# Complex Networks Master of Science in Electrical Engineering

Erivelton Geraldo Nepomuceno

Department of Electrical Engineering Federal University of São João del-Rei

March 13, 2019

## 1. Graphs and Graph Theory

- Graphs are the mathematical objects used to represent networks, and graph theory is the branch of mathematics that deals with the study of graphs.
- Introduce in 1763 by Euler, to settle a famous unsolved problem of his time: the so-called Königsberg bridge problem. [Video-03].
- Subsequent work in graph theory by Kirchhoff and Cayley had its root in the physical world.
- Kirchhoff's investigations into electric circuits led to his development of a set of basic concepts and theorems concerning trees in graphs.

# 1.1 What Is a Graph?

 Framework for the exact mathematical treatment of a complex network is a branch of discrete mathematics known as graph theory.

#### Example 1

Seven people have been invited to a party. Their names are Adam, Betty, Cindy, David, Elizabeth, Fred and George. Before meeting at the party, Adam knew Betty, David and Fred; Cindy knew Betty, David, Elizabeth and George; David knew Betty (and, of course, Adam and Cindy); Fred knew Betty (and, of course, Adam).



Figure 1: Example 1.1 (Friends at a party)

#### Example 2

The map in the figure shows 23 of Europe's approximately 50 countries. Each country is shown with a different shade of grey, so that from the image we can easily distinguish the borders between any two nations.



#### Figure 2: Example 1.2 (The map of Europe)

Prof. Erivelton (UFSJ)

Complex Networks

#### Definition 1 (Undirected graph)

A graph, more specifically an undirected graph,  $G \equiv (\mathcal{N}, \mathcal{L})$ , consists of two sets,  $\mathcal{N} \neq \emptyset$  and  $\mathcal{L}$  The elements of  $\mathcal{N} \equiv \{n_1, n_2, \dots, n_N\}$  are distinct and are called the nodes (or vertices, or points) of the graph *G*. The elements of  $\mathcal{L} \equiv \{l_1, l_2, \dots, l_K\}$  are distinct unordered pairs of distinct elements of  $\mathcal{N}$ , and are called links (or edges, or lines).

- The number of vertices N ≡ N[G] = |N|, where the symbol | · | denotes the cardinality of a set, is usually referred as the order of G, while the number of edges K ≡ K[G] = |L| is the size of G.
- In an undirected graph, each of the links is defined by a pair of nodes, *i* and *j*, and is denoted as (*i*, *j*) or (*j*, *i*).
- Two nodes joined by a link are referred to as adjacent or neighbouring.



Figure 3: Some examples of undirected graphs, namely a tree,  $G_1$ ; two graphs containing cycles,  $G_2$  and  $G_3$ ; and an undirected multigraph,  $G_4$ .

Prof. Erivelton (UFSJ)

Complex Networks



- Graph  $G_1$  is made of N = 5 nodes and K = 4 edges.
- Any pair of nodes of this graph can be connected in only one way: tree.



- Graph  $G_2$  has N = K = 4.
- By starting from one node, say node
  1, one can go to all the other nodes 2,
  3, 4, and back again to 1.
- G<sub>2</sub> contains a cycle.



- Graph  $G_2$  has N = 4 and K = 3.
- Graph *G*<sub>3</sub> contains an isolated node and three nodes connected by three links.
- *G*<sub>3</sub> is not connected.

10



- Graph *G*<sub>4</sub> has loops and multiple edges.
- $G_4$  is a multigraph.
- Node 1 is connected to itself by a loop, and it is connected to node 3 by two links.

## Density

- For a graph *G* of order *N*, the number of edges *K* is at least 0, in which case the graph is formed by *N* isolated nodes, and at mostN(N 1)/2, when all the nodes are pairwise adjacent.
- The ratio between the actual number of edges K and its maximum possible number N(N-1)/2 is known as the density of G.
- Empty graph: density equals 0.
- Complete graph: denoted as  $\mathbb{K}_N$  and its density is 1.



Figure 4: Complete graphs respectively with three, four and five nodes.  $\mathbb{K}_3$  is called triangle.

#### Box 1 (Graph Drawing)

A good drawing should highlight the properties of a graph. Example: circular layout (left) and a spring-based layout (right) based on the Kamada–Kawai algorithm



Software packages: Pajek (http://mrvar.fdv.uni-lj.si/pajek/), Gephi (https://gephi.org/) and GraphViz (http://www.graphviz.org/), NetworkX (https://networkx.github.io/), iGraph (http://igraph.org/) and SNAP (http://snap.stanford.edu/).

Prof. Erivelton (UFSJ)

# Comparison of graphs

- How to compare graphs with the same order and size?
- Two graphs  $G1 = (\mathcal{N}_1, \mathcal{L}_1)$  and  $G_2 = (\mathcal{N}_2, \mathcal{L}_2)$  are the same graph if  $\mathcal{N}_1 = \mathcal{N}_2$  and  $\mathcal{L}_1 = \mathcal{L}_2$ .
- In this case  $G_1 = G_2$ .



