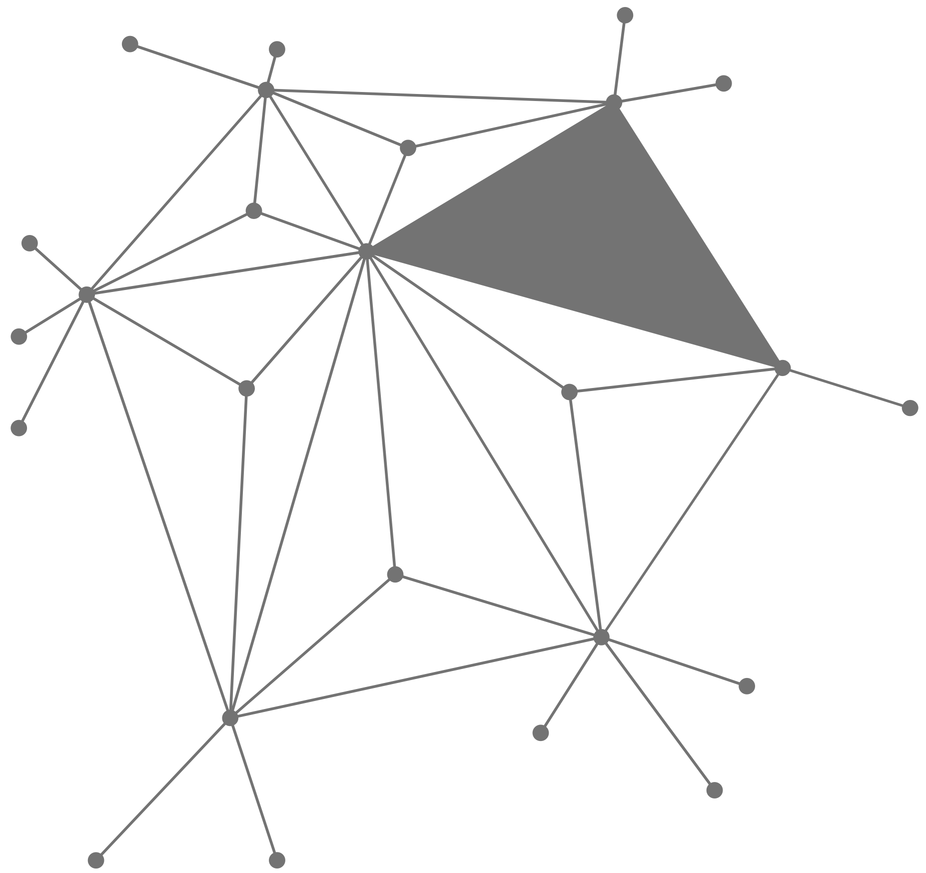
A Course on the Web Graph Chapter 1

Introduction to

GRAPH THEORY

By:
Achyuth Warrier
Ashik Stenny

Basic Definitions



Graph:

A graph G is defined as an ordered pair G=(V,E) where:

- V - the vertex set
- E - the edge set

Vertex Set(V):

Consists of finite or infinite number of elements where each element denotes a vertex (point or node) in the graph

Edge Set(E):

Set of unordered or ordered pairs of vertices representing edges or connections between them.

Example: A **Web Graph** with Web pages as vertices and edges existing between pages with direct links to each other

Types of Graphs

Finite

Finite graphs have a finite number of vertices and edges in the vertex set and edge set respectively

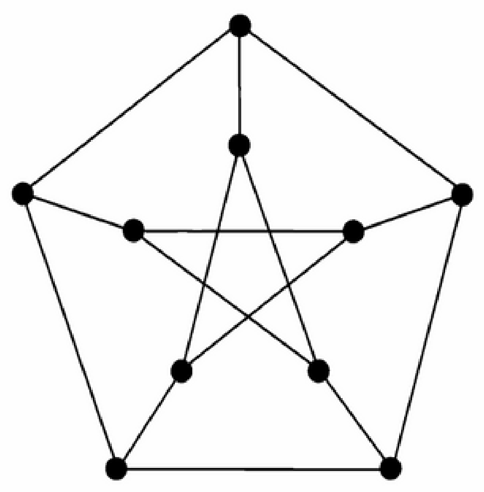
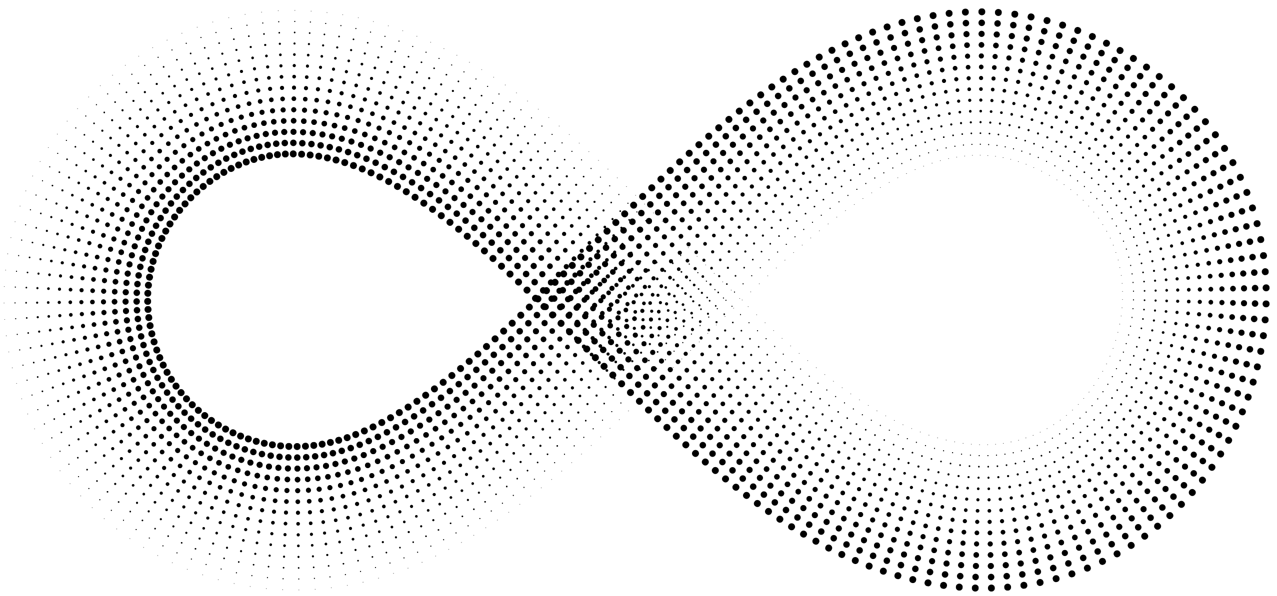


