Redes Bipartitas y Sistemas de Recomendación

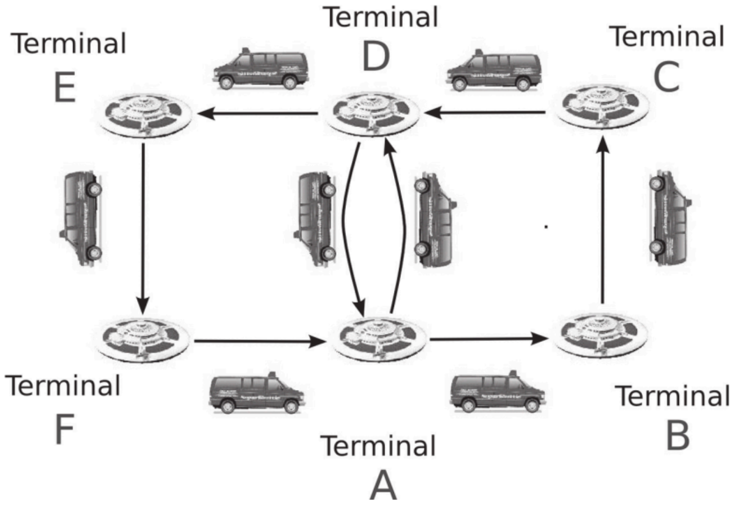
Hermilo Cortés González

July 23, 2023

Directed, Weighted and Bipartite Graphs

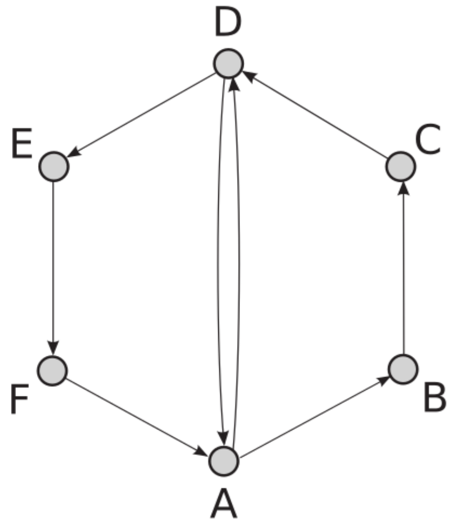
Definition (Directed graph)

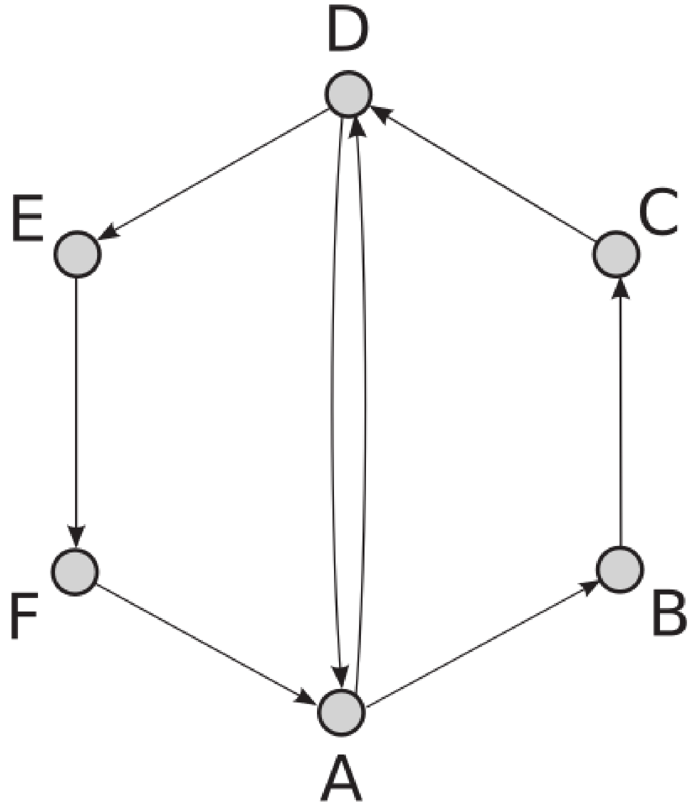
A directed graph $G \equiv (\mathcal{N}, \mathcal{L})$ consists of two sets, $\mathcal{N} \neq \emptyset$ and $\mathcal{L}$. The elements of $\mathcal{N} \equiv \{n_1, n_2, ..., n_N\}$ are the nodes of the graph G. The elements of $\mathcal{L} \equiv \{l_1, l_2, ..., l_K\}$ are distinct ordered pairs of distinct elements of $\mathcal{N}$, and are called directed links, or arcs



