Course 50.050/50.550
Advanced Algorithms

Week 3 – Lecture L03.01

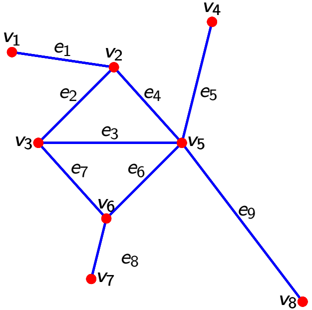
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Outline of Lecture

- Basic graph-theoretic terminology
- Formal definitions for undirected graphs and directed graphs
- Basic graph properties

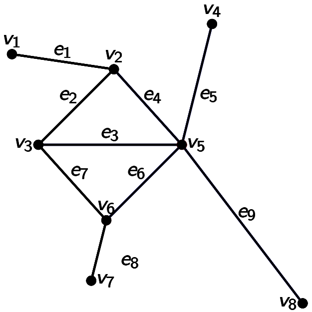
Informal intuition for graphs

Intuition: A graph can be visualized as a collection of points and some line segments joining pairs of points. The points and lines are what we call **vertices** and **edges**.



- This is a visualization of a graph.
  - There are 8 vertices drawn (depicted in **red**).
    - **Note:** The word 'vertices' is the plural of 'vertex'.
  - There are 9 edges drawn (depicted in **blue**).
- We can give names to vertices (e.g. $v_1, v_2, \dots,$ or $a, b, c \dots$).
- Similarly, we can give names to edges (e.g. $e_1, e_2, \dots$).

Informal intuition for graphs (continued)



- Two vertices are **adjacent** if they are joined by an edge.
  - We also say the two vertices are **incident** to this edge.
  - For example, v_5 and v_6 (depicted in **red**) are adjacent vertices, and they are both incident to edge e_6 (depicted in **blue**).
  - We also say that v_5 is a **neighbor** of v_6 (and vice versa).
- The **degree** of a vertex is the number of incident edges.
  - For example, v_5 (depicted in **red**) has 5 incident edges (depicted in **blue**). So v_5 has degree 5, and we write $\deg(v_5) = 5$.

Understanding graph-theoretic terminology

Important Note: **Graph-theoretic terminology is a mess!**

- There is currently still no standardization, and we would probably need several more decades before the terminology becomes more consistent across different textbooks.

Famous quote (from 1986) by famous mathematician Richard Stanley:

> The number of systems of terminology presently
> used in graph theory is equal, to a close approximation,
> to the number of graph theorists.

- Graph theory was first introduced in 1736, but the first textbook on graph theory only appeared in 1936. *(200 years later!)*
  - The second textbook on graph theory appeared in 1969. Then, since the 1970s, there was suddenly a surge in the number of graph theory textbooks, all using different terminology and notation.
- Different graph theorists seem to be trying to popularize the terminology/notation that they work with.

How do we handle this mess in this course?

- We try to use more widely accepted convention where possible.
- We will point out other commonly used terminology/notation.
- We will be consistent throughout the course.

What is a graph?

Even an innocent question like "What is a graph?" has many answers!

- First of all, do we mean a directed or undirected graph?
  - **Intuition:** For directed graphs, edges have "directions".
- Do we mean a simple graph, or do we allow non-simple graphs?
  - **Intuition:** A simple graph is a graph with no loops or multiple edges.

Definition: A simple undirected graph is an ordered pair (V, E), where

- V is a set, whose elements are called vertices.
  - Singular: vertex. (**Note: NOT vertice.**)
    - **Warning:** Some CS textbooks still use **node** to mean vertex.
  - The set V is called the vertex set of the graph.
- E is a set whose elements are called edges, such that every edge is a 2-subset of V.
  - If v_1, v_2 are distinct vertices, then $\{v_1, v_2\}$ could be an edge, and the vertices v_1, v_2 are called the endpoints of this edge.
  - The set E is called the edge set of the graph.
    - **Note:** E is a set of sets!
    - Remember: We can give names to sets, so we could for example give a name e_1 to an edge $\{v_1, v_2\}$.
    - Then, whether we write "e_1" or "$\{v_1, v_2\}$", it refers to the same edge. In other words, $e_1 = \{v_1, v_2\}$. (Note: e_1 is a set.)

