ECE 232E Lecture 1 and Lecture 2 notes

Professor Vwani Roychowdhury

April 13, 2018

In this set of lecture notes, we define graphs and study some of their structural properties and representations.

1 Definition of a graph

A graph G is defined as the tuple

$$G = (V, E, W_E, W_V)$$

where

$$\begin{aligned} V &: \textit{Set of vertices} \\ E &: \textit{Set of edges} \\ W_E &: \textit{Set of edge weights} \\ W_V &: \textit{Set of vertex weights} \end{aligned}$$

The vertex set, V, contains the vertices of the graph G

$$V = \{v_1, v_2, v_3, \cdots, v_n\}$$

The edge set, E, contains the edges of the graph and are two-element subsets of V

$$E = \{(v_1, v_5), (v_2, v_4), (v_3, v_1), \cdots\}$$

The set of edge weights, W_E, contains the weights associated with the edges of the graph

$$W_E : E \to \mathbb{R}^m$$

The set of vertex weights, W_V, contains the weights associated with the vertices of the graph

$$W_V : V \to \mathbb{R}^l$$

In order to illustrate the above definition, let's consider the graph, G_1, shown in figure 1

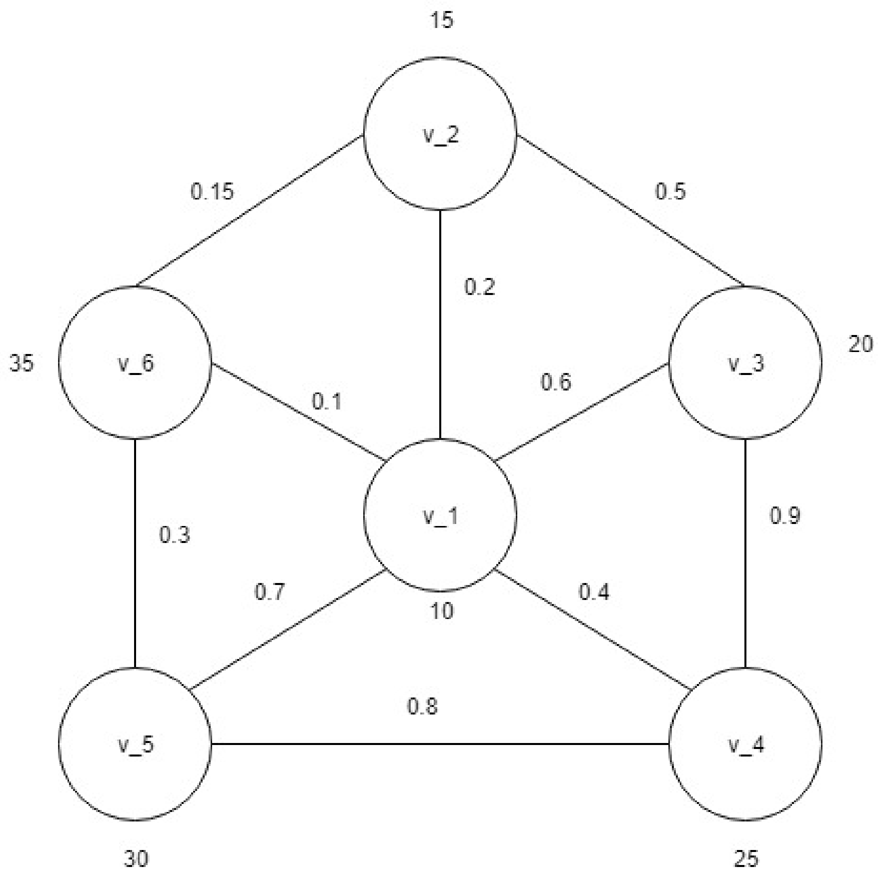

Figure 1: Graph G_1

For the graph G_1, we have

$$
\begin{aligned}
V &= \{v_1, v_2, v_3, v_4, v_5, v_6\} \\
E &= \{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5), (v_1, v_6), (v_2, v_3), (v_2, v_6), (v_6, v_5), (v_5, v_4), (v_4, v_3)\} \\
W_E &= \{0.2, 0.6, 0.4, 0.7, 0.1, 0.5, 0.15, 0.3, 0.8, 0.9\} \\
W_V &= \{10, 15, 20, 25, 30, 35\}
\end{aligned}
$$