Example. Airport shuttle

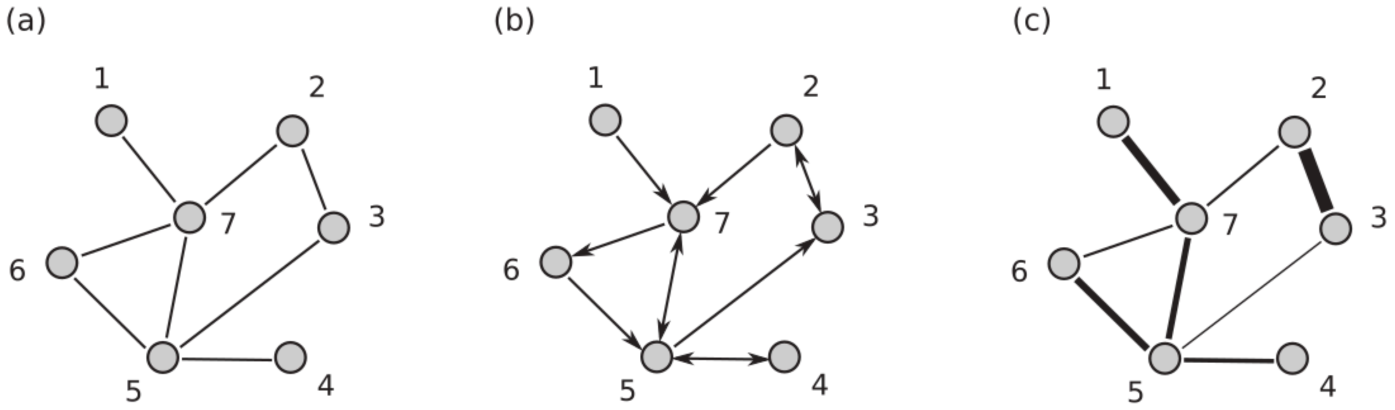
A large airport has six terminals, denoted by the letters A, B, C, D, E and F. The terminals are connected by a shuttle, which runs in a circular path, $A \rightarrow B \rightarrow C \rightarrow D \rightarrow E \rightarrow F \rightarrow A$, as shown in the figure. Since A and D are the main terminals, there are other shuttles that connect directly A with D, and vice versa. The network of connections among airport terminals can be properly described by a graph





- $N = 6$ nodes represent the terminals.
- The links indicate the presence of a shuttle connecting one terminal to another.
- In this case **it is necessary to associate a direction with each link**.
- A **directed link** is usually called an **arc**.
- The Graph has indeed K = 8 arcs.
- Notice that there can be two arcs between the same pair of nodes. For instance, arc (A, D) is different from arc (D, A).
- In a directed graph, an arc between node i and node j is denoted by the ordered pair (i, j), and we say that the link is **ingoing** in j and **outgoing** from i.
- Such an arc may still be denoted as l_{ij}.
- However, at variance with undirected graphs, this time **the order of the two nodes is important**.
- Namely, $l_{ij} \equiv (i, j)$ stands for an arc from i to j, and $l_{ij} \neq l_{ji}$, or in other terms the arc (i, j) is different from the arc (j, i).

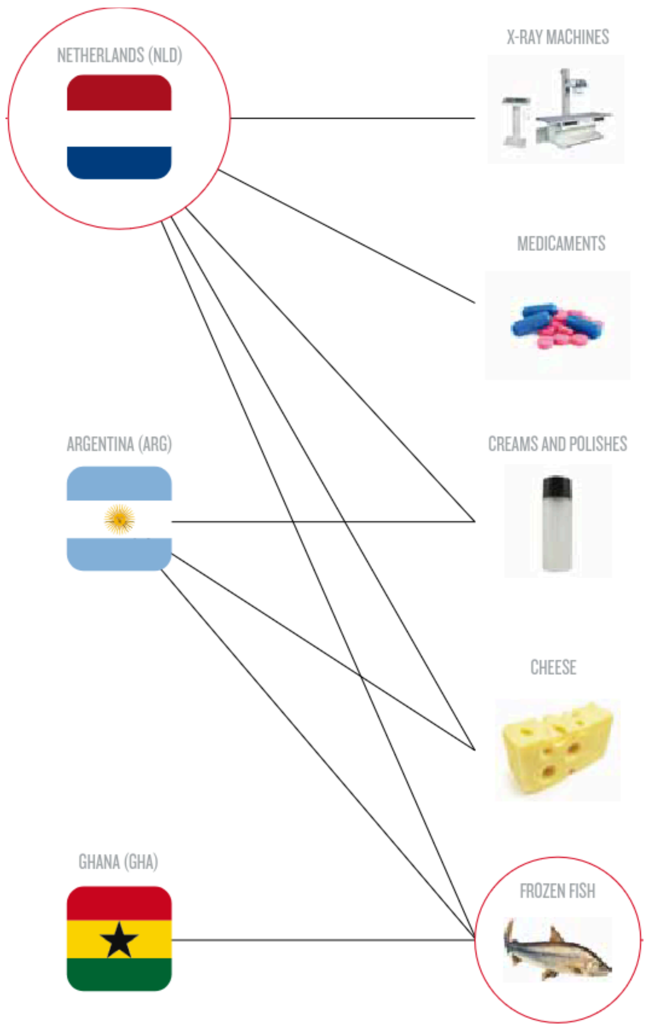
- The most basic definition is that of **undirected graph**, which describes systems in which **the links have no directionality**.
- In the case, instead, in which **the directionality of the connections is important, the directed graph definition is more appropriate**.
- Also, we often need to deal with networks displaying a **large heterogeneity in the relevance of the connections**.
- Typical examples are social systems where it is possible to measure the strength of the interactions between individuals.