Figure 1.1. The Petersen graph.

$$|E(G)| \leq \binom{|V(G)|}{2}$$

Infinite

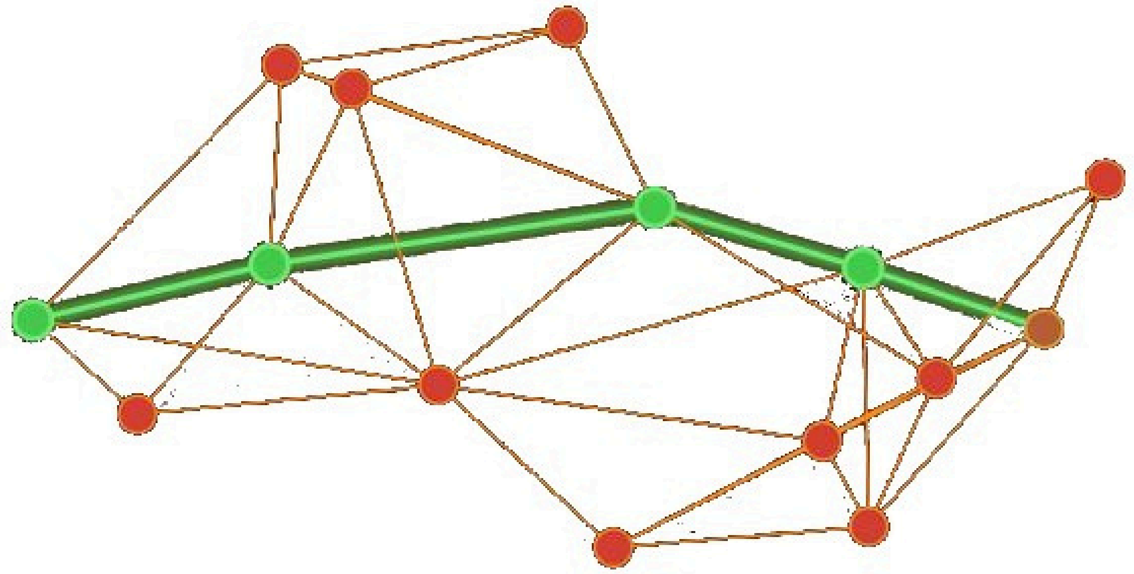
Infinite graphs have an infinite number of vertices and/or edges in the vertex set and edge set respectively



$$|E(G)| \leq |V(G)|$$

Subgraph:

A graph H is said to be a subgraph of G if
the vertex set of H is a subset of G
the edge set of H is a subset of G



Induced subgraph:

For a set S, subset of V, we define the subgraph induced by S on G(V,E) as the graph with vertex set S and an edge between any two vertices in S only if they are connected in G

Homomorphism:

A homomorphism f from vertex set of G to vertex set of H , is a function which preserves the edges, that is if there exists an edge between x and y where x and y are vertices in G then there exists an edge between f(x) and f(y) in H.

This means that every edge in G must map to some edge in H but the reverse need not be true.

Example, application of homomorphism: A k color graph coloring of a Graph G is a homomorphism to a k clique.

Graph Coloring from Homomorphism?

We will split the vertices into sets and a vertex is an element of the set only if it does not have any edges with any other vertex in the set and the neighbour set of this vertex and all other vertices in the set is a subset of single superset

After this we replace every set with a single vertex and we connect two vertices with an edge if they are subsets of the corresponding neighbour supersets of the each other

This corresponding graph will be a homomorphism.
The highest degree in this graph would be good approximate of the minimum number of colors required.

