2, v_3)$, $e_3 = (v_3, v_4)$, $e_4 = (v_4, v_5)$, $e_5 = (v_5, v_6)$, $e_6 = (v_2, v_6)$, $e_7 = (v_3, v_6)$, $e_8 = (v_1, v_5)$, and $e_9 = (v_1, v_4)$. The edges are unordered pairs of vertices, so we may write, for example, $e_1 = v_1v_2$. Two vertices linked by an edge are called *adjacent vertices*. Two edges with a common vertex are called *adjacent* edges. For example, v_1 and v_2 are adjacent vertices and e_1 and e_2 are adjacent edges. The end vertices of an edge are said to be *incident* to the edge. For example, v_1 is incident to edge e_1.

The *degree* of a vertex v, denoted by $deg(v)$, is the number of edges incident to it. In a graph with n vertices, the degree of each vertex is at most $n - 1$ since each vertex can be joined to at most $n - 1$ of the remaining vertices. The *degree sequence* of a graph is an unordered list of the degrees. The *minimum degree* is denoted by $\delta(G)$ and the *maximum degree* is denoted by $\Delta(G)$. If all the vertices have the same degree, then the graph is called a *regular graph*. For example, the prism graph in Figure 1.2 is a regular graph since every vertex has degree 3. A regular graph of degree 3 is called a *cubic graph*. The *neighborhood* of vertex v, denoted by $N(v)$, is the set of all vertices joined by an edge to v. By convention v is not in $N(v)$.

The graph with just one vertex is called the *trivial graph* and a graph with vertices, but no edges, is called an *isolated graph*. A vertex with no edges incident to it is called an *isolated vertex*.

A *directed graph* (also called *digraph*) is a graph whose edges have a direction. The directed edges, called *arcs*, are ordered pairs of vertices. The arrow on arc (u, v) points from u to v. Vertex u is called the *initial vertex* and vertex v is called the *terminal vertex*. The *outdegree* of vertex v, denoted by $outdeg(v)$, is the number of arcs that have v as the initial vertex. The *indegree* of vertex v, denoted by $indeg(v)$, is the number of arcs that have v as the terminal vertex. The *degree sequence of a digraph* with n vertices, where each vertex v_i has indegree r_i and outdegree s_i, is given by the list of ordered pairs $(r_1, s_1), (r_2, s_2), \dots, (r_n, s_n)$.

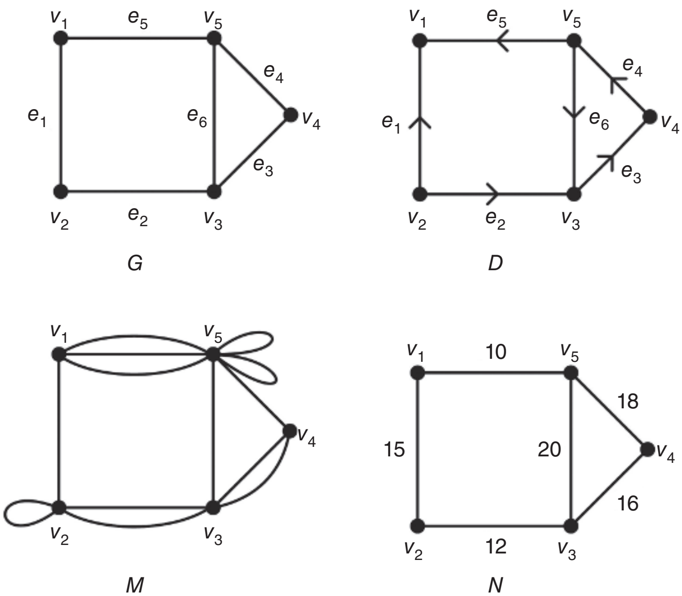


Figure 1.3 An example of a graph, multigraph, digraph, and network.

A *loop* is an edge from a vertex to itself. Multiple edges between a pair of vertices are called *parallel edges*. A *multigraph* is a graph with loops or parallel edges. The Königsburg bridge graph shown in Figure 1.1 is a multigraph.[3]

A graph that has a real number associated with each edge is called a *network*.[4] The numbers are called *weights* or *costs*. Figure 1.3 gives an example of a graph, digraph, multigraph, and network.

A complex network is a real-world network, so large and complicated that it can only be studied empirically. Vertices and edges of a complex network may not be fully known or may be constantly changing. Complex networks are everywhere. Examples include:

- The neurons in the brain linked by synapses (called a connectome);
- The Internet with wired or wireless links between computers;
- The World Wide Web with hypertext links connecting webpages;

---

3 Sometimes a graph is called a simple graph and a multigraph is called a graph. When reading a research paper, the first thing to determine is what exactly the author calls a graph.
4 The term network is used loosely in network science literature to mean graph, digraph, multigraph, network, or some graph–like structure. The rationale is that network should be used when talking about a real-world graph regardless of the type of graph.

- Networks of people where the link is based on some sort of relationship (called social networks); and
- Networks of people where the link is catching a contagious disease such as Ebola or COVID-19 (called epidemic or pandemic networks).

The study of complex networks has become an established interdisciplinary field called Network Science that is populated by physicists, biologists, computer scientists, and social scientists.

A *walk* in a graph is an ordered sequence of vertices and edges

$$v_1, e_1, v_2, e_2, \dots, v_{k-1}, e_{k-1}, v_k,$$

where $e_{i-1} = v_{i-1}v_i$ is an edge for $2 \le i \le k$.[5] Vertices and edges may be repeated in a walk. A *trail* is a walk with no edge repeated. A *path* is a walk with no vertex repeated. A *closed walk* is one in which the first and last vertex are the same. Closed trails and paths are defined in a similar manner. A closed trail is called a *circuit*. A closed path is called a *cycle*.

For example, in the prism graph shown in Figure 1.2

$$v_1e_1v_2e_2v_3e_3v_4e_3v_3e_7v_6$$

is a walk from v_1 to v_6. Notice how vertices and edges are repeated. In a trail vertices may be repeated, but not edges. For example,

$$v_2e_2v_3e_7v_6e_6v_2e_1v_1$$

is a trail from v_2 to v_1. In a path vertices are not repeated, and consequently edges are not repeated. For example,

$$v_1e_1v_2e_2v_3e_3v_4$$

is a path from v_1 to v_4 and

$$v_1e_1v_2e_2v_3e_3v_4e_9v_1$$

is a cycle. For paths and cycles we may use just vertices or just edges since there are no repetitions and therefore no cause for confusion. The aforementioned path may be written as $v_1v_2v_3v_4$ and the cycle may be written as $v_1v_2v_3v_4v_1$. Alternatively, we may refer to this particular path and cycle by their edges as $e_1e_2e_3$ and $e_1e_2e_3e_9$, respectively.

Two walks, trails, or paths are *equal* if they have exactly the same vertices and edges. The *length* of a walk, trail, path, circuit, or cycle is the number of edges in it (counting repetitions when relevant). The *girth* of a graph is defined as the length of the shortest cycle in the graph. A cycle of length 3 is called a *triangle*.

---

5 It is customary to write the phrase "$i \in \{2, 3, \dots, k\}$" in short as "$2 \le i \le k$," recognizing that in this context i is a natural number.

A circuit that crosses every edge in the graph exactly once is called an *Eulerian circuit*, and a graph with an Eulerian circuit is called an *Eulerian graph*. A cycle that crosses each vertex in the graph exactly once is called a *Hamiltonian cycle*, and a graph with a Hamiltonian cycle is called a *Hamiltonian graph*. As mentioned earlier. Euler gave a complete characterization of graphs with Eulerian circuits, however, a necessary and sufficient condition for the presence of a Hamiltonian cycle is not known.

Suppose a salesman travels to each of n cities exactly once and returns home. The n cities form the vertices of a graph and the routes between cities are the edges. The weights on the edges are the costs of traveling from one city to the other. Consider the problem of finding the cheapest route. For example, in the network shown in Figure 1.4, we could list all possible routes (that would be $5! = 120$ routes) to check which one is the cheapest route. This brute-force approach would work for five cities, but what if the traveling salesman is visiting 100 cities? The number of possible routes is 100!. Is there an efficient algorithm? No one knows. This is called the Traveling Salesman Problem, and it is one of the Clay Institute's million dollar problems.[6]

Walks, trails, paths, and cycles in a digraph are defined in a similar manner with the understanding that the direction of the edges must be taken into account while navigating the digraph. A *directed path* in a digraph is an ordered sequence of distinct vertices $v_1, v_2, \dots, v_k$, where (v_{i-1}, v_i) is an arc for $2 \leq i \leq k$. A *directed cycle* is a closed directed path.

A non-trivial graph is *connected* if there is a path between every pair of vertices. Otherwise the graph is *disconnected*. By convention the trivial graph is considered to be connected. The *distance* between a pair of distinct vertices u and v in a connected graph G, denoted by $d(u, v)$, is the length of the shortest path between u and v. The *diameter*, denoted by $diam(G)$, is the maximum distance between any pair of vertices. For example, the prism graph shown in Figure 1.2 has diameter 2.

A *connected digraph* is one whose underlying graph is connected. If, for every pair of vertices u and v, there are directed paths from u to v and from v to u, then the digraph is called *strongly connected*. The digraph D in Figure 1.3 is connected,

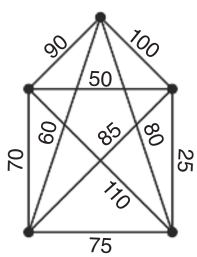


Figure 1.4 Traveling salesman network.

6 See https://www.claymath.org/millennium-problems/p-vs-np-problem.

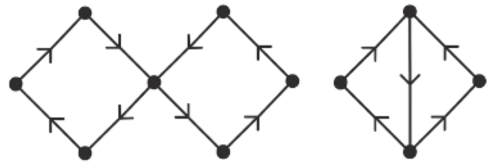


Figure 1.5 Strongly connected digraphs.

but not strongly connected. There is no directed path from v_1 to v_2. Figure 1.5 displays two strongly connected digraphs.

In a strongly connected digraph the *directed distance* between vertices u and v is the length of the shortest directed path from u to v. Note that in a digraph $d(u, v)$ is not necessarily the same as $d(v, u)$. The diameter of a strongly connected digraph is the maximum distance between any pair of vertices. The first digraph in Figure 1.5 has diameter 4 and the second has diameter 3.

The graph in which every pair of vertices is joined by an edge is called a *complete graph*. The complete graph on n vertices is denoted by K_n (see Figure 1.6). The number of edges in K_n is $\binom{n}{2} = \frac{n(n-1)}{2}$, since every pair of vertices is joined by an edge. Similarly, the digraph in which every pair of vertices u and v is joined by arcs (u, v) and (v, u) is called a *complete digraph*. The number of arcs in a complete digraph on n vertices is $n(n-1)$.

Next, let us look at some frequently used examples of graphs. These are the toy examples we will use to understand new concepts. The path P_n, for $n \geq 1$, the cycle graph C_n, for $n \geq 3$, and the wheel graph W_{n-1}, for $n \geq 4$ are three families of graphs shown in Figure 1.7. Observe that for P_n the number of edges is $m = n - 1$; for C_n the number of edges is $m = n$; and for W_{n-1} the number of edges is $m = 2(n-1)$ since the number of spokes in the wheel graph is $n - 1$.

A graph is called *bipartite* if its vertex set V can be written as $V = V_1 \cup V_2$, where V_1 and V_2 are disjoint and every edge joins a vertex in V_1 to a vertex in V_2. A graph is called *complete bipartite* if every vertex in V_1 is joined to every vertex in V_2. A complete bipartite graph with n_1 vertices in V_1 and n_2 vertices in V_2 is denoted by K_{n_1,n_2}.