An undirected (a), a directed (b), and a weighted undirected (c) graph with $N = 7$ nodes. In the directed graph, adjacent nodes are connected by arrows, indicating the direction of each arc. In the weighted graph, the links with different weights are represented by lines with thickness proportional to the weight.

Bipartite Graphs

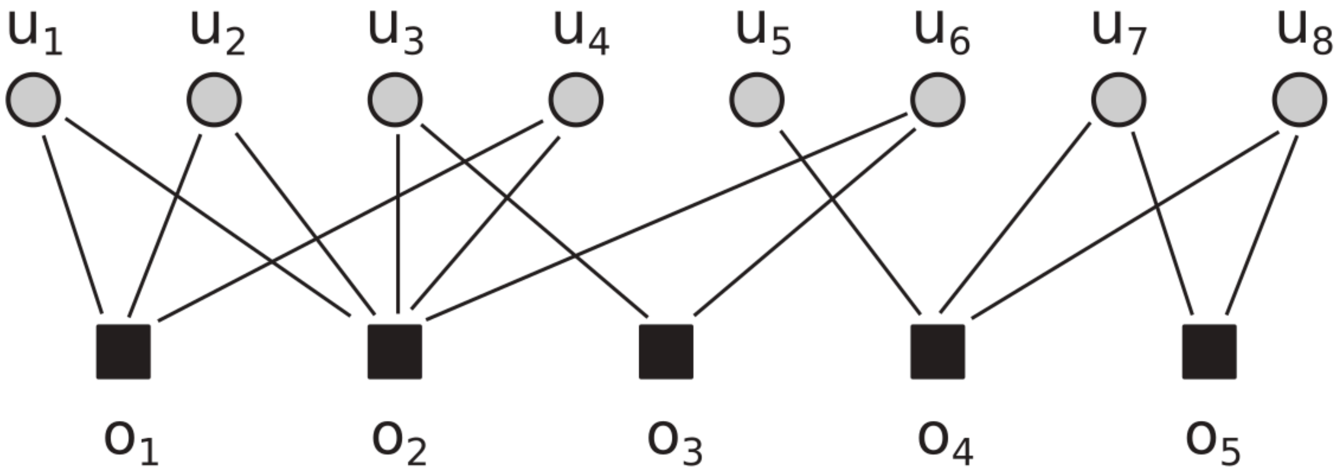
Bipartite graph is a graph whose nodes can be divided into two disjoint sets, such that every edge connects a vertex in one set to a vertex in the other set, while there are no links connecting two nodes in the same set.



Definition (Bipartite graph)

A bipartite graph, $G \equiv (\mathcal{N}, \mathcal{V}, \mathcal{L})$, consists of three sets, $\mathcal{N} \neq \emptyset, \mathcal{V} \neq \emptyset$ and $\mathcal{L}$. The elements of $\mathcal{N} \equiv \{n_1, n_2, ..., n_N\}$ and $\mathcal{V} \equiv \{v_1, v_2, ..., v_V\}$ are distinct and are called the nodes of the bipartite graph. The elements of $\mathcal{L} \equiv \{l_1, l_2, ..., l_K\}$ are distinct unordered pairs of elements, one from $\mathcal{N}$ and one from $\mathcal{V}$, and are called links or edges.