1.1 Directed graph

A graph, G, is directed if there is a direction associated with the edges of the graph. For directed graphs, the edge set, E, contains ordered two-element subsets of V. A directed graph, G_2, is shown in figure 2

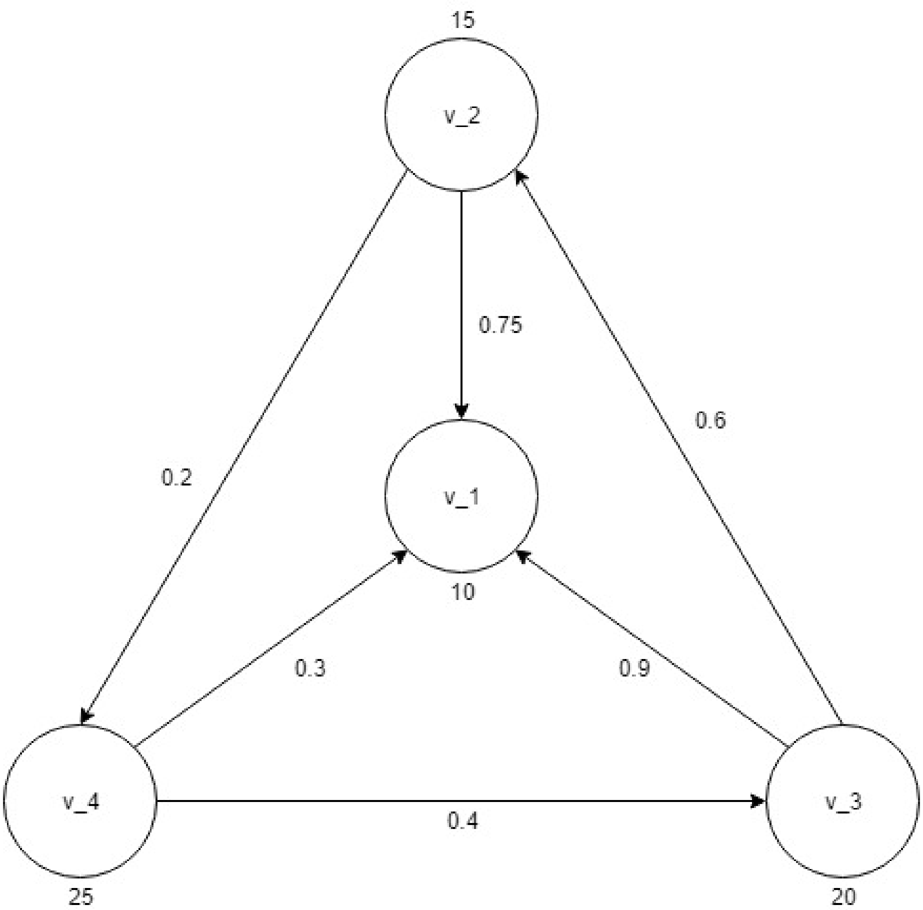

Figure 2: Directed graph G_2

For the directed graph, G_2, we have

$$\begin{aligned}
V &= \{v_1, v_2, v_3, v_4\} \\
E &= \{(v_2, v_1), (v_3, v_1), (v_4, v_1), (v_2, v_4), (v_3, v_2), (v_4, v_3)\} \\
W_E &= \{0.75, 0.9, 0.3, 0.2, 0.6, 0.4\} \\
W_V &= \{10, 15, 20, 25\}
\end{aligned}$$

2 Structural properties of a graph

Having defined a graph in the previous section, now we will study some of the structural properties of a graph. In this section, we introduce the following structural properties:

- Degree distribution
- Paths and Cycles
- Connectivity and Diameter

2.1 Degree distribution

In an undirected graph, the degree of a vertex v_i, denoted as $deg(v_i)$, is defined as

$$deg(v_i) = |\{(v_i, v_j) \in E\}|$$

For the undirected graph G_1 (shown in figure 1), we have

$$\begin{aligned}
deg(v_1) &= |\{(v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5), (v_1, v_6)\}| = 5 \\
deg(v_2) &= |\{(v_1, v_2), (v_2, v_3), (v_2, v_6)\}| = 3 \\
deg(v_3) &= |\{(v_1, v_3), (v_2, v_3), (v_4, v_3)\}| = 3 \\
deg(v_4) &= |\{(v_1, v_4), (v_4, v_3), (v_5, v_4)\}| = 3 \\
deg(v_5) &= |\{(v_1, v_5), (v_5, v_4), (v_6, v_5)\}| = 3 \\
deg(v_6) &= |\{(v_1, v_6), (v_2, v_6), (v_6, v_5)\}| = 3
\end{aligned}$$

Having defined the degree of a node, we can now define the degree distribution of a graph. Let X be a random variable denoting the degree of a randomly picked vertex of the graph. Then, the probability distribution function (pdf) of the random variable X is given below

$$\mathbb{P}(X = k) = \frac{\textit{number of vertices with degree } k}{\textit{total number of vertices}}$$

The pdf given above is the degree distribution of the graph. The degree distribution of the undirected graph G_1 is given below

$$\begin{aligned}
\mathbb{P}(X = 3) &= \frac{5}{6} \\
\mathbb{P}(X = 5) &= \frac{1}{6}
\end{aligned}$$

Here for example,

$$\mathbb{P}(X = 2) = 0$$

In a directed graph, the in-degree of a vertex v_i, denoted as $deg_{in}(v_i)$, is defined as

$$deg_{in}(v_i) = |\{(v_j, v_i) \in E\}|$$

For the directed graph G_2 (shown in figure 2), we have

$$\begin{aligned}
deg_{in}(v_1) &= |\{(v_2, v_1), (v_3, v_1), (v_4, v_1)\}| = 3 \\
deg_{in}(v_2) &= |\{(v_3, v_2)\}| = 1 \\
deg_{in}(v_3) &= |\{(v_4, v_3)\}| = 1 \\
deg_{in}(v_4) &= |\{(v_2, v_4)\}| = 1
\end{aligned}$$

Having defined the in-degree of a node, we can now define the in-degree distribution of a directed graph. Let Y be a random variable denoting the in-degree of a randomly picked vertex of the graph. Then, the probability distribution function (pdf) of the random variable Y is given below

$$\mathbb{P}(Y = k) = \frac{\textit{number of vertices with in-degree } k}{\textit{total number of vertices}}$$

The pdf given above is the in-degree distribution of the graph. The in-degree distribution of the directed graph G_2 is given below

$$\mathbb{P}(Y = 1) = \frac{3}{4}$$
$$\mathbb{P}(Y = 3) = \frac{1}{4}$$

Similarly, the out-degree of a vertex v_i, denoted as $deg_{out}(v_i)$, is defined as

$$deg_{out}(v_i) = |\{(v_i, v_j) \in E\}|$$

For the directed graph G_2 (shown in figure 2), we have

$$\begin{aligned}
deg_{out}(v_1) &= |\{\emptyset\}| = 0 \\
deg_{out}(v_2) &= |\{(v_2, v_1), (v_2, v_4)\}| = 2 \\
deg_{out}(v_3) &= |\{(v_3, v_1), (v_3, v_2)\}| = 2 \\
deg_{out}(v_4) &= |\{(v_4, v_1), (v_4, v_3)\}| = 2
\end{aligned}$$

Having defined the out-degree of a node, we can now define the out-degree distribution of a directed graph. Let Z be a random variable denoting the out-degree of a randomly picked vertex of the graph. Then, the probability distribution function (pdf) of the random variable Z is given below

$$\mathbb{P}(Z = k) = \frac{\textit{number of vertices with out-degree } k}{\textit{total number of vertices}}$$