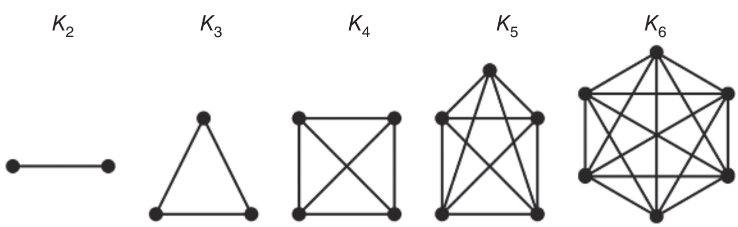


Figure 1.6 Complete graphs.

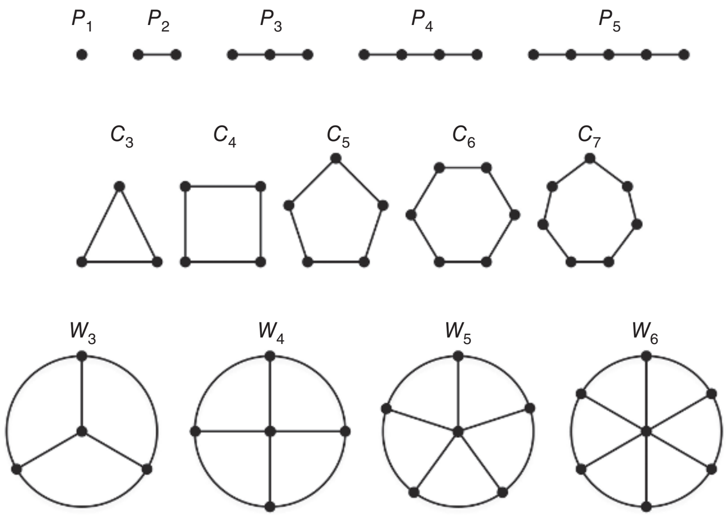


Figure 1.7 Paths, cycles, and wheels.

We can generalize this notion to t-partite graphs, where the vertex set is partitioned into t pairwise disjoint subsets in such a way that every edge joins a vertex in V_i to a vertex in V_j, for $i,j \in \{1,2,\ldots,t\}$ and $i \neq j$. A graph is called *complete t-partite* if every vertex in V_i is joined to every vertex in V_j. A complete t-partite graph with n_i vertices in V_i is denoted by $K_{n_1,\ldots,n_t}$. Figure 1.8 shows examples of bipartite and tripartite graphs. The complete bipartite graph $K_{1,n-1}$ with n vertices is called the *star graph*. See Figure 1.9.

Graphs with no cycles are called *acyclic* graphs. A connected acyclic graph is called a *tree*. An acyclic graph is called a *forest*. The terminology is evocative of trees in a forest. The term tree appears in an 1857 paper by British mathematician

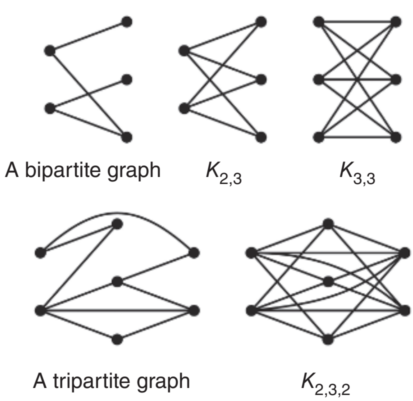


Figure 1.8 Bipartite and tripartite graphs.

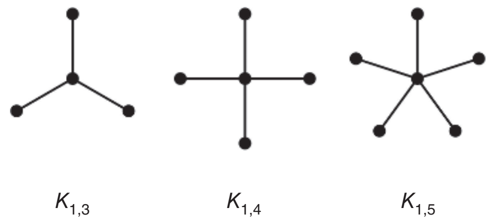


Figure 1.9 Star graphs.

Arthur Cayley titled "On the theory of the analytical forms called trees" (Cayley, 1857).[7] Cayley was trying to enumerate the isomers of saturated hydrocarbons such as methane, ethane, butane, etc., which have the form C_kH_{2k+2}. An isomer is a molecule with the same chemical formula as another molecule, but with a different arrangement of atoms and different properties. These molecules can be represented by trees as shown in Figure 1.10.

There are many ways to define a substructure of a graph. The most obvious way is to delete some vertices and edges. A graph H is a *subgraph* of a graph G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. In the definition of subgraph we understand that if an edge is in $E(H)$, then both its end vertices are in $V(H)$. A *proper subgraph* of G is a subgraph that is not the empty set or G. A subgraph H with $V(H) = V(G)$ is called a *spanning subgraph*. Figure 1.11 displays a graph G and three subgraphs H, I, and J. Observe that J is a spanning subgraph.

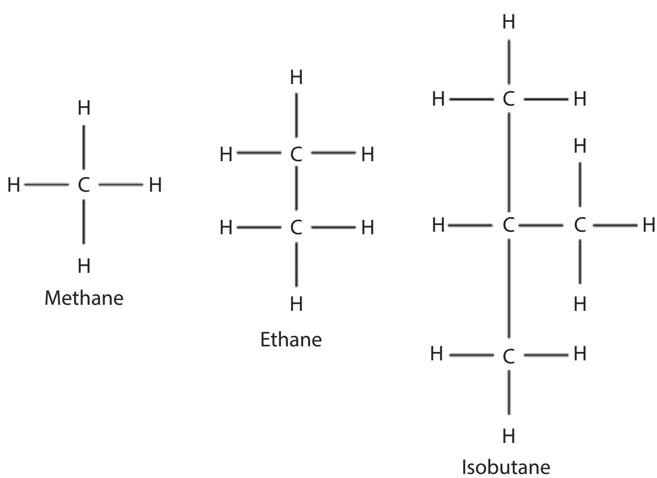


Figure 1.10 Example of trees.

7 According to Biggs et al. (1976) the concept of a tree appears nearly ten years earlier in papers by Kirchhoff (1847) and von Staudt (1847).

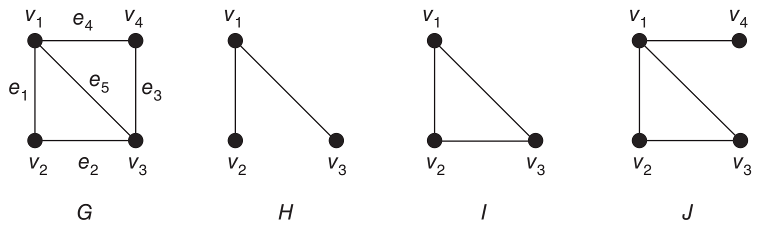


Figure 1.11 Subgraphs.

Sometimes we may want to pick a subset of vertices and look at the subgraph obtained by taking all the edges present that join those vertices. Such a subgraph is said to be "induced" by the subset of vertices. Let $U \subseteq V(G)$. The subgraph induced by U, called an *induced subgraph*, has vertex set U and edge set those edges in G incident with two vertices in U. For example, subgraph I in Figure 1.11 is induced by $\{v_1, v_2, v_3\}$, whereas subgraph H is not an induced subgraph.

A subgraph that is a complete graph is called a *clique*. A *maximal clique* in a graph is a complete subgraph that cannot be made larger by adding more vertices. A *maximum clique* is a maximal clique that has the most number of vertices.

A graph is called *planar* if it can be drawn on the plane without crossing edges. Otherwise it is called *non-planar*. Figure 1.12 gives an example of a planar graph. In the first rendition it is drawn with edges crossing. However, a re-drawing shows that it can be drawn without edges crossing. This graph is K_5 with one edge deleted. What about K_5 and $K_{3,3}$? A few attempts to draw them without crossing edges will reveal that it is impossible. This is because K_5 and $K_{3,3}$ are non-planar.

A map corresponds to a type of planar graph. Vertices can be placed on the countries and two vertices are joined by an edge if the corresponding countries share a common border. Mapmakers have known since antiquity that at most four colors are needed to color two adjacent countries with different colors (countries that meet in a point may be colored the same color). This belief came to be known as the "Four Color Conjecture." It was brought to the attention of British mathematician Augustus de Morgan in 1852 by one of his students Federick Guthrie, who attributed it to his brother Francis Guthrie (Biggs et al., 1976). De Morgan

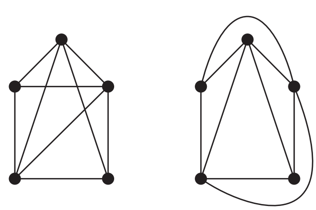


Figure 1.12 Two drawings of a planar graph.

wrote about it to Hamilton who replied saying he did not find it interesting. In 1976 two American computer scientists from the University of Illinois at Urbana Champaign, Kenneth Appel and Wolfgang Haken, proved the Four Color Conjecture. Their proof is notable in that their mathematical arguments reduced the proof to case-checking of nearly 10 000 cases for which they used a computer. Initially some mathematicians were suspicious of a proof with a computer component, but in time that attitude changed. To commemorate the occasion the university postmark had the slogan "Four colors suffice." A subsequent proof with fewer cases has also been published (Robertson et al., 1997a).

In 1930 Polish mathematician Kazimierz Kuratowski proved that every non-planar graph contains either K_5 or $K_{3,3}$ as a substructure called a subdivision (Kuratowski, 1930). A subdivision of a graph is obtained by placing vertices on edges. Klaus Wagner built on this by proving that every non-planar graph contains either K_5 or $K_{3,3}$ as a substructure called a minor (Wagner, 1937). A minor of a graph is obtained by deleting edges (and any isolated vertices formed as a result) and "contracting edges." An edge is contracted by shrinking it and fusing its end vertices into one vertex. Subdivisions and minors are described in detail in Section 1.3.

A class of graphs $\mathcal{G}$ is said to be *closed under minors* if every minor of a graph in $\mathcal{G}$ is also in $\mathcal{G}$. A graph that is not in $\mathcal{G}$ is called a *minimal excluded minor*[8] if all of its proper minors are in $\mathcal{G}$. Wagner conjectured that every infinite class of graphs closed under minors has a finite set of minimal excluded minors. This conjecture was confirmed by Neil Robertson and Paul Seymour in a series of 23 papers the first of which appeared in 1983 (Robertson and Seymour, 1983) and the last of which appeared in 2010 (Robertson and Seymour, 2010). These papers totaling over 500 pages proved the conjecture as well as dozens of related results and mapped out a thriving research genre.

In the 1930s considerable progress was made in graph theory. In addition to Kuratowski and Wagner's structural result, Karl Menger proved a major connectivity result (Menger, 1927), Frank Ramsey introduced what we now call Ramsey Theory (Ramsey, 1930), Phillip Hall solved the Marriage Problem (Hall, 1935), Hassler Whitney introduced matroids (Whitney, 1935), and Dénes König wrote the first graph theory textbook (König, 1936).[9] Less well known is a graph theoretical development in social psychology. At the 1933 meeting of the New York Medical Society, psychiatrist Jacob L. Moreno presented a graph showing the friendship of boys and girls in a fourth grade class. His work appeared in the New York Times

8 In general, a set X is called *minimal* with respect to a certain property if none of its proper subsets have the property. Note that minimal does not mean minimum.

9 In 2021, Martin Charles Golumbic published an annotated translation of Les reseaux (ou graphes), written by Andre Sainte-Lague in 1926. Golumbic called it "The Zeroth Book of Graph Theory."

in an article dated April 3, 1933 titled "Emotions mapped by Geography." He is credited with making a far-sighted statement:

> If we ever get to the point of charting a whole city or a whole nation, we would have a picture of a vast solar system of intangible structures, powerfully influencing conduct, as gravitation does in space. Such an invisible structure underlies society and has its influence in determining the conduct of society as a whole.

Moreno's friendship graph is cited by sociologists as the first example of a social network. His 1934 book *Who Shall Survive: A New Approach to the Problem of Human Interrelations* (Moreno, 1934) has, among other things, several examples of social networks (he called them sociograms) with detailed instructions for drawing them using color, size, and shape of vertices to convey information.

For small graphs it makes sense to list the degrees of the vertices and talk about the degree sequence. For large graphs the list of degrees is too long. So we talk about the frequency distribution of the degrees. The *degree distribution* of a graph is a frequency histogram with the degrees on the x-axis and the number of vertices of each degree on the y-axis. The *indegree distribution* and *outdegree distribution* of a digraph are defined in a similar manner.