What is a graph? (continued)

Definition: A simple directed graph is an ordered pair (V, E), where

- V is a set, whose elements are called vertices.
  - Same singular: vertex. (**Note: NOT vertice**.)
    - **Same warning:** Some CS textbooks still use **node** to mean vertex.
  - The set V is called the vertex set of the graph.
- E is a set whose elements are called edges, such that every edge is a 2-tuple of V consisting of two distinct vertices.
  - If v_1, v_2 are distinct vertices, then (v_1, v_2) could be an edge, and the vertices v_1, v_2 are called the endpoints of this edge.
  - Edges of a directed graph are sometimes called directed edges to emphasize that our edges are edges of a directed graph.
  - **Same terminology:** The set E is called the edge set of the graph.
    - **Note:** E is a set of tuples!
    - Remember: We can give names to tuples, so we could for example give a name e_1 to an edge (v_1, v_2).
    - Then, whether we write "e_1" or "(v_1, v_2)", it refers to the same edge. In other words, $e_1 = (v_1, v_2)$. (Note: e_1 is a tuple.)

Only distinction for directed/undirected simple graphs $(\mathbf{V}, \mathbf{E})$:

- Undirected case: Edges are sets. An edge is a 2-subset of V.
- Directed case: Edges are tuples. An edge is a 2-tuple of V.

Remarks on simple graphs

Note: A simple undirected/directed graph is an ordered pair (V, E).

- i.e. a simple graph is a 2-tuple.
  - **Intuition:** We are distinguishing two different kinds of objects: vertices and edges. We put these two different collections of objects into a tuple, rather than a set, because we want to distinguish them.
  - **Note:** When you see the phrase "simple graph" (i.e. it is not indicated whether the graph is undirected or directed), it is more common that the author means a "simple undirected graph".
    - Unfortunately, it is still relatively common to also see the usage "simple graph" when the author means a "simple directed graph".
- We can give names to tuples, so we could for example give a name G to this 2-tuple.
  - **Notation:** We usually write "Let $G = (V, E)$ be a simple graph..." to mean that G is our name for the graph, that V is the vertex set of G, and that E is the edge set of G.

Note: For simple graphs, an edge must contain two **distinct** vertices!

- Undirected case: For any $v \in V$, $\{v, v\}$ is not a 2-subset of V.
- Directed case: An edge (v_1, v_2) must satisfy $v_1 \neq v_2$.

Remarks on simple directed graphs

Let $G = (V, E)$ be a simple directed graph.

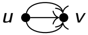
- **Recall:** An edge $e = (u, v)$ of G is a 2-tuple in $V \times V$, such that $u \neq v$. Informally, we say that e is an edge from u to v.
  - The first vertex u is called the tail of e.
  - The second vertex v is called the head of e.
  - The edge (u, v) is sometimes written as $u \to v$.
    - **Warning:** This is non-standard/amateurish/informal notation!
    - $u \to v$ is shorthand for the drawing $u \bullet\!\!\longrightarrow\!\!\bullet\, v$.
- **Note:** By definition, the edge set E is a set of tuples.
  - **Recall:** By definition, a set has no repeated elements.
  - **Recall:** Two 2-tuples (u_1, v_1), (u_2, v_2) are equal exactly when $u_1 = u_2$ and $v_1 = v_2$.
  - Thus, E cannot have repeated 2-tuples as different edges.
  - However, if u, v are distinct vertices, then (u, v) and (v, u) are **distinct** 2-tuples, so they could be two different edges in E.
    - We usually draw $\begin{tikzcd} u \bullet \arrow[r, bend left] & \bullet\, v \arrow[l, bend left] \end{tikzcd}$, and sometimes draw $u \bullet\!\!\longleftrightarrow\!\!\bullet\, v$.
    - A few authors write $u \leftrightarrow v$ as the shorthand for the drawing $u \bullet\!\!\longleftrightarrow\!\!\bullet\, v$. Again this is non-standard/amateurish/informal!

How to think about non-simple graphs?

Intuition: A graph is simple if it has no loops or multiple edges.

- Informally, a **loop** is an edge that starts from some vertex v, and ends at the same vertex v.
  - For undirected case: We draw 

; for directed case: we draw 

.
- Informally, a graph has **multiple edges** if there are multiple edges joining the same two vertices, e.g. as depicted by 

- Non-simple graphs are important! They are heavily used to model state transitions in dynamical systems.

Question: Can we modify our definitions to allow loops/multiple edges?

Problem 1: An edge set is no longer a set if we have multiple edges.
Problem 2: Even if we consider a collection of edges, two repeated edges cannot be distinguished apart.