- Many real systems are naturally bipartite. For instance, typical bipartite networks are **systems of users purchasing items such as books, or watching movies**.
- An example is shown in the next figure, where we have denoted the user-set as $U = \{u_1, u_2, \ldots, u_N\}$ and the object-set as $O = \{o_1, o_2, \ldots, o_V\}$
- In such a case we have indeed only **links between users and items, where a link indicates that the user has chosen that item**.



- Starting from a bipartite network, we can derive at least two other graphs.
- The first graph is **a projection of the bipartite graph on the first set of nodes**: the nodes are the users and two users are linked if they have at least one object in common.
- We can also assign a weight to the link **equal to the number of objects in common**.
- The weight can be interpreted as a **similarity between the two users**.

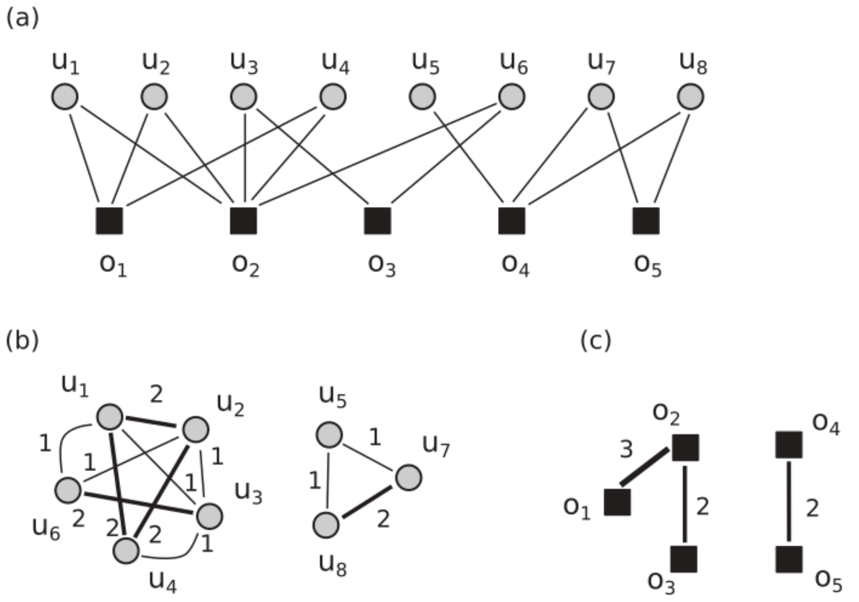
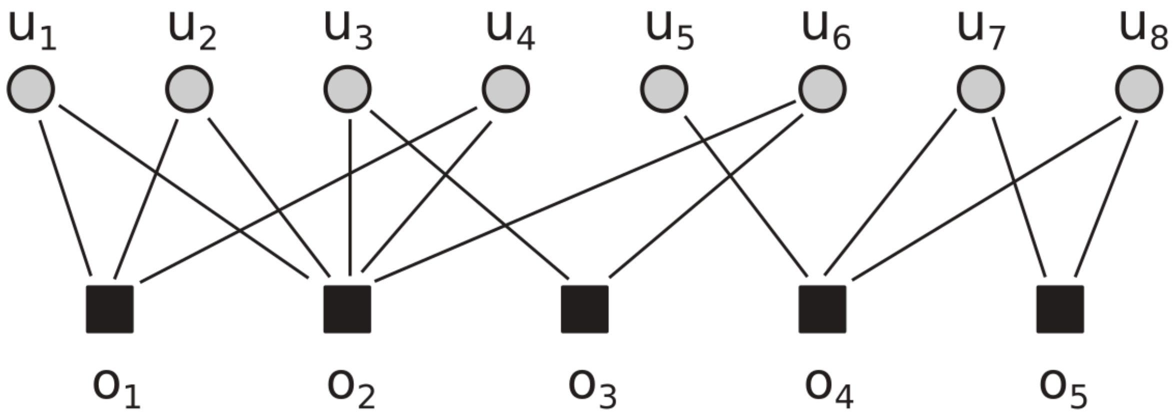


Illustration of a bipartite network of $N = 8$ users and $V = 5$ objects (a), as well as its user-projection (b) and object-projection (c). The link weights in (b) and (c) are set as the numbers of common objects and users, respectively.

Recommendation Systems

- Consider a system of users buying books or selecting other items, similar to the Figure



- Based on this, it is possible to construct **recommendation systems**, i.e. to predict the user's opinion on those objects not yet collected, and eventually to recommend some of them.
- The simplest recommendation system, known as **global ranking method (GRM)**, sorts all the objects in descending order of degree and recommends those with the highest degrees.
- Such a recommendation is based on the assumption that the most-selected items are the most interesting for the **average user**.
- A more refined recommendation algorithm, known as **collaborative filtering (CF)**, is based on similarities between users

How to Represent a Graph

- There are different ways to completely describe a graph $G = (\mathcal{N}, \mathcal{L})$ with N nodes and K links by means of a matrix.
- One possibility is to use the so-called **adjacency matrix** A.