X(G) < n if and only if there is some homomorphism from G to Kn.

Embedding

An embedding is an injective homomorphism

Definition of an Embedding Function f:
Let G = (V_G, E_G) and H = (V_H, E_H) be two graphs.
A function f: V_G to V_H is an embedding of G into H if:

1. Injectivity (One-to-One Mapping)
   - f(u) != f(v) for all u != v in V_G.
   - Ensures no vertex collapses in the mapping.
2. Edge Preservation
   - If (u, v) in E_G, then (f(u), f(v)) in E_H.
   - Ensures the adjacency structure of G is maintained in H.

In short: An embedding function preserves both vertices and edges uniquely.

Isomorphism

An isomorphism is a bijective embedding

If there exists an isomorphism between two graphs then the two graphs satisfy all the same properties.

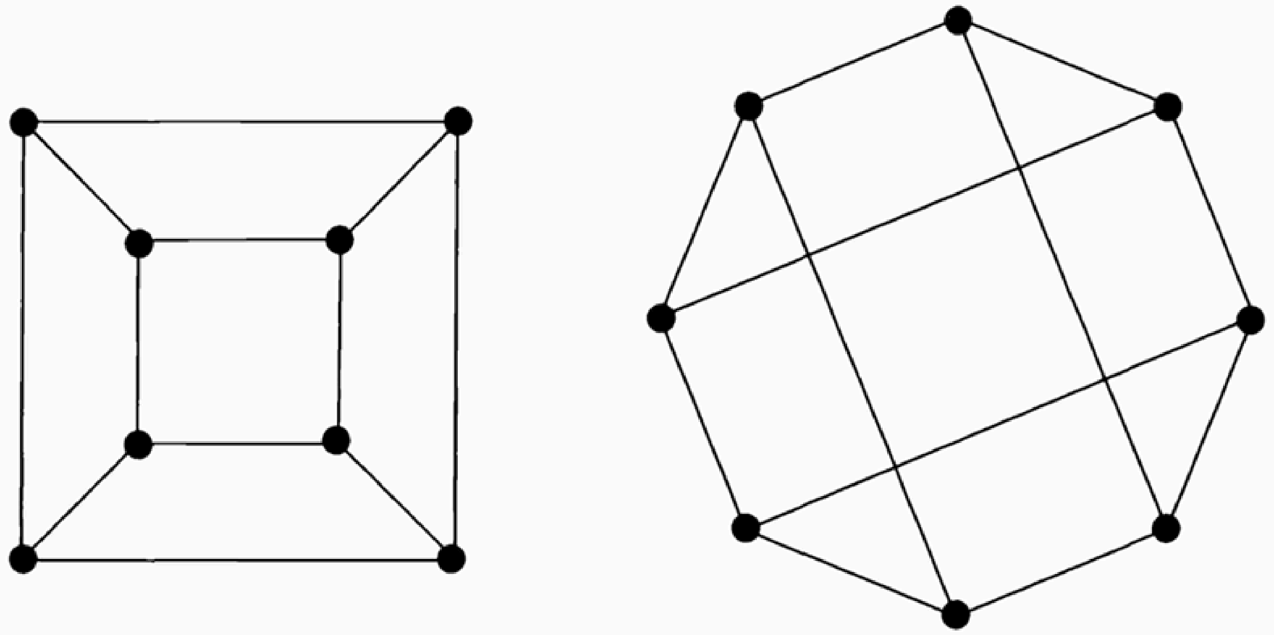


Figure 1.2. Isomorphic graphs.