In a 1965 study of the citation digraph, where vertices are scientific papers and an arc from u to v indicates u cites v, physicist and historian Derek de Solla Price observed that the degree distribution of the citation digraph followed the power law function, $f(x) = cx^{-n}$, where c and n are constants. Most papers had three to seven citations and only a few papers had many citations (Price, 1965). Graphs whose degree distributions follow the power law function are called *scale-free networks*. The name scale-free comes from the observation that since most vertices have small degree and a few have large degree, the average of the degrees is meaningless.

Meanwhile social psychologist Stanley Milgram conducted his "small-world" experiment, where he randomly selected a group of individuals and gave them letters addressed to someone they didn't know. The goal of the experiment was for each person to send the letter to a person they knew, who in turn sent the letter to a person they knew, and so on. On average Milgram found that the letters reached their destination through six connections (Milgram, 1963). Subsequently these ideas led to the notion of small-world graphs (Watts and Strogatz, 1998).

A centrality measure is a way of quantifying the most important people in a social network. This approach to determining the importance of an individual based on relations with others in the network, and not just on the individual's attributes, was evidently a paradigm shift in sociological thinking. In 1998 Sergey Brin and Larry Page, two Stanford University students, developed a new type of centrality measure called PageRank to determine the importance of a webpage (Brin and Page, 1998). Their company, Google, needs no introduction. In 2004 Harvard student Marc Zuckerberg and his company Facebook made social networks a household term. In 2016 fake news spread through networks of like-minded individuals creating echo-chambers and 2020 brought us the coronavirus pandemic and the unprecedented quarantine and lock-down. Contact-tracing to form the network of people exposed to COVID-19 entered our lexicon. The killing of unarmed black men and women created an uprising and the "Black Lives Matter" movement was born. Predictive-policing based on a smattering of mathematics including graph theory and a large dose of bias appears to have fallen out of favor, for the time being at least.

With this brief initiation into tribal knowledge, let us begin with what is universally considered the first proposition in graph theory.

Proposition 1.1.1. *The sum of the degrees of the vertices in a graph is twice the number of edges.*

Proof. Each edge has exactly two end vertices and therefore contributes exactly two to the sum of the degrees. □

Let G be a graph with n vertices and m edges. We may write Proposition 1.1.1 as

$$\sum deg(v) = 2m.$$

Proposition 1.1.2. *The number of vertices of odd degree is even.*

Proof. Let E be the set of vertices of even degree and O be the set of vertices of odd degree. By Proposition 1.1.1

$$\sum_{v \in E} deg(v) + \sum_{v \in O} deg(v) = 2m.$$

Observe that $\sum_{v \in E} deg(v)$ is even because it is a sum of even numbers. Moreover, $2m$ is clearly even, so $\sum_{v \in O} deg(v)$ must be even. Since each vertex has odd degree, the only way to get an even sum is if the number of odd degree vertices is even. □

Proposition 1.1.2 is called the Handshake Lemma. At a party, consider the people as vertices and assume that two people shaking hands are linked by an edge. We can be certain that the number of people who greet an odd number of acquaintances is even.

Proposition 1.1.3. *In a digraph the sum of the indegrees is equal to the sum of the outdegrees and this sum is equal to the number of arcs.*

Proof. Every arc leaves exactly one vertex and enters exactly one vertex. So every arc contributes one to the sum of the indegrees and one to the sum of the outdegrees. □

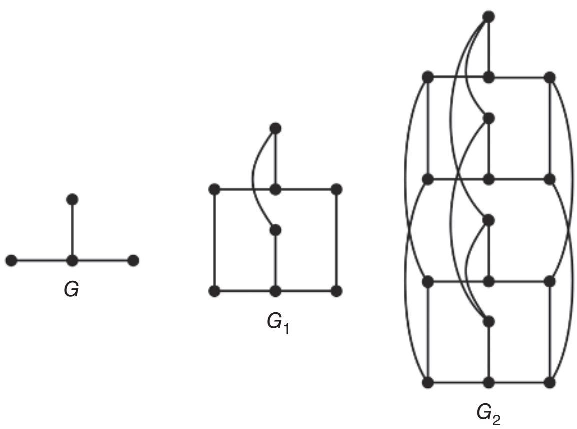
Let G be a digraph with n vertices and m arcs. We may write Proposition 1.1.3 as

$$\sum indeg(v) = \sum outdeg(v) = m.$$

The last proposition in this section is a straightforward example of a constructive proof (König, 1936).

Proposition 1.1.4. *Let G be a graph with maximum degree $\triangle(G)$. There exists a regular graph of degree $\triangle(G)$ that contains G as an induced subgraph.*

Proof. If G is a regular graph, then there is nothing to prove. Therefore suppose G is not a regular graph. Let G' be a copy of G and link the corresponding vertices in G and G' whose degrees are less than $\triangle(G)$. Call the resulting graph G_1. If G_1 is regular, then G_1 is the required regular graph containing G as an induced subgraph. If not, let G_1' be a copy of G_1 and link corresponding vertices whose degrees are less than $\triangle(G)$. Continue like this until we arrive at a regular graph G_k of degree $\triangle(G)$, where $k = \triangle(G) - \delta(G)$, as shown in the following diagram. □



A graph G and a regular graph with G as an induced subgraph.

Note that the proof of Proposition 1.1.4 gives a method for constructing a regular graph containing G as an induced subgraph, but not necessarily the smallest regular graph containing G.

Returning to complex networks, the graph that consists of neurons as vertices and the synapses that connect neurons as edges is called a *connectome*. Depending on the context, it refers to the neurons in the brain or to the neurons in the entire nervous system. The idea that the nervous system is composed of a finite number of nerve cells linked to other cells forming a digraph goes back to Spanish pathologist Santiago Ramón y Cajal in the beginning of the twentieth century. At that time, the "Neuron doctrine" was controversial. Camillo Golgi, inventor of a silver staining technique in 1873, had proposed that the nervous system was a continuous single thread. Cajal proposed a discrete model for the brain of a chicken (Garcia-Lopez et al., 2010). The controversy was settled in favor of Cajal after the electron microscope was invented in 1931.

The first creature whose nervous system was completely mapped was a tiny nematode called *Caenorhabditis elegans* (*C. elegans*). It has 302 neurons that belong to two distinct and independent nervous systems: a large somatic nervous system with 282 neurons (including three isolated neurons that we will disregard) and a small pharyngeal nervous system with 20 neurons. A digraph of the large somatic nervous system with 279 vertices and 2993 directed edges is shown in Figure 1.13 and its indegree and outdegree distributions are shown in Figure 1.14.[10]

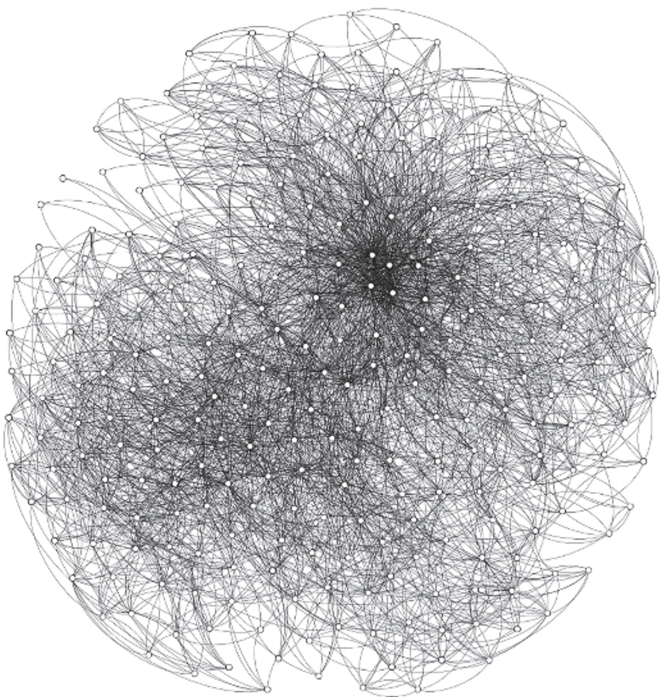


Figure 1.13 *C. elegans* connectome.

10 The data is taken from the Worm Atlas Project (https://wormatlas.org/neuronalwiring.html) and drawn in Gephi 0.9.2 using the Fruchterman-Reingold layout. The data includes a

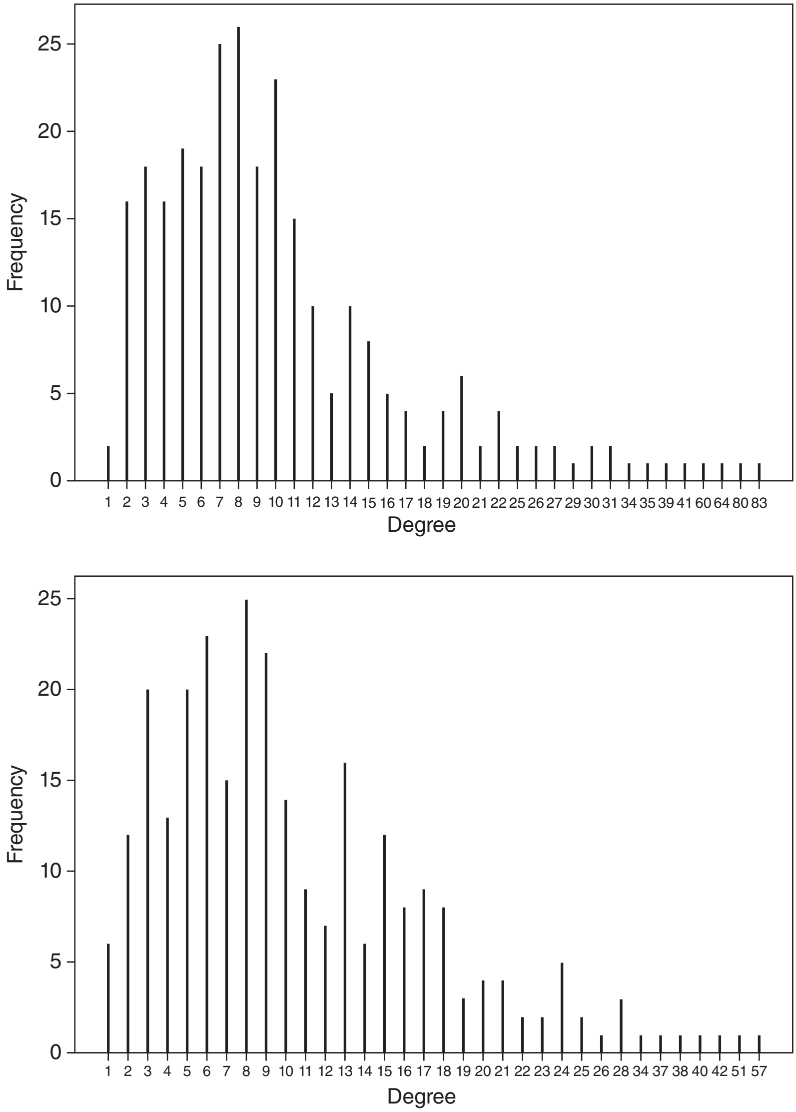


Figure 1.14 *C. elegans* in-degree (top) and out-degree (bottom) distributions.

---

spreadsheet named NeuronConnect.xls, which lists 6418 links between 279 distinct neurons. Once duplicate edges are removed, 2993 distinct directed links remain. Of the 302 neurons in the *C. elegans* nervous system, 20 are part of the pharyngeal system and are not included, two have no connections, and one connects only to muscle tissue, leaving 279. The cited data source

An examination of a large graph begins with an analysis of its degree distribution since a visual representation may not be of much use. The *C. elegans* connectome is a scale-free network. A few vertices have high degree and most have small degree. Vertices of high degree are called *hubs*. The phenomenon that results in scale-free networks is called *preferential attachment*. The term conveys how newly added vertices connect preferentially to existing vertices maintaining and enhancing hubs.