Figure 5: Graphs (a) and (b) are the same graph, since their edges are the same. Graphs (b) and (c) are isomorphic, since there is a bijection between the nodes that preserves the edge set

Prof. Erivelton (UFSJ)

## Isomorphism

#### Definition 2 (Isomorphism)

Two graphs,  $G_1 = (\mathcal{N}_1, \mathcal{L}_1)$  and  $G_2 = (\mathcal{N}_2, \mathcal{L}_2)$ , of the same order and size, are said to be isomorphic if there exists a bijection  $\phi : \mathcal{N}_1 \to \mathcal{N}_2$ , such that  $(u, v) \in \mathcal{L}_1 iff(\phi(u), \phi(v)) \in \mathcal{L}_2$ . The bijection  $\phi$  is called an isomorphism.

- The adjacency is preserved.
- In this case we write  $G_1 \simeq G_2$ .



Figure 6: Two unlabelled graphs, namely the cycle  $\mathbb{C}_4$  and the star graph  $\mathbb{S}_4$ , and one possible labelling of such two graphs.

- It is difficult to check whether two unlabelled graphs are isomorphic, because there are N! possible ways to label the N nodes of a graph.
- This is known as the isomorphism problem.
- There is no algorithms to check if two generic graphs are isomorphic in polynomial time.

#### Definition 3 (Automorphism)

Given a graph  $G = (\mathcal{N}, \mathcal{L})$ , an automorphism of G is a permutation  $\phi : \mathcal{N} \to \mathcal{N}$  of the vertices of G so that if  $(u, v) \in \mathcal{L}$  then  $(\phi(u), \phi(v)) \in \mathcal{L}$ . The number of different automorphisms of G is denoted as  $a_G$ .



Figure 7: All possible automorphisms of graph with  $\mathbb{C}_4$ .

Prof. Erivelton (UFSJ)

Complex Networks

#### Definition 4 (Subgraph)

A subgraph of  $G = (\mathcal{N}, \mathcal{L})$  is a graph  $G' = (\mathcal{N}', \mathcal{L}')$  such that  $\mathcal{N}' \subseteq \mathcal{N}'$ and  $\mathcal{L}' \subseteq \mathcal{L}'$  If G' contains all links of G that join two nodes in  $\mathcal{N}'$ , then G' is said to be the subgraph induced or generated by  $\mathcal{N}'$ , and is denoted as  $G = G[\mathcal{N}']$ .



Figure 8: A graph *G* with N = 6 nodes and three subgraphs of *G*. (c) represents the graph  $G_6$ , induced by the neighbours of node 6. It is called a subgraph of the neighbours.

## 1.2 Directed, Weighted and Bipartite Graphs



Figure 9: Example 1.4 (Airport shuttle) A large airport has six terminals, denoted by the letters A, B, C, D, E and F.

#### Definition 5 (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  $\mathcal{N} \equiv \{n_1, n_2, \dots, n_N\}$  are the nodes of the graph *G*. The elements of  $\mathcal{L} \equiv \{l_1, l_2, \dots, l_K\}$  are distinct ordered pairs of distinct elements of  $\mathcal{N}$ , and are called directed links, or arcs.

- 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*.
- An arc is denoted as *l<sub>ij</sub>*.
- However, we may have  $I_{ij} \neq I_{ji}$ .

## Data Set 1



Figure 10: Elisa's kindergarten network describes N = 16 children between three and five years old, and their declared friendship relations. The network given in this data set is a directed graph with K = 57 arcs and is shown in the figure. A graph reciprocity is  $r = 34/57 \approx 0.6$ .

Prof. Erivelton (UFSJ)



Figure 11: 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.

#### Definition 6 (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  $\mathcal{N} \equiv \{n_1, n_2, \dots, n_N\}$  and  $\mathcal{V} \equiv \{v_1, v_2, \dots, v_N\}$ are distinct and are called the nodes of the bipartite graph. The elements of  $\mathcal{L} \equiv \{l_1, l_2, \dots, l_K\}$  are distinct unordered pairs of elements, one from  $\mathcal{N}$  and one from  $\mathcal{V}$ , and are called links or edges. (a)



Figure 12: 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 the number of common objects and users.

Prof. Erivelton (UFSJ)

Complex Networks

# **1.3 Basic Definitions**

### Definition 7 (Node degree)

The degree  $k_i$  of a node *i* is the number of edges incident in the node. If the graph is directed, the degree of the node has two components: the number of outgoing links  $k_i^{out}$ , referred to as the out-degree of the node, and the number of ingoing links  $k_i^{in}$ , referred to as the in-degree of node *i*. The total degree of the node is then defined as  $k_i = k_i^{out} + k_i^{in}$ .

- $\{k_1, k_2, \ldots, k_N\}$  is called the degree sequence.
- The average degree  $\langle k \rangle$  of a graph is defined as

$$\langle k \rangle = rac{1}{N} \sum_{i=1}^{N} k_i$$

and is equal to  $\langle k \rangle = 2K/N$ .