The pdf given above is the out-degree distribution of the graph. The out-degree distribution of the directed graph G_2 is given below

$$\mathbb{P}(Z = 0) = \frac{1}{4}$$
$$\mathbb{P}(Z = 2) = \frac{3}{4}$$

2.1.1 Property of degree sequence

The degree sequence, $\{deg(v_i)\}_{i=1}^{|V|}$, in an undirected graph satisfies the following equality

$$\sum_{i=1}^{|V|} deg(v_i) = 2|E|$$

The in-degree and out-degree sequence, $\{deg_{in}(v_i)\}_{i=1}^{|V|}$ and $\{deg_{out}(v_i)\}_{i=1}^{|V|}$, in a directed graph satisfies the following equality

$$\sum_{i=1}^{|V|} deg_{in}(v_i) = \sum_{i=1}^{|V|} deg_{out}(v_i)$$

2.2 Paths and Cycles

A path in a graph G is a sequence of nodes $[v_1, v_2, v_3, \cdots, v_k]$, $k \geq 1$, such that $(v_j, v_{j+1}) \in E$ for $j = 1, 2, \cdots, k-1$. For example, $[v_2, v_3, v_1, v_4]$ is a path in G_1. For a directed graph, the direction of the edges needs to be taken into account while constructing a directed path. For example, $[v_3, v_2, v_4, v_1]$ is a directed path in G_2.

A simple cycle in a graph G is a sequence of nodes $[v_1, v_2, v_3, \cdots, v_k]$, $k \geq 1$, such that $(v_j, v_{j+1}) \in E$ for $j = 1, 2, \cdots, k-1$ and $v_1 = v_k$ and $v_i \neq v_j$. Therefore, a simple cycle is a path with the starting and end nodes repeated and all the interior nodes are distinct. For example,$[v_3, v_2, v_6, v_5, v_4, v_3]$ is a simple cycle in G_1. For a directed graph, the direction of the edges needs to be taken into account while constructing a directed cycle. For example, $[v_3, v_2, v_4, v_3]$ is a directed cycle in G_2.

2.2.1 Acyclic graphs

If a graph G has no cycles, then it is called an acyclic graph. If a directed graph G has no cycles, then it is called a Directed Acyclic Graph (DAG). DAG's are a important sub-class of graphs and will be studied in detail in this course.

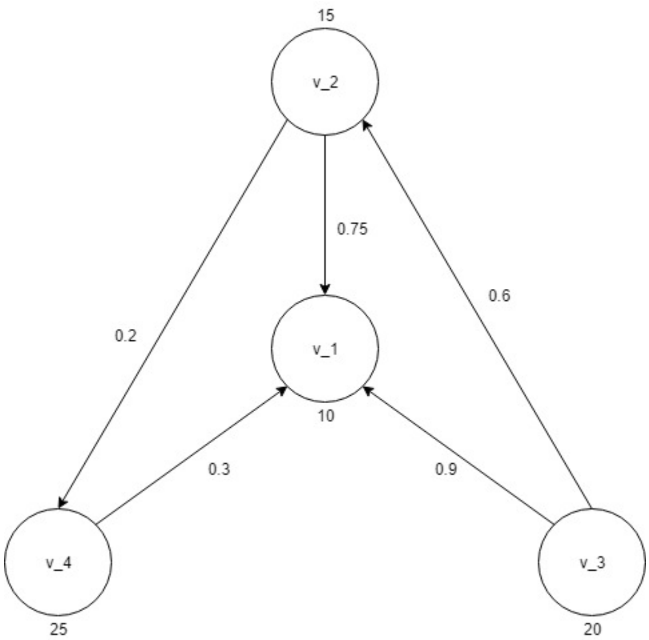

Figure 3: Directed acyclic graph (DAG)

2.3 Connectivity and Diameter

The notion of connectivity in graphs varies depending on whether the edges are directed or undirected.

2.3.1 Connectivity in undirected graphs

An undirected graph G is connected if there exists a path between any pair of vertices in G. For example, G_1 is connected because there exists a path between any pair of vertices in G_1.