Let G be a graph with n vertices and m edges. The *mean degree* of G (often called *average degree*), denoted by $d(G)$, is given by

$$d(G) = \frac{1}{n} \sum deg(v).$$

It follows from Proposition 1.1.1 that

$$d(G) = \frac{2m}{n}.$$

The *density* of G is the number of edges divided by the maximum possible number of edges in the graph. In symbols,

$$density(G) = \frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}.$$

The corresponding definitions for a digraph are similar except that the maximum number of arcs in a digraph is $n(n-1)$. Let D be a digraph with n vertices and m arcs. The mean degree of D is

$$d(D) = \frac{m}{n}$$

and the density of D is

$$density(D) = \frac{m}{n(n-1)}.$$

For example, the *C. elegans* connectome (the large somatic nervous system with 279 vertices and 2993 directed edges) has mean degree and density as follows:

$$d(G) = \frac{m}{n} = \frac{2993}{279} \approx 10.728,$$

$$density(G) = \frac{m}{n(n-1)} = \frac{2993}{279(279-1)} \approx 0.039.$$

Other connectomes can be analyzed in a similar manner.

In an epidemic network or a pandemic network, the nodes are living entities like people, animals, or plants and a directed edge from one node to another indicates transmission of a contagion. The average degree is denoted by R_0 and is called the *reproduction number* because when a disease is transmitted to a person

---

does not include data on the pharyngeal system; that data can be obtained from the WormWeb project, at http://wormweb.org/details.html.

it reproduces itself. In order for an outbreak to show signs of ending, R_0 must become less than 1. For example, the 2009 outbreak of the H1N1 virus (swine flu) had an R_0 value between 1.4 and 1.6, whereas the 1918 outbreak had an R_0 value between 1.4 and 2.8. Seasonal strains of influenza have an R_0 value between 0.9 and 2.1 (Coburn et al., 2009). The jury is still out on the R_0 value of the COVID-19 pandemic network.[11]

We will end this section with some additional information on connectomes. The function of a neuron is to receive information from some neurons, process that information, and send it to other neurons. A neuron consists of three parts: the cell body (soma); the axon, which looks like a long tail and transmits messages from the nucleus; and the dendrites which look like fine branches of a tree and receive messages for the nucleus.[12] Cajal classified neurons based on the shapes and sizes of the nucleus, axon, and dendrites. He compared the shapes of neurons that he saw to forms in nature such as flowers, trees, branches, roots, etc. and was the first to recognize that neurons were discrete members of what he called the "garden of neurons." However, Sebastian Seung writes in his book *Connectome: How the brain's wiring makes us who we are*, that Cajal's approach to classifying neurons was "much as a nineteenth-century naturalist would have classified different species of butterflies" (Seung, 2012). Seung raises the possibility that there may be another way of classifying neurons.

The brain is divided into different regions based on function. The process of identifying parts of the brain involved in different functions dates back to 1891 when French neurosurgeon Paul Broca and German neurologist Carl Wernicke located the region in the brain involved in language processing. Broca's area, located in the lower part of the frontal lobe, is involved in the formation of sentences. Wernicke's area, located in the temporal lobe, is involved in the comprehension of speech. Wernicke hypothesized the existence of a bundle of long axons connecting Broca's and Wernicke's regions. He predicted that damage to these links will leave both speech production and comprehension intact, yet it would be impossible for a person to hear someone talking and repeat their words. Subsequently, several patients with this problem called "conduction aphasia" were found. Thus Wernicke established that links between regions are just as important as the regions.[13]

The network consisting of regions as vertices and links between regions as edges is called the *regional connectome*. The goal of the Human Connectome Project

11 The article "A guide to R - the pandemic's misunderstood metric" describes the pitfalls of using R_0 in making policy (Adam, 2020). See also "Case clustering emerges as key pandemic puzzle" (Kupferschmidt, 2020).

12 Robert Stufflebaum's *Mind project* located at http://www.mind.ilstu.edu/.

13 See http://thebrain.mcgill.ca.

launched in 2010 by the National Institutes of Health is to obtain a thorough understanding of the regional connectome. It is however a shortcut, Seung says, describing the approach of dividing the brain into regions as follows:

> Neuroscientists continue to argue over their classification. Their disagreements are a sign of a more fundamental problem: It's not even clear how to properly define the concepts of "brain region" and "neuron type." In Plato's dialogue *Phaedrus*, Socrates recommends "division…according to the natural formation, where the joint is, not breaking any part as a bad carver might." This metaphor vividly compares the intellectual challenge of taxonomy to the more visceral activity of cutting poultry into pieces. Anatomists follow Socrates literally, dividing the body by naming its bones, muscles, organs, and so on. Does his advice also make sense for the brain?

Seung proposes developing a regional classification of the brain based on the manner in which neurons are linked. He says linked groups of neurons have a common "connectional fingerprint," and these substructures could then be used to define the "regions" of the brain. It is an intriguing idea; one that may benefit from the application of graph structure theorems.

1.2 Isomorphism

We say two graphs G_1 and G_2 are *isomorphic* if there is a bijective function (a one-to-one and onto function) from $V(G_1)$ to $V(G_2)$ that preserves adjacencies. Otherwise they are *non-isomorphic*. In symbols we write $G_1 \cong G_2$, if there is a bijective function f from $V(G_1)$ to $V(G_2)$ such that

$$(u, v) \in E(G_1) \text{ if and only if } (f(u), f(v)) \in E(G_2).$$

Figure 1.15 shows three pairs of isomorphic graphs: $G_1 \cong G_2$, $G_3 \cong G_4$, and $G_5 \cong G_6$. It is easy to see G_1 is isomorphic to G_2. The bijective function f from $V(G_1)$ to $V(G_2)$ given by $f(1) = b$, $f(2) = c$, $f(3) = d$, $f(4) = a$ is an isomorphism because it preserves adjacencies. Preserving adjacencies means edge $(1, 2)$ is mapped to edge (b, c), edge $(2, 3)$ is mapped to edge (c, d), edge $(3, 4)$ is mapped to edge (d, a), edge $(1, 4)$ is mapped to (b, a), and edge $(1, 3)$ is mapped to (b, d).

The isomorphism between the second pair G_3 and G_4 is less obvious. The bijective function g from $V(G_1)$ to $V(G_2)$ given by $g(1) = a$, $g(2) = c$, $g(3) = e$, $g(4) = b$, $g(5) = d$, and $g(6) = f$ is the required isomorphism.

The isomorphism between the third pair G_5 and G_6 is given by the bijective function h from $V(G_1)$ to $V(G_2)$ given by $h(1) = a$, $h(2) = e$, $h(3) = c$, $h(4) = b$, $h(5) = d$.

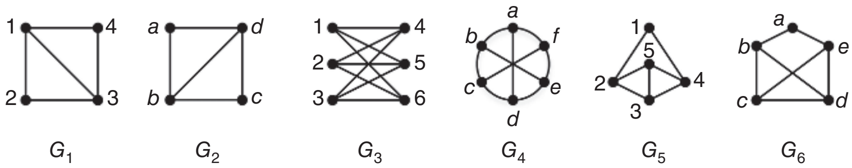


Figure 1.15 Three pairs of isomorphic graphs.

The relation "is isomorphic to" is an equivalence relation on the set of graphs.[14] To see this, observe that isomorphism is:

(1) Reflexive since a graph is isomorphic to itself by the identity function.
(2) Symmetric since if $G_1 \cong G_2$ by the bijection f, then $G_2 \cong G_1$ by the inverse function f^{-1} that exists and is a bijection.
(3) Transitive since if $G_1 \cong G_2$ and $G_2 \cong G_3$ by the bijections f and g, respectively, then $G_1 \cong G_3$ by the composite function $g \circ f$ that exists and is a bijection.

Isomorphism partitions the set of graphs into subsets called equivalence classes.[15] Two graphs G_1 and G_2 are in the same equivalence class if and only if $G_1 \cong G_2$. All graphs in the same equivalence class are isomorphic to each other and any one graph from the class serves to represent the entire class. Figure 1.16 shows the non-isomorphic graphs on 1, 2, 3, and 4 vertices.

An isomorphism from a graph to itself is called an *automorphism*. For example, consider the graph G_1 in Figure 1.15. The function $f : \{1,2,3,4\} \longrightarrow \{1,2,3,4\}$ given by $f(1) = 3, f(2) = 4, f(3) = 1, f(4) = 2$ is an automorphism of G_1.

An *invariant* of a graph G is a property that is the same for any graph isomorphic to G. For example, the number of vertices, the number of edges, and the degree sequence are graph invariants. To determine if two graphs G_1 and G_2 are isomorphic, we first check as many invariants as possible to see if they differ on G_1 or G_2. However, if all the known invariants match, there is no guarantee that the two graphs are isomorphic until we find the bijection that preserves adjacencies.

14 The Cartesian product of two sets A and B, denoted by $A \times B$, is the set of ordered pairs (a, b), such that $a \in A$ and $b \in B$. We write $A \times B = \{(a, b) \mid a \in A,\ b \in B\}$. A relation R is a subset of $A \times B$. A relation R is reflexive if $(x, x) \in R$; symmetric if $(x, y) \in R \iff (y, x) \in R$; antisymmetric if $(x, y) \in R$ and $(y, x) \in R \implies x = y$; and transitive if $(x, y) \in R$ and $(y, z) \in R \implies (x, z) \in R$. An equivalence relation on a set S is a relation $\sim$ on $S \times S$ that is reflexive, symmetric, and transitive. For example, the equality relation on the set of integers $\mathbb{Z}$ is an equivalence relation.

15 A partition of a set S is a family of pairwise disjoint subsets of S whose union is S. An equivalence relation imposes a partition on S, where each subset consists of elements that are related to each other. Conversely, every partition of S gives rise to an equivalence relation $\sim$, where $x \sim y \Leftrightarrow x$ and y lie in the same sub set of the partition.

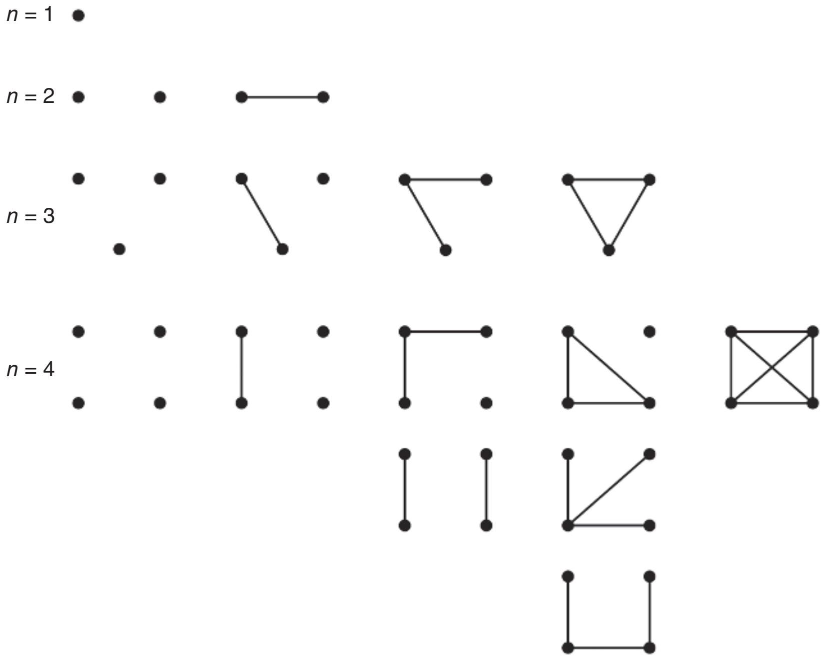


Figure 1.16 The non-isomorphic graphs on $n \leq 4$ vertices.

The "brute-force" way to do this would be to check $n!$ possible mappings in the worst case.