- e.g. in the directed case, there could be two copies of (u, v), but we cannot tell them apart.

Solution: We still use an edge set $E = \{e_1, e_2, \ldots, e_m\}$, where we use different names for different edges to tell them apart.

- **New idea:** We introduce a function from E to sets of vertices.

Formal definition for undirected graphs

Definition: An undirected graph is a 3-tuple $G = (V, E, \phi)$, where

- V is a set, whose elements are called vertices.
  - The set V is called the vertex set of G.
- E is a set whose elements are called edges.
  - The set E is called the edge set of G.
- ϕ is a function with domain E, such that every edge $e \in E$ is mapped to a subset of V of cardinality 1 or 2.
  - **Note:** The image of ϕ is contained in $\mathcal{P}(V)$, i.e. $\phi(E) \subseteq \mathcal{P}(V)$.
    - **Recall:** $\mathcal{P}(V)$, the power set of the set V, contains all subsets of V.
  - ϕ is called the incidence function of G.
    - **Note:** No standardized symbol to represent this incidence function.
    - Other common symbols used include: ψ, I, i.
  - An edge e is called a loop if $|\phi(e)| = 1$, and a non-loop if $|\phi(e)| = 2$.
  - If $v \in V$ and $e \in E$, such that $v \in \phi(e)$, then we say that v is incident to e, or equivalently, that v is an endpoint of e.

Use of terminology: "Let $G = (V, E, \phi)$ be an undirected graph..."

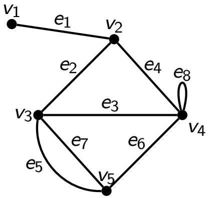
- **Intuition:** V is the set of names for the vertices, E is the set of names for the edges, and ϕ encodes information about how the edges relate to the vertices.

Example of non-simple undirected graph

Consider the undirected graph $G = (V, E, \phi)$, where

- $V = \{v_1, v_2, v_3, v_4, v_5\}$ is the vertex set of G.
- $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$ is the edge set of G.
- $\phi : E \to \mathcal{P}(V)$ is the incidence function of G given below.

$\phi(e_1) = \{v_1, v_2\}$;
$\phi(e_2) = \{v_2, v_3\}$;
$\phi(e_3) = \{v_3, v_4\}$;
$\phi(e_4) = \{v_2, v_4\}$;
$\phi(e_5) = \{v_3, v_5\}$;
$\phi(e_6) = \{v_4, v_5\}$;
$\phi(e_7) = \{v_3, v_5\}$;
$\phi(e_8) = \{v_4\}$.



Some Observations:

- e_8 is a loop, since $|\phi(e_8)| = 1$.
- e_8 has only v_4 as its single endpoint, since $\phi(e_8) = \{v_4\}$.
- e_7 has two endpoints v_3 and v_5, since $\phi(e_7) = \{v_3, v_5\}$.
- v_1 is incident to e_1, since $v_1 \in \phi(e_1)$.

Formal definition for directed graphs

Definition: A directed graph is a 3-tuple $G = (V, E, \phi)$, where

- V is a set, whose elements are called vertices.
  - The set V is called the vertex set of G.
- E is a set whose elements are called edges (or directed edges).
  - The set E is called the edge set of G.
- ϕ is a function $\phi : E \to V \times V$.
  - **Note:** The image of ϕ contains ordered pairs of vertices.
  - ϕ is called the incidence function of G.
    - **Note:** No standardized symbol to represent this incidence function.
    - Other common symbols used include: ψ, I, i.
  - An edge e is called a loop if $\phi(e) = (v, v)$ for some $v \in V$, and called a non-loop otherwise.
  - If $v \in V$ and $e \in E$, such that $v \in \phi(e)$, then we say that v is incident to e, or equivalently, that v is an endpoint of e.
    - If $\phi(e) = (v_1, v_2)$, then v_1 is called the tail of e, and v_2 is called the head of e. We also say that e is an edge from v_1 to v_2.

Use of terminology: "Let $G = (V, E, \phi)$ be a directed graph..."

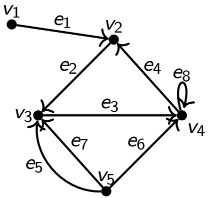
- **Same Intuition:** V is the set of names for the vertices, E is the set of names for the edges, and ϕ encodes information about how the edges relate to the vertices.