Definition (Adjacency matrix)

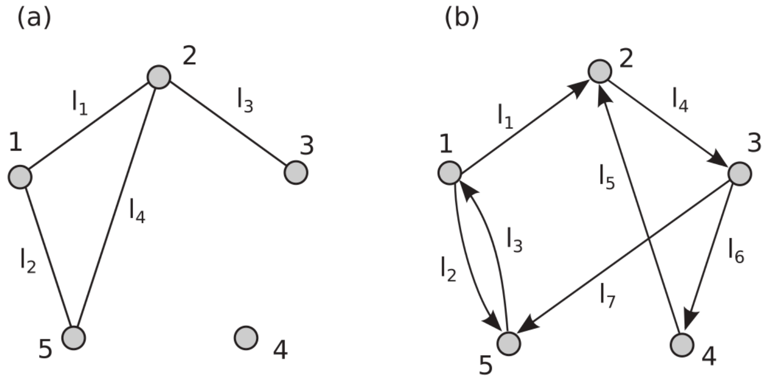
The adjacency matrix A of a graph is a $N \times N$ square matrix whose entries a_{ij} are either ones or zeros according to the following rule:

$$a_{ij} = \begin{cases} 1 & \text{iff } (i, j) \in \mathcal{L} \\ 0 & \text{otherwise} \end{cases}$$

- In practice, for an undirected graph, entries a_{ij} and a_{ji} are set equal to 1 if there exists the edge (i, j), while they are zero otherwise.
- Thus, in this case, the adjacency matrix is symmetric.
- If instead the graph is directed, $a_{ij} = 1$ if there exists an arc from i to j.
- Notice that in both cases it is common convention to set $a_{ii} = 0, \forall i = 1, ..., N$.

Example

Consider the two graphs in the figure below. The first graph is undirected and has $K = 4$ links, while the second graph is directed and has $K = 7$ arcs.



The adjacency matrices associated with the two graphs are respectively:

$$A_u = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad A_d = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Example

$$A_u = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

$$A_d = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- A_u is symmetric and contains $2K$ non-zero entries.
- The number of ones in row i, or equivalently in column i, is equal to the degree of vertex i.
- The adjacency matrix of a directed graph is in general not symmetric. This is the case of A_d.
- This matrix has K elements different from zero, and the number of ones in row i is equal to the number of outgoing links k_i^{out}, while the number of ones in column i is equal to the number of ingoing links k_i^{in}.

- A **Bipartite Graph** can be represent by means of a slightly different definition of the adjacency matrix than that given above, and which in general is not a square matrix.
- It can be described by an $N \times V$ adjacency matrix A, such that entry $a_{i\alpha}$, with $i = 1, \ldots, N$ and $\alpha = 1, \ldots, V$, is equal to 1 if node i of the first set and node α of the second set are connected, while it is 0 otherwise.
- Using this representation of a bipartite graph in terms of an adjacency matrix, we show in the next example **how to formally describe a commonly used method to recommend a set of objects to a set of users**.

Recommendation systems: collaborative filtering

- Consider a bipartite graph of users and objects such as that shown in Figure 1.8, in which the existence of the link between node i and node α denotes that user u_i has selected object o_α.
- A famous personalised recommendation system, known as **collaborative filtering** (CF), is based on the construction of a $N \times N$ **user similarity matrix** $S = \{s_{ij}\}$.
- The similarity between two users u_i and u_j can be expressed in terms of the adjacency matrix of the graph as:

$$s_{ij} = \frac{\sum_{\alpha=1}^{V} a_{i\alpha} a_{j\alpha}}{\min\left\{k_{u_i}, k_{u_j}\right\}},$$

- Where $k_{u_i} = \sum_{\alpha=1}^{V} a_{i\alpha}$ is the degree of user u_i i.e. the number of objects chosen by u_i[324].
- Based on the similarity matrix S, we can then construct an $N \times V$ **recommendation matrix** $R = \{r_{ia}\}$.
- In fact, for any user–object pair u_i, o_α, if u_i has not yet chosen o_α, i.e. if $\alpha_{i\alpha} = 0$, , we can define a recommendation score $r_{i\alpha}$ measuring to what extent u_i may like o_α, as :

$$r_{i\alpha} = \frac{\sum_{j=1, j \neq i}^{N} s_{ij} a_{j\alpha}}{\sum_{j=1, j \neq i}^{N} s_{ij}}.$$