Next, let us study another definition of "sameness." A graph on n vertices with distinct labels attached to the vertices such as $v_1, \dots, v_n$ or simply $1, 2, \dots, n$ is called a *labeled graph*. Two labeled graphs G_1 and G_2 are *label isomorphic* or *identical* if there is a bijection from $V(G_1)$ to $V(G_2)$ that preserves labeled adjacencies. Otherwise they are *non-identical*. Figure 1.17 shows the non-identical graphs on 1, 2, and 3 vertices.

The next result determines precisely the number of non-identical graphs on n vertices.

Proposition 1.2.1. *The number of non-identical graphs on n vertices is* $2^{\binom{n}{2}}$.

Proof. The maximum number of edges in a graph with n vertices is $\binom{n}{2}$. Each edge may or may not be in the graph. So there are $2^{\binom{n}{2}}$ non-identical graphs. [16] □

16 There are some combinatorial ideas embedded in the argument worth elucidating, especially the well-known combinatorial identity $\sum_{k=0}^{t} \binom{t}{k} = 2^t$. To prove this identity, count the same thing

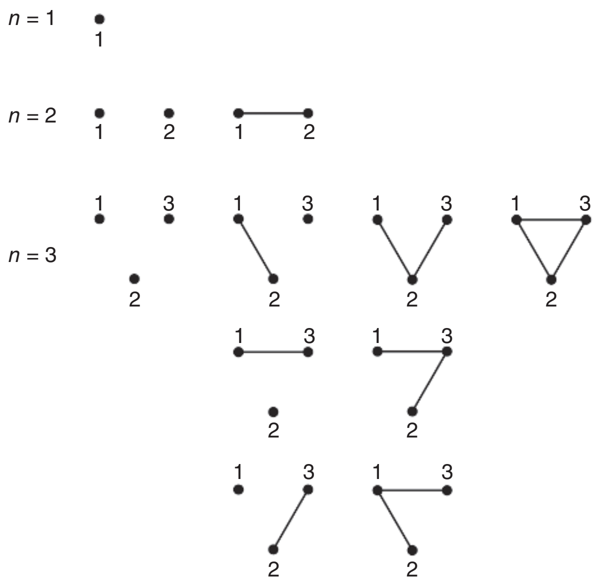


Figure 1.17 The non-identical graphs on $n \leq 3$ vertices.

Isomorphism is not quite relevant for large graphs in the sense that it is unlikely that two large graphs arising from real world situations would be isomorphic. One approach adopted by network scientists is to compare and contrast complex networks based on their rough shapes. Finding the approximate shape of a complex network is no small task. For example, the World Wide Web (the digraph formed with webpages as vertices, where two pages are joined by an arc if one page links to another) evidently has a distinct bow-tie shape according to Broder et al. (2000).[17] Bacterial metabolic networks (digraphs consisting of chemical reactions and their links to each other) also exhibit a bow-tie shape (Csete and Doyle, 2004). Figure 1.18 displays some typical shapes of large graphs.

---

in two different ways. Let S be a set of t elements. The expression on the left counts the number of k-element subsets of S, where k ranges from 0 to t. The number of subsets of S can be counted in a different way. Let X be a subset of S. Each element of S can be inside X or outside X, leading to two choices per element. So the number of ways of constructing X is 2^t. In the case of graphs, the number of non-identical graphs with n vertices and m edges, where $m \leq n$ is $\binom{\binom{n}{2}}{m}$. To obtain the total number of non-identical graphs on n vertices sum over all possible m and use the combinatorial identity to get $\sum_{m=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{m} = 2^{\binom{n}{2}}$.

17 Later research showed that the bow-tie shape of the Web graph was not entirely correct. While the Web graph does have a large central core, the bow-tie structure was the result of the Web crawling techniques used (Meusel et al., 2015).

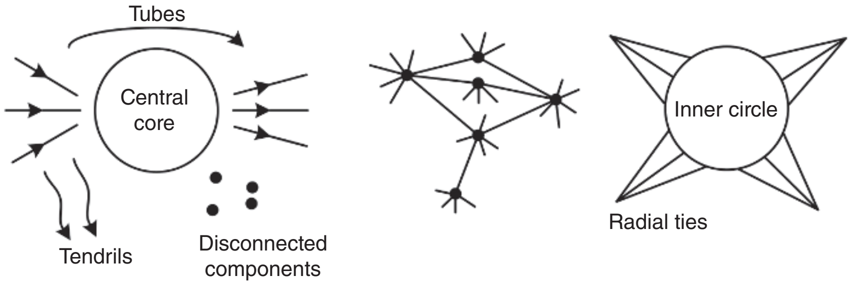


Figure 1.18 Shapes of graphs.

Easley and Kleinberg (2010) describe how the spread of email recommendations for a Japanese comic book and the spread of a tuberculosis outbreak have a similar appearance (see the second diagram in Figure 1.18). Social contagions, unlike biological contagions, involve decision-making by the individual. However, as Easley and Kleinberg note "the network-level dynamics are similar, and insights from the study of biological epidemics are also useful in thinking about the processes by which ideas spread on social networks."

In a *collaboration graph* people are represented as vertices and two people are joined by an edge if they collaborated on a project. In mathematics a famous collaboration graph revolves around Paul Erdős, who wrote hundreds of papers with collaborators. The 511 mathematicians who coauthored a paper with Erdős are assigned Erdős number 1, mathematicians who coauthored a paper with a coauthor of Erdős are assigned Erdős number 2, and so on. Erdős himself is assigned Erdős number 0. An edge is drawn between two mathematicians if they collaborated on a paper together (possibly they jointly coauthored a paper with Erdős). The Erdős collaboration graph is evolving, although the number of mathematicians with Erdős number 1 is fixed at 511 since Erdős died in 1996.

The shape of this collaboration graph looks like the third diagram in Figure 1.18. It graph has a densely connected core along with loosely coupled radial branches reaching out from the core. The close knit inner circle of collaborators share a common "mathematical tongue" that is critical to proving new results. The radial ties bring diversity in the form of new collaborators.[18]

The entire Erdős-1 collaboration graph is presented in Figure 1.19. It is drawn using Gephi 0.9.2 using the Force-Atlas layout algorithm. Vertices with higher degree are darker. Note that Erdős does not appear in the Erdős-1 collaboration

18 Jerold Grossman and his collaborators created a website on the Erdős Number Project located at http://www.oakland.edu/enp/. The Erdős -1 collaboration graph and its description were originally prepared by Valdis Kreb based on Grossman's data and is located at http://www.orgnet.com/Erdos.html.

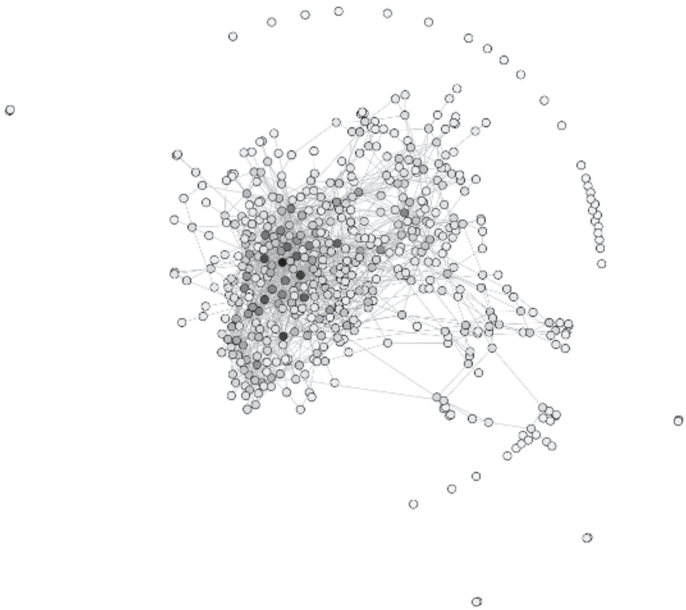


Figure 1.19 The Erdős-1 collaboration graph.

graph and coauthors of Erdős who did not publish a paper with any other coauthor of Erdős appear as isolated vertices in the graph.

We will end this section with the Reconstruction Conjectures. The subgraph obtained by deleting one vertex v is called a *vertex-deletion* and is denoted by $G - v$. The subgraph obtained by deleting one edge e is called an *edge-deletion* and is denoted by $G\backslash e$. When a vertex is deleted, all the edges incident with the vertex are removed. When an edge is deleted, only the link between the two end vertices is removed, while the end vertices remain. Any isolated vertices that may be formed remain intact. If this is not the case it will be specified.

Suppose G is a graph with n vertices labeled $v_1, \ldots, v_n$. Consider the labeled subgraphs of vertex-deletions: $G - v_1, G - v_2, \ldots, G - v_n$. This multiset of vertex-deletions is called the *deck* of vertex-deletions. If you are shown only the deck of vertex-deletions, can you reconstruct G? The Vertex Reconstruction Conjecture states that two graphs (with at least 3 vertices) that have the same deck of vertex-deletions are isomorphic (Kelly, 1957; Ulam, 1960). Figure 1.20 displays two non-isomorphic graphs on 3 vertices and their decks of vertex-deletions. Observe that the decks are different.

Now suppose we remove all vertex labels from the vertex-deletions and keep only non-isomorphic graphs. This set of unlabeled subgraphs is called the set of vertex-deletions. The vertex-deletions in Figure 1.20, although different when

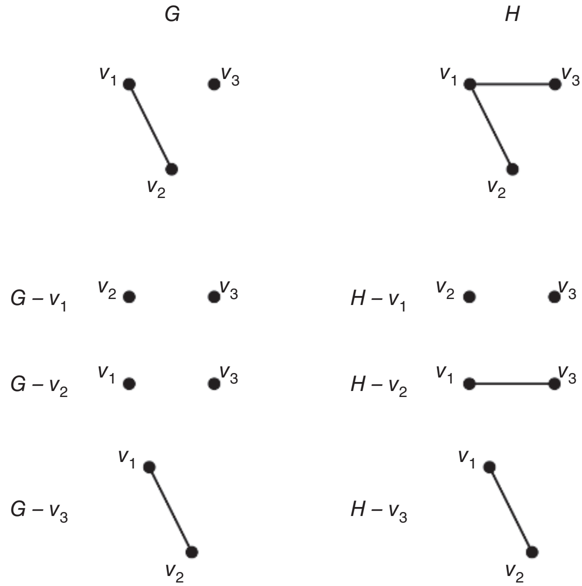


Figure 1.20 Two non-isomorphic graphs and their decks of vertex-deletions.

viewed as multisets, are the same when viewed as sets since $G - v_1 \cong G - v_2$ and $H - v_2 \cong H - v_3$. Can G be reconstructed from its set of vertex-deletions? This harder question, known as the *Set Reconstruction Conjecture*, was proposed by Harary (1964).

One approach to tackling the Vertex Reconstruction Conjecture is to see how much information about the graph can be recovered from its deck of vertex-deletions. Proposition 1.2.2 lists three graph invariants that can be reconstructed.

Proposition 1.2.2. *The number of vertices, edges, and the degree sequence of a graph can be reconstructed from its deck of vertex-deletions.*

Proof. Let G be a graph with n vertices $v_1, \dots, v_n$ and m edges. The number of vertices is clearly reconstructable since it is the number of graphs in the deck. To see that the number of edges of G is reconstructable, let m_i be the number of edges in $G - v_i$, where $1 \leq i \leq n$. Each edge appears in $n - 2$ of the vertex-deletions. Therefore the number of edges is

$$m = \frac{1}{n-2} \sum_{i=1}^{n} m_i.$$

To find $deg(v_i)$, observe that in $G - v_i$, the missing edges are precisely $deg(v_i)$. Therefore $deg(v_i) = m - m_i$. □

Another approach to tackling the Vertex Reconstruction Conjecture is to prove that it holds for special classes of graphs; for instance regular graphs. In fact, if G is a regular graph of degree t, it can be reconstructed from a single vertex-deletion $G - v$. To see this, observe that the vertices in $G - v$ are of two types: those with degree t and those with degree $t - 1$. Moreover, there will be t vertices of degree $t - 1$. To obtain G add a new vertex adjacent to the t vertices of degree $t - 1$.