• Example: Compute the node degrees in Elisa's kindergarten.

#### Definition 8 (Walks, trails, paths and geodesics)

A walk W(x, y) from node x to node y is an alternating sequence of nodes and edges (or arcs)  $W = (x \equiv n_0, e_1, n_1, e_2, ..., e_l, n_l \equiv y)$  that begins with x and ends with y, such that  $e_i = (n_{i-1}, n_i)$  for i = 1, 2, ..., l. Usually a walk is indicated by giving only the sequence of traversed nodes:  $W = (x \equiv n_0, n_1, ..., n_l \equiv y)$ . The length of the walk,  $l = \ell(W)$ , is defined as the number of edges (arcs) in the sequence. A trail is a walk in which no edge (arc) is repeated. A path is a walk in which no node is visited more than once. A shortest path (or geodesic) from node x to node y is a walk of minimal length from x to y, and in the following will be denoted as P(x, y).

#### Definition 9 (Graph distances)

In an undirected graph the distance between two nodes x and y is equal to the length of a shortest path P(x, y) connecting x and y. In a directed graph the distance from x to y is equal to the length of a shortest path P(x, y) from x to y.

#### Definition 10 (Circuits and cycles)

A circuit is a closed trail, i.e. a trail whose end vertices coincide. A cycle is a closed walk, of at least three edges (or arcs)  $W = (n_0, n_1, ..., n_l), l \ge 3$ , with  $n_0 = n_l$  and  $n_i, 0 < i < l$ , distinct from each other and from  $n_0$ . An undirected cycle of length *k* is usually said a *k*-cycle and is denoted as  $C_k$ .  $C_3$  is a triangle ( $C_3 = \mathcal{K}_3$ ),  $C_4$  is called a quadrilater,  $C_5$  a pentagon, and so on. • Example 1.7



- Find a walk of length 5 from node 5 back to node 5.
- Find a trail of length 7 starting from node 5.
- Find path from node 5 to node 2.
- Find the geodesic from node 5 to node 2.

• Example 1.7



- Find a walk of length 5 from node 5 back to node 5.
  - Answer: (5, 6, 4, 2, 4, 5)
- Find a trail of length 7 starting from node 5.
  - Answer: (5, 6, 4, 5, 1, 2, 4)
- Find path from node 5 to node 2.

▶ Answer: (5, 4, 3, 2)

• Find the geodesic from node 5 to node 2.

Answer: (5, 1, 2), (5, 6, 2), (5, 4, 2)

• Example 1.7



- Find a walk of length 5 from node 5 back to node 5.
  - Answer: (5, 6, 4, 2, 4, 5)
- Find a trail of length 7 starting from node 5.
  - Answer: (5, 6, 4, 5, 1, 2, 4)
- Find path from node 5 to node 2.
  - Answer: (5, 4, 3, 2)
- Find the geodesic from node 5 to node 2.
  - Answer: (5, 1, 2), (5, 6, 2), (5, 4, 2)

# Definition 11 (Connectedness and components in undirected graphs)

Two nodes *i* and *j* of an undirected graph *G* are said to be connected if there exists a path between *i* and *j*. *G* is said to be connected if all pairs of nodes are connected; otherwise it is said to be unconnected or disconnected. A component of *G* associated with node *i* is the maximal connected induced subgraph containing *i*, i.e. it is the subgraph which is induced by all nodes which are connected to node *i*.

#### Box 2 (Path-Finding Behaviours in Animals)

Finding the shortest route is extremely important also for animals moving regularly between different points. How can animals, with only limited local information, achieve this? Ants, for instance, find the shortest path between their nest and their food source by communicating with each othervia their pheromone, a chemical substance that attracts other ants. Initially, ants explore all the possible paths to the food source. Ants taking shorter paths will take a shorter time to arrive at the food. This causes the quantity of pheromone on the shorter paths to grow faster than on the longer ones, and therefore the probability with which any single ant chooses the path to follow is quickly biased towards the shorter ones. The final result is that, due to the social cooperative behaviour of the individuals, very quickly all ants will choose the shortest path.

Definition 12 (Connectedness and components in directed graphs)

Two nodes *i* and *j* of a directed graph *G* are said to be strongly connected if there exists a path from *i* to *j* and a path from *j* to *i*. A directed graph *G* is said to be strongly connected if all pairs of nodes (i, j) are strongly connected.

- Do with two different examples the Matlab codings (Introduction -Slides 17 to 20.). Deadline: March 21.
- Read and prepare individually a ten-minute talk about a free-choice paper. Please, follow the recommendations:
  - Search for papers with "complex networks" AND preferred subject.
  - Do not worry to understand the whole paper.
  - Talk should be uploaded by March 27.
  - Talk and discussion will happen on March 28.
- Prepare a talk to explain one of the packages (Box 1). Do it in pairs. Use the Dataset 1 to give some examples. Deadline: April 3 (upload), April 4 (talk).