Automorphism

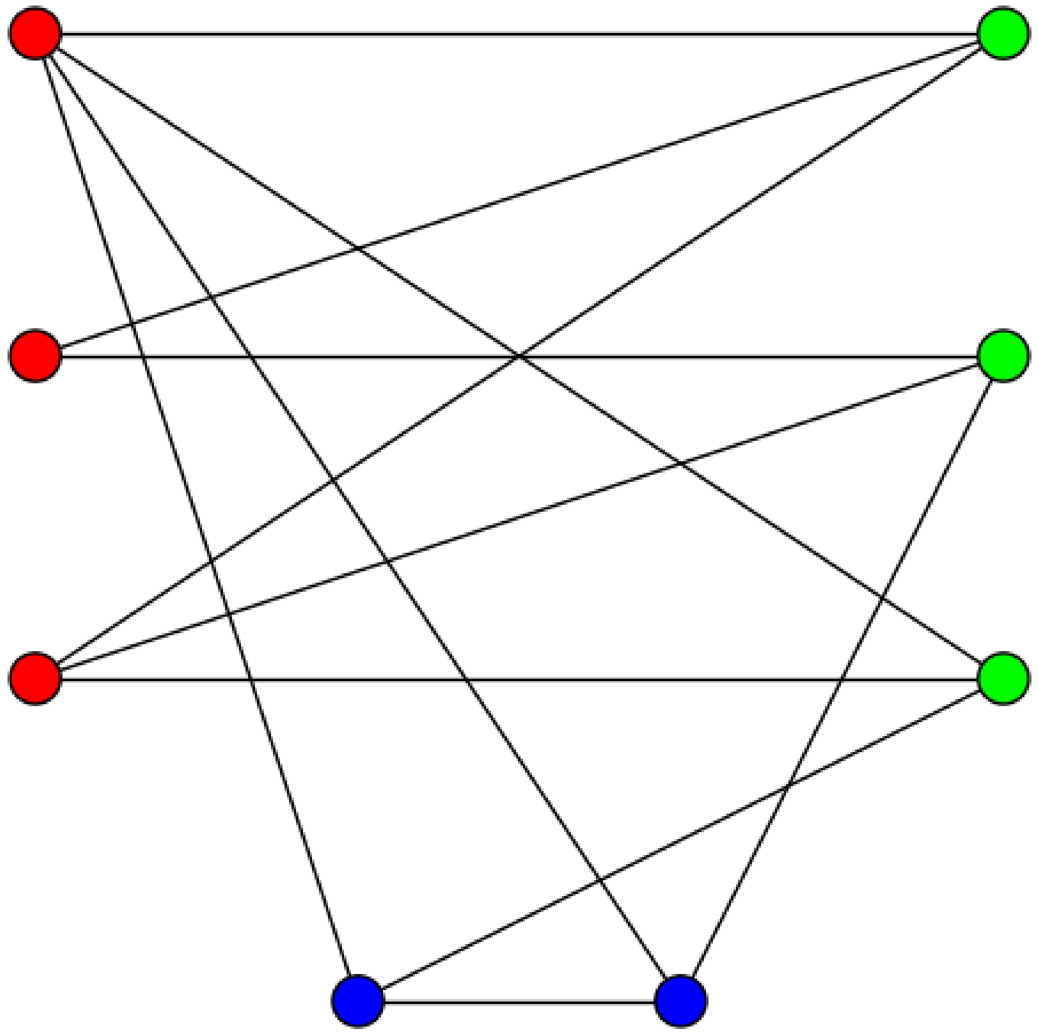
An automorphism of a graph G is an isomorphism from G to itselff

Walk:

A walk in a graph consists of an alternating sequence of vertices and edges x0,ei,xi,...,et,xt so that for all 1 < i < t, ei = xi - i xi .

Degree:

The degree of a vertex in a graph is the number of edges incident to it.



Path:

A path is an open walk with no repeated vertex. A cycle is a closed walk with no repeated vertex.

Cycle:

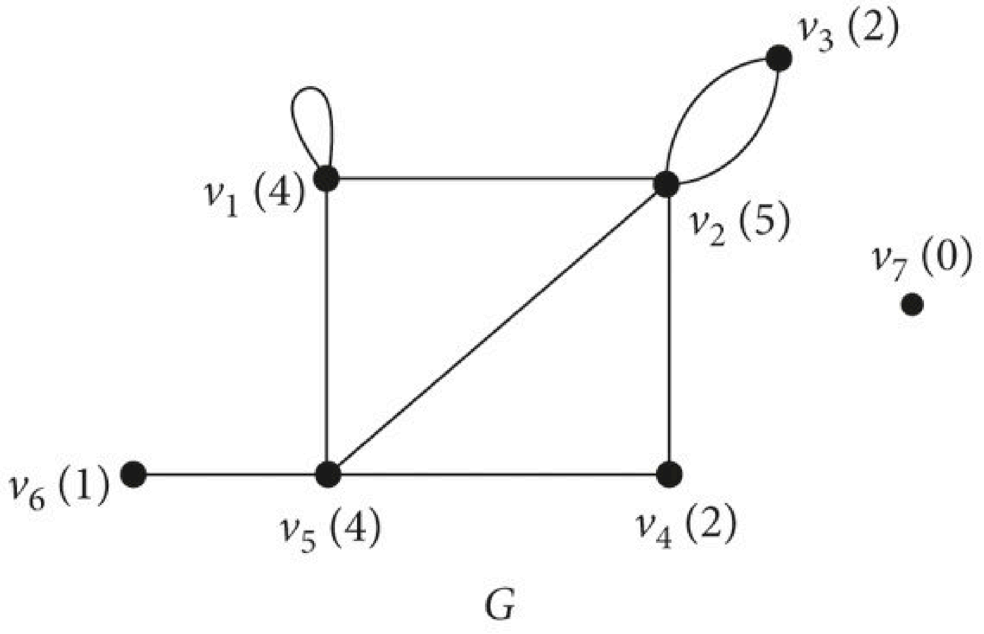
A cycle is a closed walk with no repeated vertex.