The Edge Reconstruction Conjecture is analogous to the Vertex Reconstruction Conjecture. Suppose G is a graph with m edges labeled $e_1, \ldots, e_m$. Consider the labeled subgraphs of edge-deletions: $G\backslash e_1, G\backslash e_2, \ldots, G\backslash e_m$. This multiset of edge-deletions is called the deck of edge-deletions. If you are shown only this deck of edge-deletions, can you reconstruct G? The Edge Reconstruction Conjecture states that graphs (with at least 6 edges) that have the same decks of edge-deletions are isomorphic (Harary, 1964). A corresponding conjecture also exists for sets of edge-deletions.

1.3 Constructions and Minors

We begin this section by describing three ways of combining two graphs G_1 and G_2 to obtain a new graph. The *join* of G_1 and G_2, denoted by $G_1 + G_2$, has vertex set $V(G_1) \cup V(G_2)$ and every vertex in $V(G_1)$ is joined to every vertex in $V(G_2)$. To construct the join draw the two graphs next to each other and connect every vertex in one to every vertex in the other. For example, the wheel graph W_{n-1}, shown in Figure 1.7, may be viewed as the join of the single vertex and the cycle C_{n-1} with $n - 1$ vertices. Figure 1.21 gives another example of a join.

The *Cartesian product* of two graphs G_1 and G_2, denoted by $G_1 \times G_2$, has vertex set

$$V(G_1) \times V(G_2) = \{(v_i, v_j) \mid v_i \in V(G_1),\ v_j \in V(G_2)\}.$$

Two vertices (v_i, v_j) and (v_k, v_l) are adjacent if $v_i v_k \in E(G_1)$ and $j = l$ or $v_j v_l \in E(G_2)$ and $i = k$. Figure 1.22 displays the Cartesian product of the path graphs P_m and P_n and the Cartesian product of P_3 and the cycle C_3.

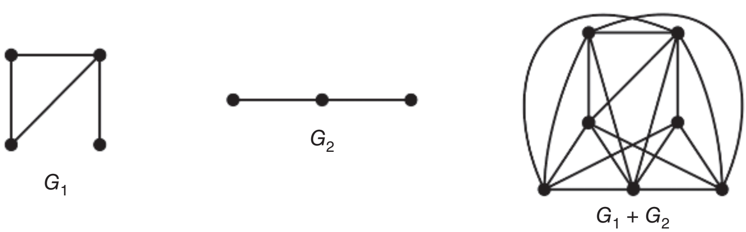


Figure 1.21 An example of a join.

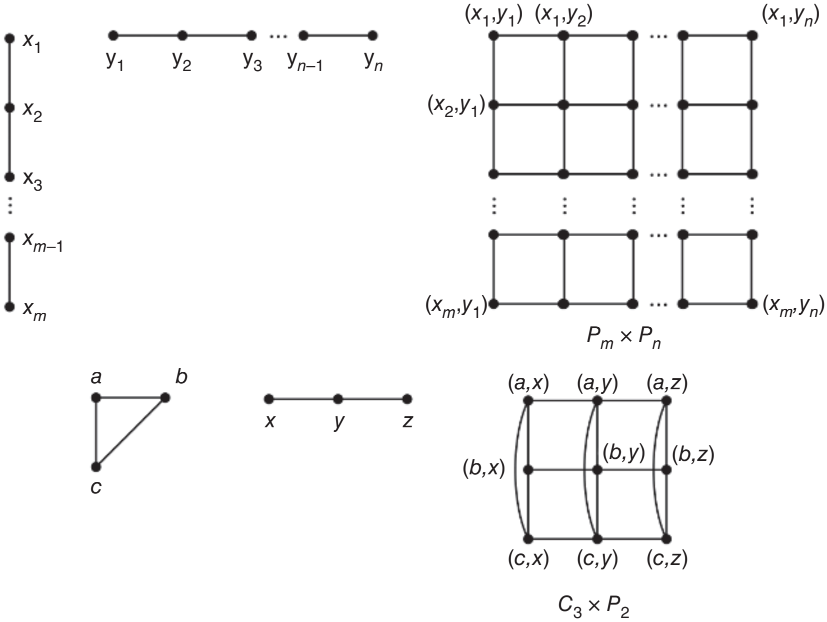


Figure 1.22 Two examples of Cartesian products.

Using the Cartesian product we can define the n-dimensional "cubes," Q_n, recursively as: $Q_1 = P_2$; and for $n \geq 2$, $Q_n = Q_{n-1} \times P_2$. Figure 1.23 gives a nice visualization of the four-dimension analog of the cube, called a tesseract. Start with P_2 whose vertices are labeled a and b. To obtain $Q_2 = Q_1 \times P_2$ draw two copies of P_2

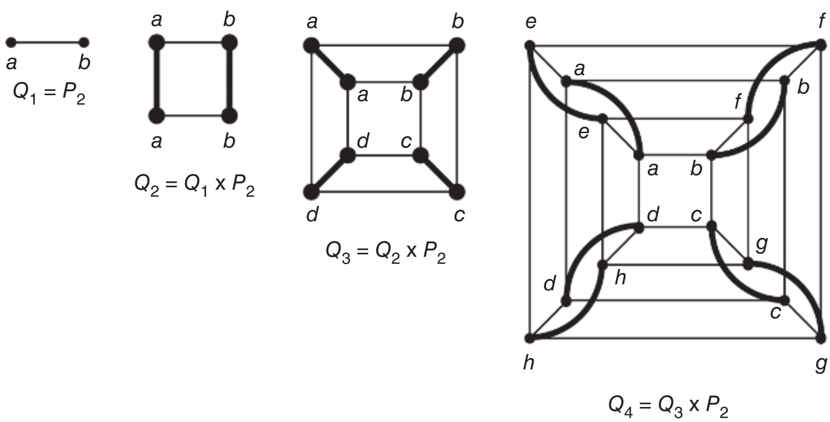


Figure 1.23 The four-dimensional cube.

next to each other and join corresponding vertices. The edges joining corresponding vertices are shown in bold. To obtain $Q_3 = Q_2 \times P_2$ draw two copies of Q_2 and join corresponding vertices. Similarly, to obtain $Q_4 = Q_3 \times P_2$ draw two copies of Q_3 and join corresponding vertices. We can assign coordinates to each of the vertices so that each adjacent vertex differs by exactly one coordinate. The vertices of Q_2 may be labeled $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$. Similarly, the vertices of Q_3 may be assigned 3-tuples, and the vertices of Q_4 may be assigned 4-tuples. Viewed in this manner Q_4 is the four-dimensional cube.

Suppose each of G_1 and G_2 has a clique K_t, where $t \geq 1$. The *clique sum* of G_1 and G_2 across K_t is formed by placing two copies of K_t, one on top of the other, and identifying the vertices and edges. For example, the clique sum of G_1 and G_2 across K_1 (a vertex) is obtained by attaching G_1 and G_2 at the specified vertex. The clique sum across K_2 (an edge) is obtained by attaching G_1 and G_2 along the specified edge. The clique sum across K_3 (a triangle) is obtained by attaching G_1 and G_2 along the specified triangle, and so on. For $t \geq 2$, if in addition the identified edges of K_t are removed, then the clique sum is called a t-sum.[19] Figure 1.24 gives examples of clique sums across K_2 and K_3 with all the edges removed.

The *complement* of a graph G, denoted by $\overline{G}$, is the graph that has vertex set $V(G)$ and a pair of vertices is joined by an edge in $\overline{G}$ if there is no edge joining them in G. Figure 1.25 gives three examples of graphs and their complements. Some graphs like P_4 and C_5 are isomorphic to their complements. Such graphs are called *self-complementary*. The next result gives a necessary condition for a graph to be self-complementary.

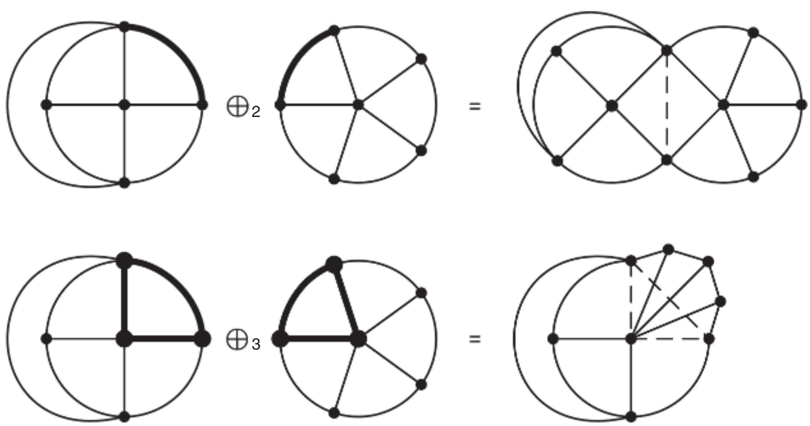


Figure 1.24 $(K_5 \backslash e) \oplus_2 W_5$ and $(K_5 \backslash e) \oplus_3 W_5$.

19 The terms “clique sum” and “t-sum” depend on how authors define them. Some authors allow a subset of the identified edges to stay.

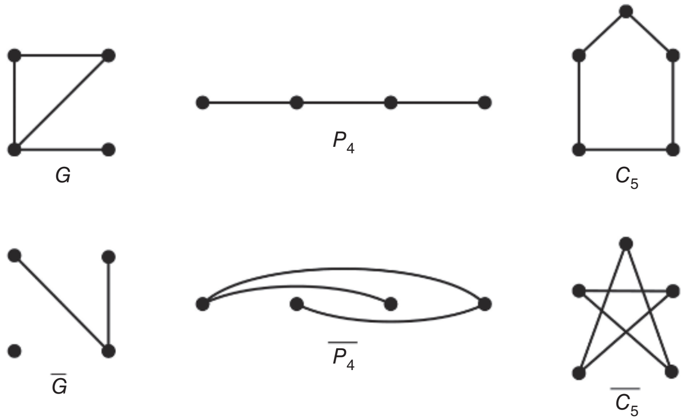


Figure 1.25 Examples of complements.

Proposition 1.3.1. *If G is a self-complementary graph on n vertices, then $n = 4k$ or $n = 4k + 1$, for some integer $k \geq 1$.*

Proof. Suppose G has m edges and let $\overline{G}$ be its complement. Observe that

$$|E(G)| + |E(\overline{G})| = \frac{n(n-1)}{2}.$$

Since G is self-complementary, $|E(G)| = |E(\overline{G})| = m$. So $2m = \dfrac{n(n-1)}{2}$ and $m = \dfrac{n(n-1)}{4}$. Since m is a natural number, n or $n-1$ must be a multiple of 4. Therefore $n = 4k$ or $n = 4k + 1$, for some integer $k \geq 1$. □

A *subdivision* of a multigraph G is obtained from G by replacing edges with paths of length at least 2 and loops by cycles. Figure 1.26 shows a multigraph G and a graph H and their subdivisions.

A *series-parallel network* is a multigraph obtained from a loop or a single edge by subdividing edges or adding edges in parallel. Subdivisions and series-parallel networks are related, but not exactly the same. A series-parallel network begins with a loop. Edges are added in parallel to existing edges and vertices are placed on edges. A subdivision may be constructed from any graph H, by placing vertices on edges. For example, the subdivision of the multigraph G in Figure 1.26 is a series-parallel network. However, the subdivision of graph H is not a series-parallel network.