Example of non-simple directed graph

Consider the directed graph $G = (V, E, \phi)$, where

- $V = \{v_1, v_2, v_3, v_4, v_5\}$ is the vertex set of G.
- $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$ is the edge set of G.
- $\phi : E \to V \times V$ is the incidence function of G given below.

$\phi(e_1) = (v_1, v_2)$;
$\phi(e_2) = (v_2, v_3)$;
$\phi(e_3) = (v_3, v_4)$;
$\phi(e_4) = (v_4, v_2)$;
$\phi(e_5) = (v_5, v_3)$;
$\phi(e_6) = (v_5, v_4)$;
$\phi(e_7) = (v_5, v_3)$;
$\phi(e_8) = (v_4, v_4)$.



Some Observations:

- e_8 is a loop, since $\phi(e_8) = (v_4, v_4)$, whose both entries are equal.
  - v_4 is both the head and the tail of e_8.
- e_8 has only v_4 as its single endpoint, since $\phi(e_8) = (v_4, v_4)$.
- e_7 has two endpoints v_3 and v_5, since $\phi(e_7) = (v_5, v_3)$.
  - v_5 is the tail of e_7, while v_3 is the head of e_7.
- v_1 is incident to e_1, since v_1 is an entry of $\phi(e_1) = (v_1, v_2)$.

Reconciling different definitions for graphs

Definition: Let $G = (V, E, \phi)$ be an undirected graph or a directed graph. We say that G is simple if E has no loops, and ϕ is injective. (Non-simple graphs are sometimes called multigraphs or pseudographs.)

- We now have two definitions for "simple undirected graph", and two definitions for "simple directed graph".
- **Question:** In either case (undirected or directed), can you see how the two definitions are equivalent?
  - Starting with the 2nd definition with graph (V, E, ϕ), we can directly rename each edge $e \in E$ by its image $\phi(e)$, to get the edge set as defined in the 1st definition.
  - Starting with the 1st definition with graph (V, E), we can define an incidence function $\phi : E \to \mathcal{P}(V)$ via the identity map.
    - Undirected case: $\{u, v\} \mapsto \{u, v\}$
    - Directed case: $(u, v) \mapsto (u, v)$.
- **Intuition:** If E and $\phi(E)$ are exactly the same, the information encoded by ϕ is superfluous.

Remark: When we write "Let $G = (V, E)$ be a simple graph...", we are already assuming that we are using the 1st definition.

Why have such formal definitions?

By being exposed to this level of formalism, it allows you to:

- Read up more graph theory literature by youself.
  - You can read latest research papers without being confused.
- Be aware how to code/store non-simple graphs.
  - In 50.004, we learned how to use adjacency lists/matrices to store the graph structure of simple graphs. These graph representations no longer work for non-simple graphs.
  - Instead, we can use incidence lists/matrices, or even a Python dictionary with key-value pairs of the form "$e : \phi(e)$".
- Be able to easily understand generalizations of graphs (and how to code them), e.g. mixed graphs, hypergraphs, etc.

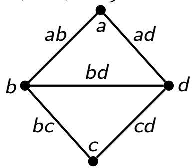
Only if you are interested: Recall that in our definition of a undirected graph $G = (V, E, \phi)$, the incidence function $\phi : E \to \mathcal{P}(V)$ must satisfy the condition that $\phi(e)$ has cardinality either 1 or 2 for all edges $e \in E$. If we instead allow $\phi(e)$ to be any non-empty subset of V, then we get what is called a hypergraph.

- Intuitively, hypergraphs are generalized graphs where the edges are allowed to be non-empty subsets of the vertex set.
- Hypergraphs are used in natural languge processing (NLP) models, and are used to model correlations or non-binary relations in large networks.

More on graph terminology

- Directed graphs are *sometimes* called digraphs.
  - Non-simple directed graphs are *sometimes* called directed multigraphs or multidigraphs.
  - For a directed edge (u, v), the tail v_1 is *sometimes* called the initial vertex, and the head v_2 is *sometimes* called the terminal vertex, or (confusingly), the end vertex.
- Frequently, we could see "Let G be a graph...", where the 2-tuple (V, E) or 3-tuple (V, E, ϕ) is not explicitly indicated.
  - In this context, it is very common to denote the vertex set of G by $V(G)$, and the edge set of G by $E(G)$.
- For a simple undirected graph G, an edge $\{u, v\}$ is *sometimes* abbreviated as uv. We think of uv as the name for the set $\{u, v\}$.