Girth:

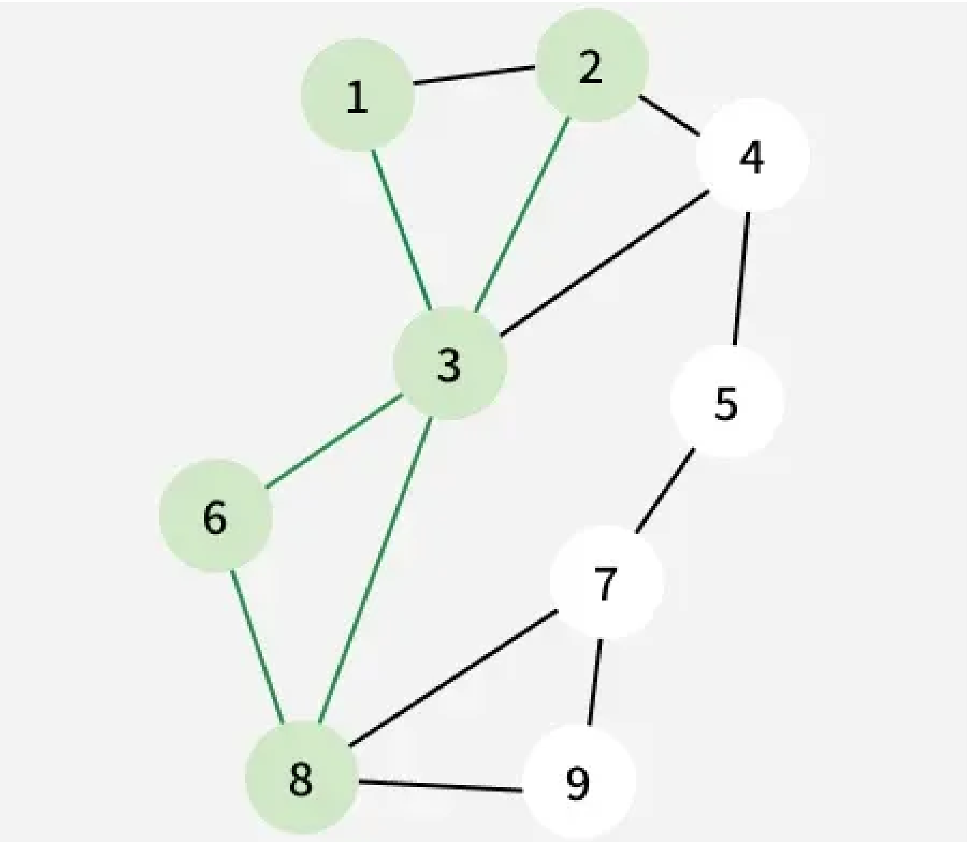
Minimum length of a cycle in a path

THEOREM 1.1. *If G is a graph, then*

$$2|E(G)| = \sum_{u \in V(G)} \deg(u).$$



LEMMA 1.2. *If G is a graph, then G contains a path of length $\delta(G)$, and if $\delta(G) \geq 2$, then $g(G) \leq \delta(G) + 1$.*

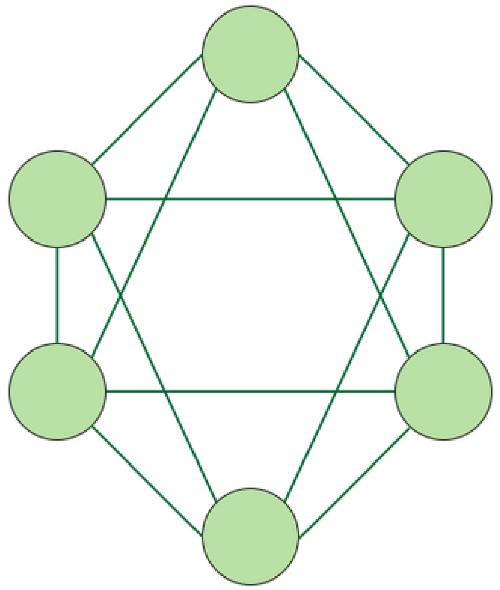


Proofs:

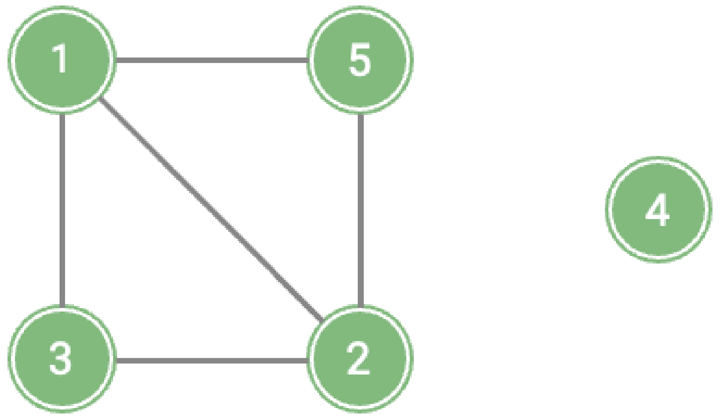
Proof. Let P be a path of maximum length r in G, with consecutively joined vertices $u_0, \ldots, u_r$. As all the neighbours of u_r are in P, we have that $r \geq \delta(G)$.