Let $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ be a family of sets. The *intersection graph* of $\mathcal{S}$ is the graph formed by treating the n sets in $\mathcal{S}$ as vertices and joining two vertices by

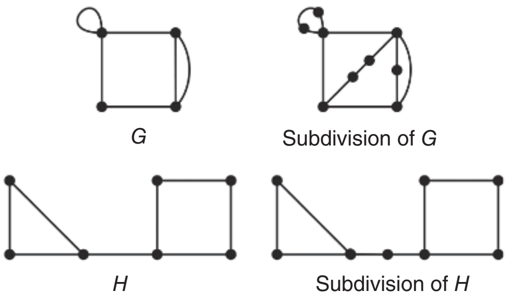


Figure 1.26 Examples of a subdivision.

an edge if their corresponding sets have a non-empty intersection. If the sets are intervals of real numbers, then the intersection graph is called an *interval graph*.

The *line graph* of a graph G, denoted by $L(G)$, is obtained from G by placing a vertex on each edge and joining vertices if the corresponding edges are adjacent. Figure 1.27 gives an example of a graph and its line graph. The presence of isolated vertices in G has no impact on $L(G)$. Moreover, the line graph of a disconnected graph consists of the line graphs of each connected component. Thus we may assume the graph is connected. The next result gives the number of vertices and edges in a line graph.

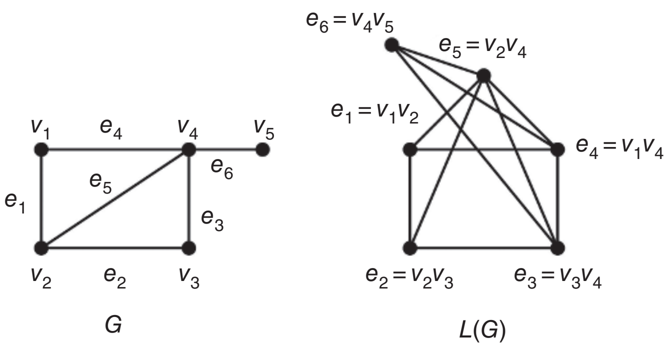


Figure 1.27 Line graphs.

Proposition 1.3.2. *Let G be a connected graph with $n \geq 2$ vertices and $m \geq 1$ edges. The number of vertices in $L(G)$ is m and the number of edges is $-m + \frac{1}{2}\sum_{v \in V} [deg(v)]^2$.*

Proof. By definition the number of vertices in $L(G)$ is m. To obtain the number of edges, observe that each of the $\binom{deg(v)}{2}$ pairs of edges adjacent to v gives rise to an edge in $L(G)$. Therefore the number of edges in $L(G)$ is $\sum_{v \in V} \binom{deg(v)}{2}$. Simplifying this expression gives

$$\sum_{v\in V}\binom{deg(v)}{2} = \frac{1}{2}\sum_{v\in V} deg(v)[deg(v)-1] = \left[\frac{1}{2}\sum_{v\in V}[deg(v)]^2\right] - m. \qquad \square$$

A graph G is a *line graph* if $G \cong L(H)$ for some graph H. Line graphs have several interesting properties. For example, each vertex in G is in at most two cliques. To see this, observe that the edges incident to a vertex in H form a complete subgraph in G. Since each edge of H has two end vertices, it is in the neighborhood of exactly two vertices. Therefore each vertex in G is in at most two cliques. Not all graphs are line graphs. For example, the star graph $K_{1,3}$ is not a line graph, since the vertex that is joined to the other three vertices is in three cliques.

A class of graphs $\mathcal{G}$ is said to be *closed under induced subgraphs* if every induced subgraph of a graph in $\mathcal{G}$ is also in $\mathcal{G}$. For example, an induced subgraph of a line graph is also a line graph. A *forbidden induced subgraph* is a graph that is not in $\mathcal{G}$, but all of its proper induced subgraphs are in $\mathcal{G}$. *A priori* there is no reason to believe the list of forbidden induced subgraphs for line graphs is finite. However, Lowell Beineke proved precisely such a result (Beineke, 1970), which we will state without proof.

Theorem 1.3.3. *Let G be a connected graph. Then $G = L(H)$ for some graph H if and only if G has no induced subgraph isomorphic to one of the nine graphs in Figure 1.28.*

Minors are substructures that are a bit more complicated than subgraphs. First we must understand the contraction operation. To *contract* an edge e with end vertices u and v, collapse the edge and identify the end vertices u and v as one vertex. The resulting graph, denoted by G/e, is called an *edge-contraction*. Note that parallel edges and loops may be formed in the contraction process. A graph H is a *minor* of a graph G, if H can be obtained from G by deleting edges or contracting edges. We write $H = G\backslash X/Y$, where X is the set of edges deleted and Y is the set of edges contracted. The focus is entirely on edges. Figure 1.29 shows examples of deletions and contractions.

An edge with an end-vertex of degree 1 is called a pendant edge. If e is a loop or a pendant edge, then $G\backslash e = G/e$. If H is obtained from G only by deleting edges and any resulting isolated vertices, then H is called a *deletion-minor* of G. Similarly, if H is obtained from G only by contracting edges, then H is called a *contraction-minor* of G.

A minor is not necessarily a subgraph. Consider the Petersen graph P shown in Figure 1.30 with 10 vertices and 15 edges.[20] It does not contain K_5 as a subgraph

20 The Petersen graph is a useful example to test any conjectures that come to mind. Julius Petersen attributes it to J.J. Sylvester Petersen (1898). It appears in Kempe (1886) for the first time according to Biggs et al. (1976).

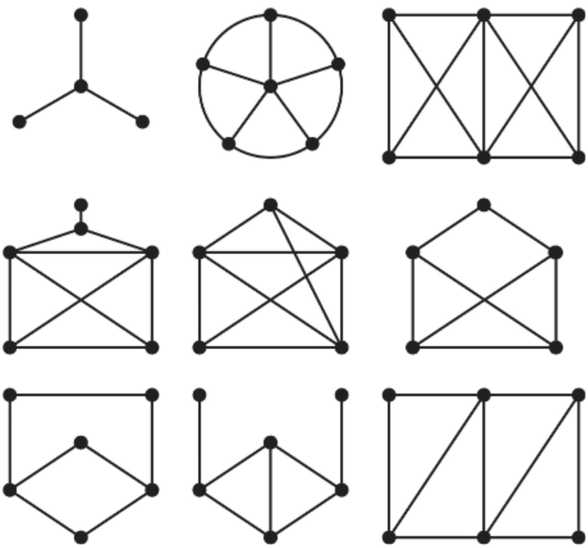


Figure 1.28 Forbidden induced subgraphs for line graphs.

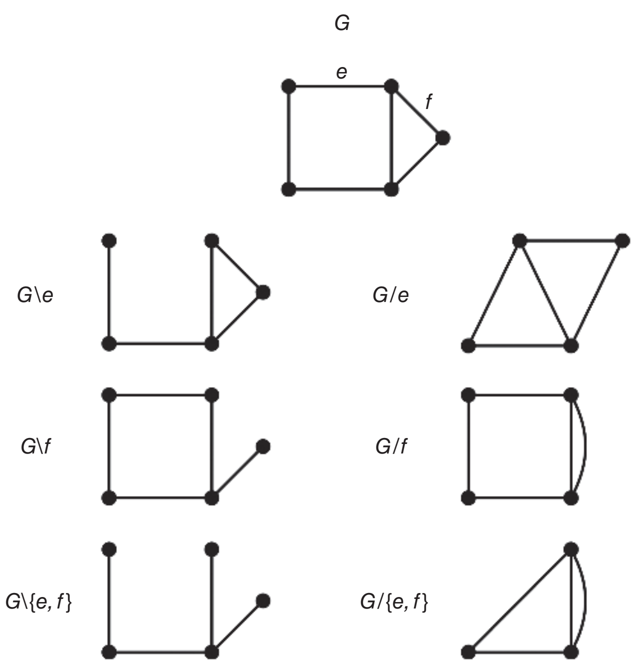


Figure 1.29 Example of edge-deletions and edge-contractions.

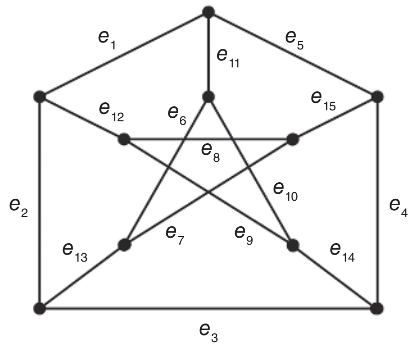


Figure 1.30 Petersen graph.

since it has no vertex of degree 4. However, it does contain K_5 as a minor. To see this, contract the five edges e_{11}, e_{12}, e_{13}, e_{14}, and e_{15} that join the outer pentagon to the inner star; the resulting minor is K_5. We write

$$P/\{e_{11}, e_{12}, e_{13}, e_{14}, e_{15}\} \cong K_5.$$

A subgraph is not necessarily a minor. If the subgraph is formed only by deleting edges then the isolated vertices stay, and the subgraph is not a minor. This is the only situation where a subgraph is not a minor.

Let $\mathcal{M}$ be a class of graphs closed under minors. That is, if $G \in \mathcal{M}$, then $G \backslash e \in \mathcal{M}$ and $G/e \in \mathcal{M}$ for every edge e of G. Recall from Section 1.1 that a minimal excluded minor for $\mathcal{M}$ is a graph that is not in $\mathcal{M}$, but all of its proper minors are in $\mathcal{M}$. For example, the class of forests is closed under minors, since a minor of a forest is also a forest. The class of trees, however, is not closed under minors since removal of an edge in a tree results in a disconnected graph. The next result is an example of a straightforward excluded minor result.

Proposition 1.3.4. *A graph G is a forest if and only if G has no minor isomorphic to K_3.*

Proof. The smallest graph that is not a forest is K_3. Since the class of forests is closed under minors, if G is a forest, then G has no K_3-minor. Conversely, suppose G has no K_3-minor. Suppose, if possible, G is not a forest. Then G has a cycle C. A minor isomorphic to K_3 may be obtained by deleting all the edges outside C (and any isolated vertices formed as a result) and contracting all except three edges of C. This is a contradiction to the hypothesis. Therefore G is a forest. □

The class of series-parallel networks is closed under minors and has an excluded minor characterization: A connected graph is a series parallel network if and only if it has no minor isomorphic to K_4. It is easy to see that K_4 is a minimal excluded minor for series-parallel networks by drawing all the connected series-parallel networks up to 4 vertices. The other direction requires additional results and appears in Section 6.5.

The set of graphs $\mathcal{G}$ with the relation "is a minor of" is a partially ordered set.[21] To see this, observe that the minor relation is:

- Reflexive since a graph is a minor of itself.
- Antisymmetric since if G_1 is a minor of G_2 and G_2 is a minor of G_1, then $G_1 \cong G_2$.
- Transitive since if G_1 is a minor of G_2 and G_2 is a minor of G_3, then G_1 is a minor of G_3.

A pair of elements a, b is *incomparable* if neither $a \prec b$ nor $b \prec a$. A sequence of pairwise incomparable elements is called an *antichain*. The Graph Minor Theorem established that the set of graphs $\mathcal{G}$ with the minor relation has no infinite antichain. Thus for every infinite set of graphs, one of its members is isomorphic to a minor of another. This is the main result in "Graph Minors XX" (Robertson and Seymour, 2004). Diestel (2017) describes this theorem as "one which dwarfs any other result in graph theory and may doubtless be counted among the deepest theorems that mathematics has to offer."

Theorem 1.3.5. **(*Graph Minor Theorem*)** *For every infinite set of graphs, one of its members is isomorphic to a minor of another.*

To prove Theorem 1.3.5 the authors observed that if $\{G_1, G_2, \dots\}$ is an infinite antichain, then none of $G_2, G_3, \dots$ have a minor isomorphic to G_1. Therefore it sufficies to prove that the class of graphs with no minor isomorphic to a specific graph H has no infinite antichain. To prove this statement, they developed a structure theorem for the class of graphs with no H-minor, first when H is planar (Robertson and Seymour, 1986), and subsequently when H is non-planar (Robertson and Seymour, 2003).