An undirected graph G is disconnected if there exists at least one pair of vertices that does not have a path between them. For example, the graph shown in figure 4 is disconnected between there does not exist a path between v_1 and v_8.

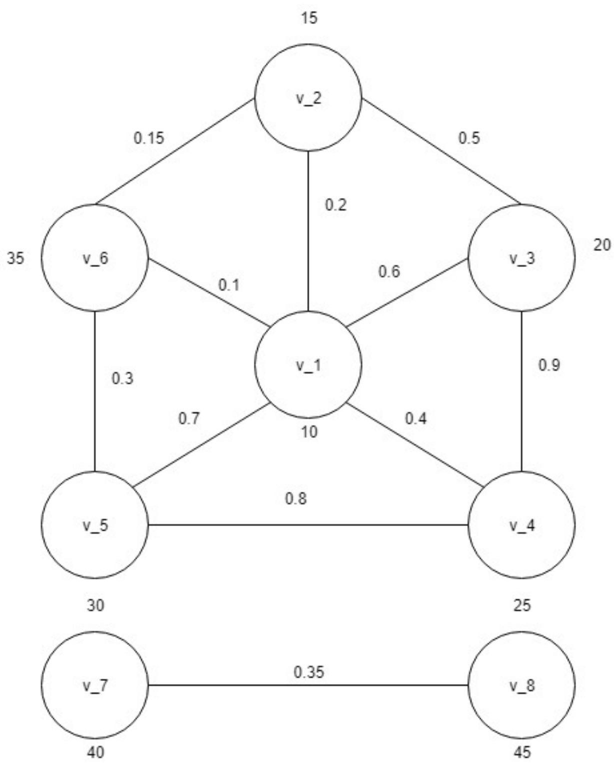

Figure 4: Disconnected graph

A disconnected graph G consists of connected components $G_1, G_2, G_3, \cdots, G_k$, where each subgraph G_i is a connected graph. For example, the disconnected graph shown in figure 4 has 2 connected components: subgraph consisting of the vertices $\{v_1, v_2, v_3, v_4, v_5, v_6\}$ and subgraph consisting of the vertices $\{v_7, v_8\}$. The connected component with the highest number of vertices is called the Giant Connected Component (GCC).

2.3.2 Connectivity in directed graphs

There are different measures of connectivity in directed graphs. In one of the measure, a directed graph is connected if the underlying undirected graph is connected. In another measure, a directed graph G is connected if for every pair of vertices (v_i, v_j) there is directed path from v_i to v_j or from v_j to v_i. The graph is said to be strongly connected if there exists a directed path from both v_i to v_j and v_j to v_i. For example, the directed graph G_2 shown in figure 2 is connected but not strongly connected since there is no directed path from v_1 to v_2.

2.3.3 Diameter

The distance between any pair of vertices (v_i, v_j) in a graph G, denoted as $d(v_i, v_j)$, is the length of the shortest path between v_i and v_j. For example, the distances between the nodes in G_1 are

$$
\begin{aligned}
d(v_1, v_2) &= 1 \\
d(v_1, v_3) &= 1 \\
d(v_1, v_4) &= 1 \\
d(v_1, v_5) &= 1 \\
d(v_1, v_6) &= 1 \\
d(v_2, v_3) &= 1 \\
d(v_2, v_4) &= 2 \\
d(v_2, v_6) &= 1 \\
d(v_3, v_4) &= 1 \\
d(v_3, v_5) &= 2 \\
d(v_3, v_6) &= 2 \\
d(v_4, v_5) &= 1 \\
d(v_4, v_6) &= 2 \\
d(v_5, v_6) &= 1
\end{aligned}
$$

Having defined the distance between the vertices, now we can define the diameter of a connected graph. The diameter of a connected graph G, denoted by $dia(G)$, is defined as the maximum distance between the vertices of the graph G. The mathematical expression for the diameter is given below

$$dia(G) = \max_{(v_i, v_j)} \quad d(v_i, v_j) \tag{1}$$

Using the above definition of the diameter, we have

$$
\begin{aligned}
dia(G_1) &= max(1, 1, 1, 1, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1) \\
dia(G_1) &= 2
\end{aligned}
$$

If the graph G is disconnected then there exists at least one pair of vertices (v_i, v_j) for which $d(v_i, v_j) = \infty$ and therefore $dia(G) = \infty$.