Assume that $\delta(G) \geq 2$. Let u_m be the vertex joined to u_0 with maximum index, where $m \geq 2$. Then the cycle formed by joining $u_0 u_m$ to the path from u_0 to u_m has length at least $\delta(G) + 1$. $\square$

Connected Graphs



Disconnected Graphs



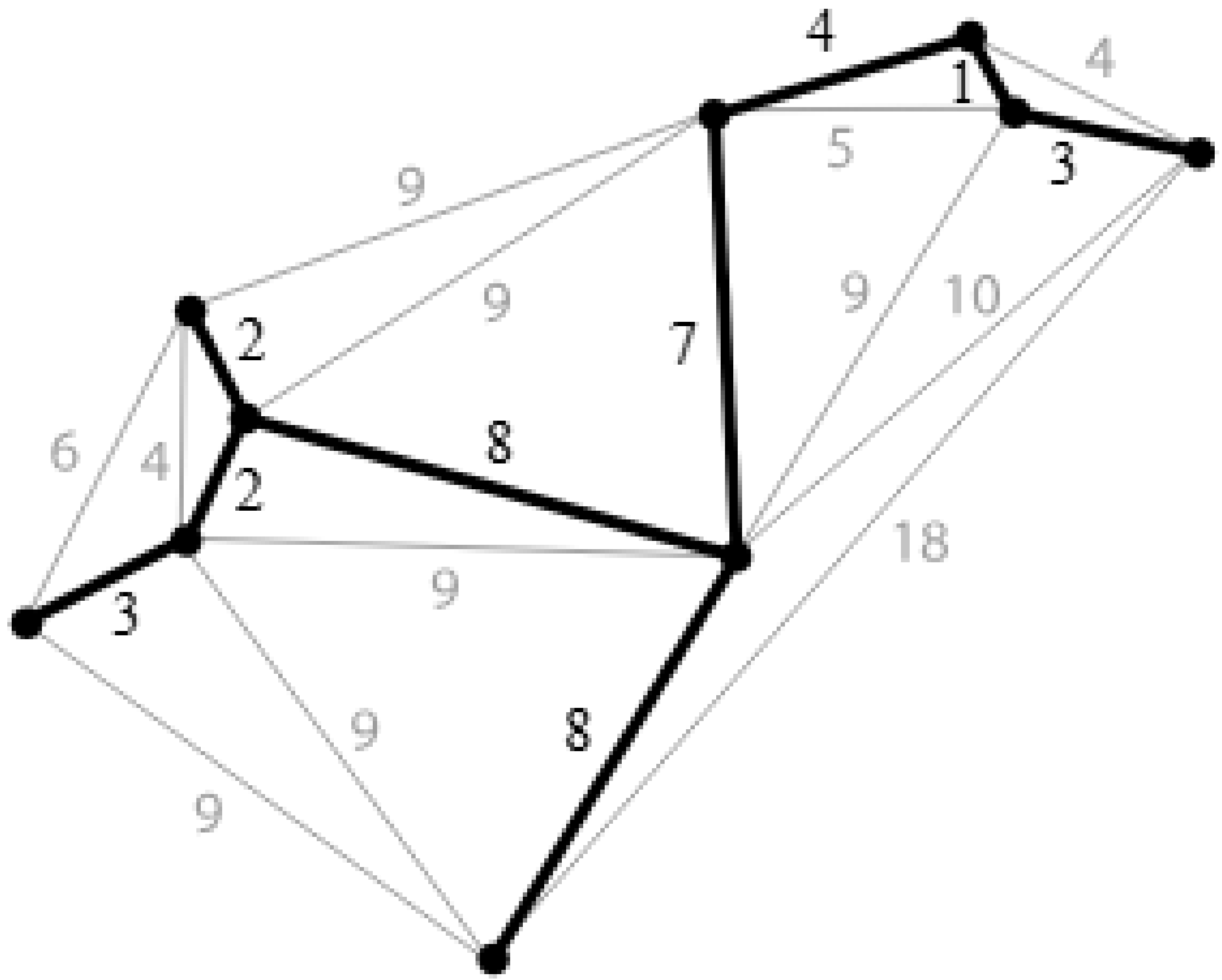
Distance - d(u,v)

Shortest path length between the vertices u and v

Diameter - Diam(G)

Supremum of all distances between distinct pairs of vertices.

Spanning Tree



A spanning tree of a connected, undirected graph GG is a subgraph that includes all of GG's vertices with the minimum number of edges required to keep it connected

THEOREM 1.3. *A graph G is connected if and only if it contains a spanning tree.*

Theorem 1.4. *A graph G is a tree if and only if any one of the following properties is satisfied.*

(1) *Every pair of distinct vertices of G is connected by a unique path.*
(2) *The graph G is connected, but deleting any edge disconnects G.*
(3) *The graph G is acyclic, but adding any edge to G forms a cycle.*
(4) *The graph G is connected, and has $|V(G)| - 1$ many edges.*

Hamilton Cycle:

A spanning subgraph that is also a cycle

Ray:

A Ray in an infinite graph is an infinite one way path with the elements of N in it and edges i(i+1) for all i in N.