Example: Let G be a simple undirected graph with $V(G) = \{a, b, c, d\}$ and $E(G) = \{ab, ad, bd, bc, cd\}$. Then G can be depicted as follows.



Adjacency, neighborhoods, degree

Definition: Let $G = (V, E, \phi)$ be an **undirected** graph.

- If two vertices u, v in V are both incident to a common edge, then we say that u and v are adjacent; we also say that u is a neighbor of v, and that v is a neighbor of u.
  - If $e \in E$ is a loop incident to vertex x, then x is adjacent to x itself, and x is a neighbor of x.
  - Given a vertex $x \in V$, the set of all neighbors of x is called the neighborhood of x, or the open neighborhood of x. This set is commonly denoted by $N_G(x)$, or $N(x)$ (if the context is clear).
    - Other commonly used notation: $\Gamma_G(x)$, $\Gamma(x)$.
  - Given a vertex $x \in V$, the closed neighborhood of x is the set consisting of x together with all neighbors of x. This set is commonly denoted by $N_G[x]$, or $N[x]$ (if the context is clear).
    - Other commonly used notation: $\Gamma_G[x]$, $\Gamma[x]$.
    - **Fact:** $N_G[x] = \{x\} \cup N_G(x)$.
- For any vertex $v \in V$, the degree of v, denoted by $\deg_G(v)$ or simply $\deg(v)$, is the number of edges in E incident to v.
  - **Convention:** Every loop incident to v contributes $+2$ to $\deg(v)$.
  - **Convention:** In general, for any notation with the graph G as a subscript, this subscript can be omitted if the context is clear.
- **Fact:** If G is simple, then $\deg_G(x) = |N_G(x)|$ for all $x \in V$.

Mess in terminology for directed graphs

Let $G = (V, E, \phi)$ be a **directed** graph.

- **Warning!** There is no "usual" definition for the adjacency of vertices in directed graphs! Here are 3 common definitions:
  1. [Rosen] If $e \in E$ is an edge such that $\phi(e) = (u, v)$, then we say that u is adjacent to v, and v is adjacent from u.
  2. If $e \in E$ is an edge such that $\phi(e) = (u, v)$, then we say that u is adjacent to v. *(i.e. tail is adjacent to head)*
     - For this definition, v is not adjacent to u unless there exists $e' \in E$ such that $\phi(e') = (v, u)$.
  3. If $e \in E$ is an edge such that $\phi(e) = (u, v)$, then we say that v is adjacent to u. *(i.e. head is adjacent to tail)*
     - For this definition, u is not adjacent to v unless there exists $e' \in E$ such that $\phi(e') = (v, u)$.
- **Warning!** There is no "usual" definition for neighborhoods in directed graphs.
  - *(Let's use "neighborhoods" only for undirected graphs.)*
- **Warning!** There is no "usual" definition for neighbors of vertices in directed graphs.
  - However, some authors use "in-neighbor" and "out-neighbor".
  - *(Let's use "neighbors" only for undirected graphs.)*

In-degree, out-degree

Definition: Let $G = (V, E, \phi)$ be a **directed** graph, and suppose there exists an edge $e \in E$ such that $\phi(e) = (u, v)$.

- We say that u is an in-neighbor or incoming neighbor of v.
- We say that v is an out-neighbor or outgoing neighbor of u.
  - **Intuition:** For a directed edge (u, v), which can be drawn as $u \bullet\!\!\longrightarrow\!\!\bullet\, v$, it is an edge **from** u **to** v, so we think of u as an incoming neighbor of v that is "pointing inwards to" v, and similarly, we can thinking of v as an outgoing neighbor of u that is "pointing outwards from" u.

Definition: Let $G = (V, E, \phi)$ be a **directed** graph, and let $x \in V$.

- The in-degree of x is the number of edges with x as the head.
  - Usually denoted by $\deg_G^-(x)$, and sometimes by $\mathrm{indeg}_G(x)$.
- The out-degree of x is the number of edges with x as the tail.
  - Usually denoted by $\deg_G^+(x)$, and sometimes by $\mathrm{outdeg}_G(x)$.
- **Note:** Every loop $e \in E$ that is incident to x contributes $+1$ to $\deg_G^-(x)$, and contributes $+1$ to $\deg_G^+(x)$.
  - This is because $\phi(e) = (x, x)$, where x is both an in-neighbor and an out-neighbor, of x itself.