3 Graph representation

A graph G can be represented using a node-node incidence matrix, $A \in \mathbb{R}^{|V|\times|V|}$. The entries of A are given by

$$A_{ij} = \begin{cases} 1, (v_i, v_j) \in E \\ 0, \text{otherwise} \end{cases}$$

The node-node incidence matrix for the graph G_1 is given below

$$A_{G_1} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Since G_1 is undirected so A_{G_1} is symmetric.

The node-node incidence matrix for the directed graph G_2 is given below

$$A_{G_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Since G_2 is directed so A_{G_2} is not symmetric.

3.1 Counting paths using node-node incidence matrix

Node-Node incidence matrix multiplication can be used to count the number of paths in a graph. In order to show this, let's consider the i^{th} and j^{th} column of the incidence matrix A, denoted as a_i and a_j. Then $(A^2)_{ij}$ is given by

$$(A^2)_{ij} = a_i^T a_j = \sum_{k=1}^{n} A_{ik} A_{kj}$$

Now,

$$A_{ik} A_{kj} = \begin{cases} 1, A_{ik} = A_{kj} = 1 \\ 0, \text{otherwise} \end{cases}$$

$A_{ik} = 1$ implies $(v_i, v_k) \in E$ and $A_{kj} = 1$ implies $(v_k, v_j) \in E$. Therefore, $A_{ik}A_{kj} = 1$ implies that there is a path of length 2 between vertices v_i and v_j. Hence we have the following result,

$$(A^2)_{ij} = \text{number of paths of length 2 between } v_i \text{ and } v_j$$

Similarly, by induction we can show that

$$(A^n)_{ij} = \text{number of paths of length n between } v_i \text{ and } v_j$$

As an example, let's consider $A^2_{G_1}$

$$A^2_{G_1} = \begin{bmatrix} 5 & 2 & 2 & 2 & 2 & 2 \\ 2 & 3 & 1 & 2 & 2 & 1 \\ 2 & 1 & 3 & 1 & 2 & 2 \\ 2 & 2 & 1 & 3 & 1 & 2 \\ 2 & 2 & 2 & 1 & 3 & 1 \\ 2 & 1 & 2 & 2 & 1 & 3 \end{bmatrix}$$

We have $A^2_{G_1}(3,1) = 2$ implying that there are 2 paths of length 2 between v_3 and v_1. From the graph G_1, we can see that the 2 paths of length 2 are (v_3, v_2, v_1) and (v_3, v_4, v_1).

3.2 Logical operations on the node-node incidence matrix

Logical operations can be used to check for the existence of paths of specific length. In order to show this let's consider the logical multiplication of the node-node incidence matrix

$$(A \bigotimes A)_{ij} = \bigoplus_{k=1}^{n} (A_{ik} \bigodot A_{kj})$$

Then,

$$(A \bigotimes A)_{ij} = \begin{cases} 1, \text{if there exists at least one path of length 2 between } v_i \text{ and } v_j \\ 0, \text{otherwise} \end{cases}$$

3.2.1 Connectivity check using logical operations

We can check for the connectivity of a graph using logical operations. Let's define the reachability matrix B in the following manner

$$B = A \bigoplus A^2 \bigoplus A^3 \bigoplus A^4 \bigoplus \cdots \bigoplus A^n$$

Then,

$$B_{ij} = \begin{cases} 1, \text{if there exists at least one path of length} \leq n \text{ between } v_i \text{ and } v_j \\ 0, \text{otherwise} \end{cases}$$

If the graph G is connected then $B_{ij} = 1$ for $i \neq j$. This property is called the transitive closure of graphs and can be used to check whether a graph is connected or disconnected.