Hamilton Path:

It is a spanning ray in an infinite graph

The **chromatic number** of G, written **x(G)**, is the minimum cardinal n with the property that V(G) may be partitioned into n independent sets;

$\chi(G) = 2$ => Graph is **bipartite**

THEOREM 1.5. *A graph G is bipartite if and only if it has no odd cycle as a subgraph.*

The **clique number** of G, written **w(G)**, is the order of a largest clique (more precisely, the supremum of all orders of cliques), while the **independence** (or stability) **number** of G, written **a(G)**, is the order of a largest independent set.

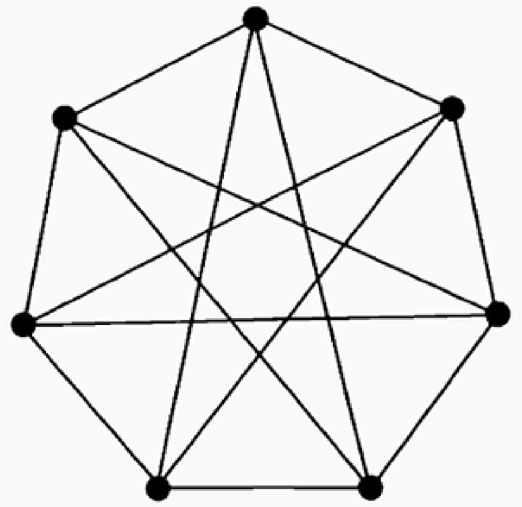


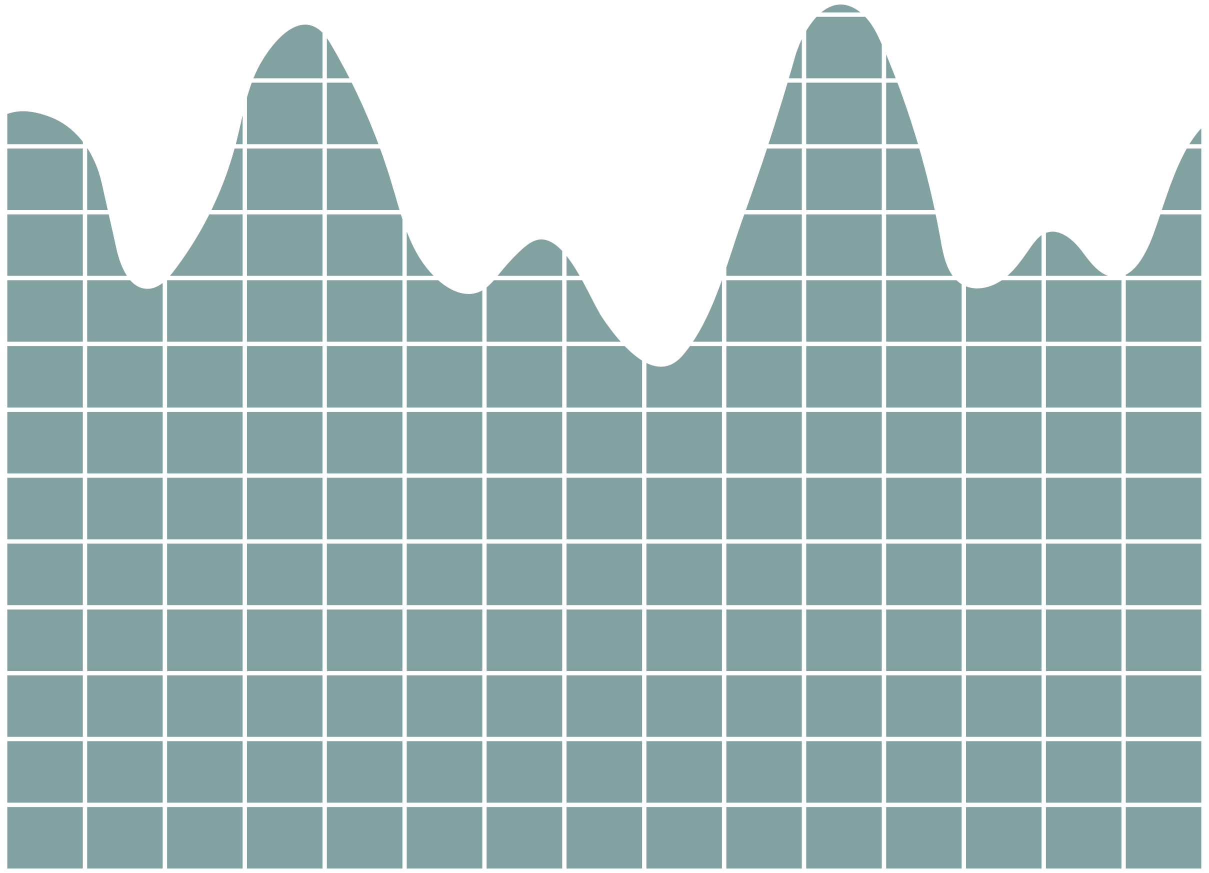
Figure 1.3. A graph G with $\omega(G) = \alpha(G) = 3$ and $\chi(G) = 4$.