Fact: If $G = (V, E, \phi)$ is a simple directed graph, then for all $x \in V$, $\deg^-(x)/\deg^+(x)$ is the number of in-neighbors/out-neighbors of x.

Sum of in-degrees and sum of out-degrees

Theorem: Let $G = (V, E, \phi)$ be a **directed** graph. Then, we have
$$|E| = \sum_{v \in V} \deg^-(v) = \sum_{v \in V} \deg^+(v).$$

Proof: (by double counting) Define the following two sets:

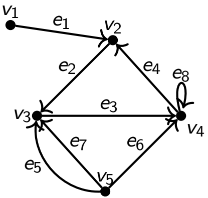
- $S_{\text{head}} := \{(e, v) \in E \times V | v \text{ is the head of } e\}$.
- $S_{\text{tail}} := \{(e, v) \in E \times V | v \text{ is the tail of } e\}$.

Every edge has exactly one head and one tail, so $|S_{\text{head}}| = |S_{\text{tail}}| = |E|$.
Next, we now count $|S_{\text{head}}|$ and $|S_{\text{tail}}|$ via a different approach.

- Every $v \in V$ is the head of exactly $\deg^-(v)$ edges.
- Every $v \in V$ is the tail of exactly $\deg^+(v)$ edges.

Thus, $|S_{\text{head}}| = \sum_{v \in V} \deg^-(v)$ and $|S_{\text{tail}}| = \sum_{v \in V} \deg^+(v)$. $\square$

Example:



There are 8 edges.
$\deg^-(v_1) = 0, \quad \deg^+(v_1) = 1;$
$\deg^-(v_2) = 2, \quad \deg^+(v_2) = 1;$
$\deg^-(v_3) = 3, \quad \deg^+(v_3) = 1;$
$\deg^-(v_4) = 3, \quad \deg^+(v_4) = 2;$
$\deg^-(v_5) = 0, \quad \deg^+(v_5) = 3.$

Degree sum formula

Corollary: (**degree sum formula**) Suppose that $G = (V, E, \phi)$ is an **undirected** graph. Then, we have $2|E| = \sum_{v \in V} \deg(v)$.

- **Idea of proof:** Convert G into a directed graph by randomly choosing a direction for every edge in G.

Proof: Starting with G, we construct a directed graph $G' = (V, E, \phi')$ with the same vertex set V and the same set E of names for the edges, but with a new incidence function ϕ', given as follows:

- If $e \in E$ is a non-loop, then randomly choose one of the vertices in the pair $\phi(e)$ to be the new head (say v), and let the other vertex in $\phi(e)$ be the new tail (say u), so that $\phi'(e) := (u, v)$.
- If $e \in E$ is a loop, with $\phi(e) = \{v\}$, then define $\phi'(e) := (v, v)$.

Note: For every vertex $v \in V$, $\deg_{G'}(v) = \deg^-_G(v) + \deg^+_G(v)$.

- Thus, the degree sum formula follows from the theorem given on the previous slide. $\square$

Note: The degree sum formula is sometimes called the handshaking lemma, but "**handshaking lemma**" usually refers to another theorem (see next slide).

Handshaking lemma

Theorem: (**handshaking lemma**) An undirected graph has an even number of vertices of odd degree.

- **Informal:** At a party, where people go around to shake hands, the number of people who shake an odd number of other people's hands is even.

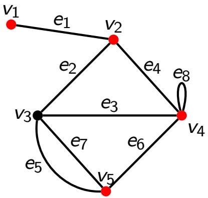
Proof: If G is an undirected graph, then by the degree sum formula,

$$2|E(G)| = \sum_{v \in V(G)} \deg(v) = \sum_{\substack{v \in V(G) \\ \deg(v) \text{ is even}}} \deg(v) + \sum_{\substack{v \in V(G) \\ \deg(v) \text{ is odd}}} \deg(v).$$

Note that $2|E(G)|$ is even. The blue term is a sum of even integers and hence also even. Thus, the green term must be even, which means there must be an even number of vertices of odd degree. $\square$

Example:

$\deg(v_1) = 1;$
$\deg(v_2) = 3;$
$\deg(v_3) = 4;$
$\deg(v_4) = 5;$
$\deg(v_5) = 3.$



Vertices with odd degree are colored in red.

Summary

- Basic graph-theoretic terminology
- Formal definitions for undirected graphs and directed graphs
- Basic graph properties