In Section 1.1 we described the Graph Minor Theorem as stating that "every infinite class of graphs closed under minors has a finite list of minimal excluded minors." This statement is equivalent to the statement in Theorem 1.3.5. To see this, observe that an infinite set of minimal excluded minors for a

21 A relation $\prec$ on a set S is a partial order if it is reflexive, antisymmetric, and transitive. The set S is called a partially ordered set or poset. Reflexive, antisymmetric, and transitive operations were defined in Section 1.2. For example, the set of natural numbers with the divisibility relation and the set of all subsets of a set with the usual subset relation are partially ordered sets.

minor-closed class would be an infinite antichain. Theorem 1.3.5 implies that every minor-closed class of graphs has a finite list of minimal excluded minors.

The converse is also fairly straightforward. Suppose every minor-closed class of graphs has a finite list of minimal excluded minors. Let $\mathcal{H}$ be any set of graphs and consider the class of graphs with no minor isomorphic to any graph in $\mathcal{H}$. Since this class is closed under minors, it has a finite list of minimal excluded minors $G_1, \ldots, G_t$. By definition $G_1, \ldots, G_t$ are in $\mathcal{H}$, but no proper minor of $G_1, \ldots, G_t$ is in $\mathcal{H}$. This means that every graph in $\mathcal{H}$ has a minor isomorphic to at least one of $G_1, \ldots, G_t$. Therefore $\mathcal{H}$ has at least two graphs such that one is a minor of the other.

A *quasi order* is a relation that is reflexive and transitive. It is not necessarily antisymmetric. A quasi order is a *well-quasi ordering* if every sequence of elements $x_1, x_2, \ldots$ contains an increasing pair $x_i \prec x_j$, where $i < j$. So a well-quasi ordering has no infinite antichain nor an infinite strictly decreasing sequence. Clearly the class of graphs $\mathcal{G}$ with the minor relation has no infinite strictly decreasing sequence because deleting and contracting edges in a graph eventually results in the empty graph. The hard part is to show that $\mathcal{G}$ has no infinite antichain, which is Theorem 1.3.5. The Graph Minor Theorem is often presented as the following statement: The class of graphs $\mathcal{G}$ under the minor relation is well-quasi ordered.

Exercises

1.1 Suppose a graph has 15 edges, 3 vertices of degree 4, and all other vertices have degree 3. How many vertices does the graph have?

1.2 Is it possible to have a group of nine people where each person is acquainted with exactly five of the other people? Either draw a graph that demonstrates this or explain why it is not possible.

1.3 Prove that a graph on $n \geq 2$ vertices has at least two vertices of the same degree.

1.4 Figure 1.31 displays six graphs and three digraphs.
(a) For each graph find the degree sequence, mean degree, and density.
(b) For each digraph find the indegree sequence, outdegree sequence, mean indegree, mean outdegree, and density.

1.5 Draw all the non-isomorphic connected graphs on 5 vertices.[22]

22 The On-Line Encyclopedia of Integer Sequences (http://oeis.org) has the number of non-isomorphic connected graphs. Check to see the largest value of n for which these numbers are known.

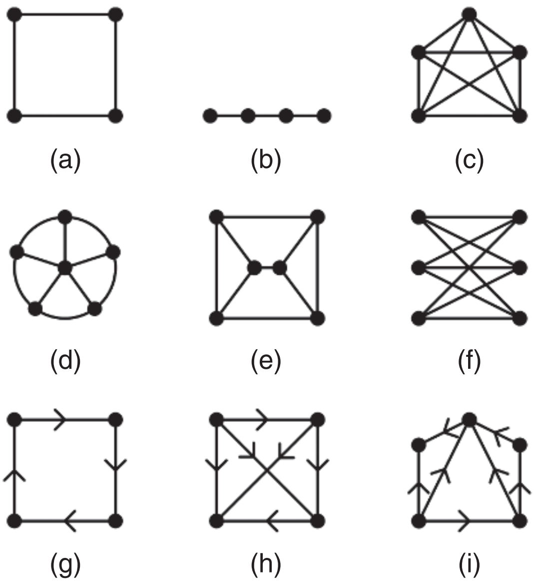


Figure 1.31 Examples of graphs and digraphs.

1.6 Draw all the non-identical graphs on 4 vertices and 3 edges.

1.7 How many labeled isomorphic graphs are there on 10 vertices? How many of them have 15 edges?

1.8 Are the pairs of graphs in Figure 1.32 isomorphic? If yes, give the isomorphism. If no, find an invariant that differs for them.

1.9 Give an example of two non-isomorphic graphs with the same degree sequence.

1.10 Show that the Petersen graph has a $K_{3,3}$-minor.

1.11 For the graphs in Figure 1.31:
(a) Find the decks of vertex-deletions and edge-deletions.
(b) Find minors isomorphic to W_3 and W_4, if possible.
(c) Find an induced subgraph isomorphic to $K_{1,3}$, if possible.
(d) Find their complements.
(e) Find their line graphs.

1.12 Illustrate $P_3 + W_4$, $P_3 \times W_4$, and $W_4 \oplus_3 W_4$.

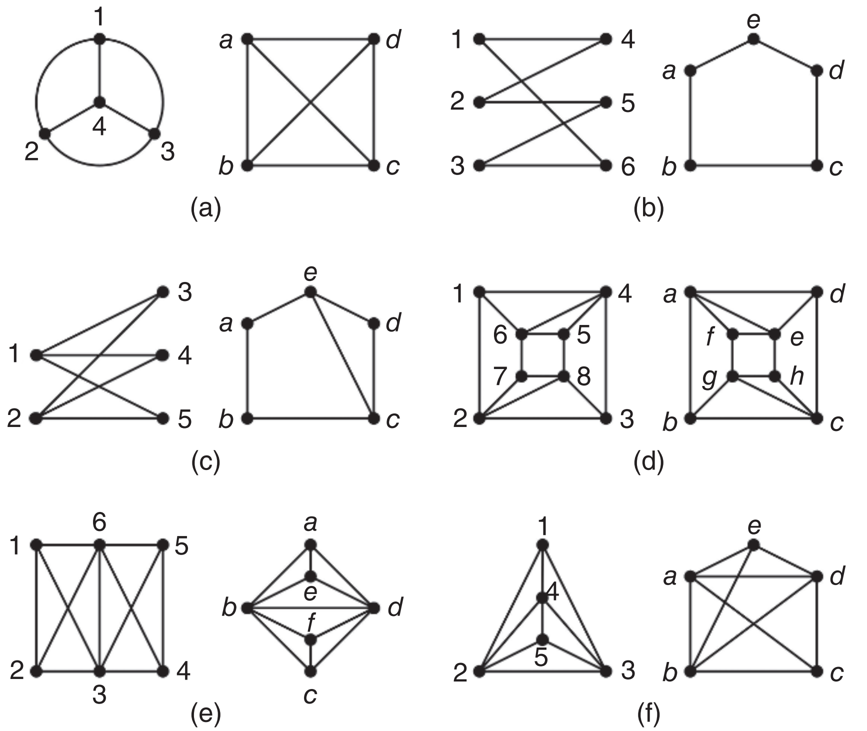


Figure 1.32 Pairs of graphs for isomorphism testing.

1.13 The *tensor product* of two graphs G and H, denoted by $G \otimes H$, has vertex set $V(G) \times V(H)$ and pairs (g, h) and (g, h') are adjacent if and only if gg' is an edge in G and hh' is an edge in H. Illustrate $P_4 \otimes P_4$, $P_2 \otimes C_3$, and $P_2 \otimes W_4$.

1.14 Show that each of the nine graphs in Figure 1.28 is not a line graph.

1.15 Figure 1.33 contains the nine graphs shown in Figure 1.28 as induced subgraphs. Find each of them.

Topics for Deeper Study

1.16 Give an example to show that digraphs are not reconstructable from their decks of vertex-deletions (Stockmeyer, 1977).

1.17 Line graphs have many equivalent characterizations and interesting properties. Let G be a connected graph.

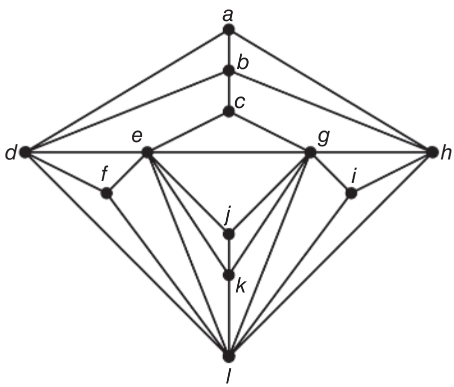


Figure 1.33 A graph that contains all nine forbidden induced subgraphs for line graphs.

(a) Prove that G is a line graph if and only if the edges of G can be partitioned into cliques so that every vertex is in at most two cliques (Krausz, 1943).
(b) What sort of graphs are isomorphic to their line graphs? For example, $L(C_n) \cong C_n$. As it turns out, cycles are the only graphs with this property. Let G be a connected graph with n vertices. Prove that $G \cong L(G)$ if and only if $G \cong C_n$.
(c) It is easy to check that $L(C_3) \cong L(K_{1,3})$ even though $C_3 \ncong K_{1,3}$. Are there other such graphs? This was a question that Hassler Whitney answered in his paper introducing line graphs (Whitney, 1932a). Let G_1 and G_2 be connected graphs neither of which is C_3 or $K_{1,3}$. Prove that $G_1 \cong G_2$ if and only if $L(G_1) \cong L(G_2)$.
(d) Prove Theorem 1.3.3: A graph is a line graph if and only if it does not have any of the nine graphs in Figure 1.28 as induced subgraphs (Beineke, 1970).
(e) Suppose G is a graph with $\delta(G) \geq 5$. Prove that G is a line graph if and only if G has no induced subgraph isomorphic to the first six graphs in Figure 1.28 (Metelsky and Tyshkevich, 1997).

1.18 The set of automorphisms of a graph together with the composition operation form a group called the *automorphism group* of the graph. Explore the connection between graphs and groups (Chartrand et al., 2011).

1.19 Isomorphism and label isomorphism are not the only ways to determine if two graphs are similar. Ronald Graham notes that "Among a variety of fundamental themes running through Stan Ulam's mathematical research, one that particularly intrigued him was that of similarity (Graham, 1987). He was constantly fascinated by the problem of quantifying exactly how alike (or different) two mathematical objects or structures were." Ulam's idea for measuring the difference between k objects was to break them into as few as possible equal pieces. Let $\mathcal{G}_{n,m}$ be the set of graphs with m edges and at most n vertices An *Ulam decomposition* of k graphs $G_1, \ldots, G_k \in \mathcal{G}_{n,m}$ is k sets of partitions

$$(E_{11}, \ldots, E_{1t}), (E_{21}, \ldots, E_{2t}), \ldots, (E_{k1}, \ldots, E_{kt})$$

where, as graphs, $E_{1j} \cong E_{2j} \cong \ldots E_{kj}$ for $1 \leq j \leq t$. Let $U(G_1, \ldots, G_k)$ be the minimum value of t for which an Ulam decomposition of $G_1, \ldots, G_k$ exists. Let $U_k(n)$ be the maximum value of $U(G_1, \ldots, G_k)$, where the maximum is taken over all sets of k graphs from $\mathcal{G}_{n,m}$. This measure $U_k(n)$ is a measure of maximum dissimilarity for a set of k graphs in $\mathcal{G}_{n,m}$. The first result in the following text appears in Chung et al. (1979) and the second and third in Chung et al. (1981)

(a) Show that $U_2(n) \leq \frac{2}{3}n + c_2$, where c_2 is a constant.
(b) Show that $U_3(n) \leq \frac{3}{4}n + c_3$, where c_3 is a constant.
(c) Show that $U_k(n) \leq \frac{3}{4}n + c_k$, where c_k is a constant.
(The constant factor of $\frac{3}{4}$ did not increase for $k \geq 3$ suggesting that $\mathcal{G}_{n,m}$ is 3-dimensional in the sense that once there are three graphs that are maximally dissimilar, adding further graphs does not change anything.)