THEOREM 1.6. *If G is a graph, then* $\omega(G) \leq \chi(G)$, *while*

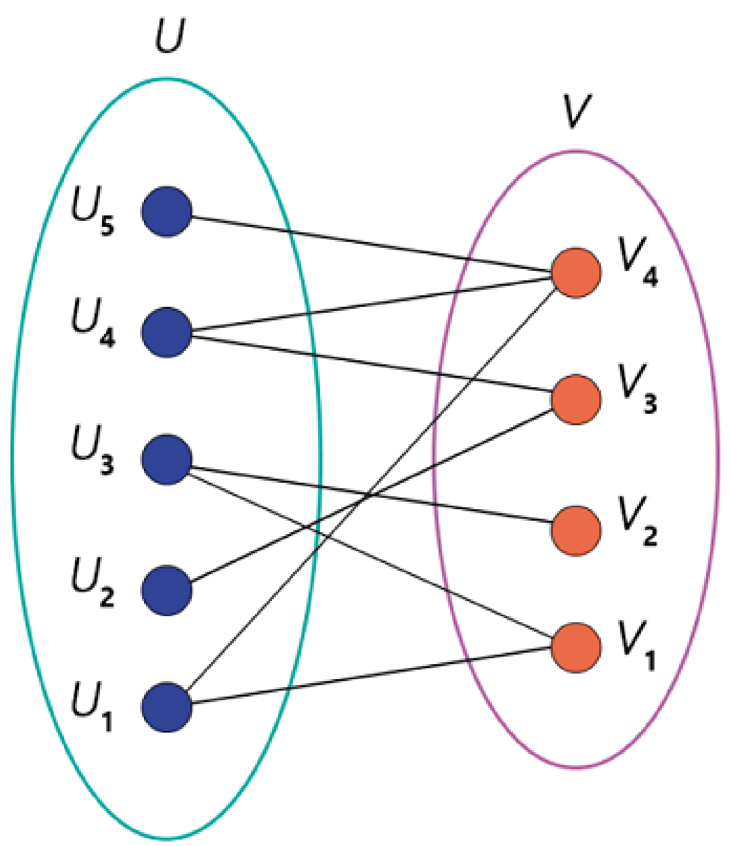
$$\chi(G) \geq |V(G)|/\alpha(G).$$

PROBABILITY THEORY

- Probability theory is much more than just gambling—it has significant applications in graph theory and theoretical computer science.
- We will explore how concepts like expectation, variance, and concentration are fundamental to analyzing models of the web graph.
- To keep things both straightforward and useful, our focus will be limited to discrete probability spaces.



A (discrete) **probability space S** consists of a triple (S, F , P). The set S, called the sample space, is non-empty and countable (although we usually take S to be finite)



The **set F** is the collection of all **subsets** of S

P is a function F -> R and is called the **probability measure**

(1) For all events A, $\mathbb{P}(A) \in [0,1]$, and $\mathbb{P}(S) = 1$.

(2) If $(A_i : i \in I)$ is a countable set of events that are pairwise disjoint, then

$$\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbb{P}(A_i).$$

Some Lemmas on Probability Theory

If $(S, \mathcal{F}, \mathbb{P})$ *is a probability space and* $A, B \in \mathcal{F}$, *then*

(1) $\mathbb{P}(\varnothing) = 0.$

(2) $\mathbb{P}(S \backslash A) = 1 - \mathbb{P}(A).$

(3) *If* $A \subseteq B$, *then* $\mathbb{P}(A) \leq \mathbb{P}(B).$

(4) *If* $(A_i : i \in I)$ *is a countable set of events, then*

$$\mathbb{P}\left(\bigcup_{i \in I} A_i\right) \leq \sum_{i \in I} \mathbb{P}(A_i).$$

Conditional Probability

If $\mathbb{P}(B) > 0$, then the *conditional probability* that A occurs given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Independent Events

Two events A and B are said to be independent if the occurrence of one does not affect the probability of the other occurring. Mathematically, this means:

$P(A \cap B) = P(A) \cdot P(B)$

where:

$P(A)$ = Probability of event A occurring.

$P(B)$ = Probability of event B occurring.

$P(A \cap B)$ = Probability that both A and B occur.

Random variables

Let (**Ω,F,P**) be a probability space

A **random variable** is a function X:Ω→R such that for every real number x, the set {ω∈Ω | X(ω)≤x} is in F. This ensures that probability calculations involving X are **well-defined**.

THEOREM 1.8. *Suppose that X, Y, and X_i, where $1 \leq i \leq n$, are random variables defined on a probability space. Then the following properties hold.*

(1) *(Linearity of expectation) Let c_i, where $1 \leq i \leq n$, be real numbers. Then*

$$\mathbb{E}\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i \mathbb{E}\left(X_i\right).$$

(2) *(Monotonicity) If $X \leq Y$ (that is, $X(s) \leq Y(s)$ for all $s \in S$), then $\mathbb{E}(X) \leq \mathbb{E}(Y)$.*

Theorem 1.9 (Markov's inequality). *Let $X \geq 0$ be a random variable on a probability space. If c is a positive real number, then*

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}(X)}{c}.$$

Theorem 1.10 (Chebyshev's inequality). *Let X be a random variable on a probability space. If c is a positive real number, then*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

THANK YOU