Complex Networks

Theory and Applications

Guido CALDARELLI

[1] IMT Alti Studi Lucca, Italy

http://www.guidocaldarelli.com/index.php/lectures

What

What exactly am I going to learn?

Goal of this course is to make you expert of complex networks
Goal of this lecture is to introduce you the first basic concepts

- What is a network
- What is the degree of a vertex
- What is the distance in a graph, what is the diameter
- What is assortativity

Definitions

Graph

A Graph is a mathematical object composed by vertices connected by edges

Networks

We use the term Network (Web) for any real realization of a Graph

Adjacency Matrix

The structure of any graph can be represented by means of a matrix.

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Two simple graphs and their adjacency matrices. Note that for a directed graph (right) the matrix is not symmetric (i.e. it changes if we swap the rows with columns).

Some definitions

- Density is the number of lines in a simple network, expressed as a proportion of the maximum possible number of lines.
- A complete network is a network with maximum density.
- Degree of a vertex is the number of lines incident with it.
- Two vertices are adjacent if they are connected by a line.
- The indegree of a vertex is the number of arcs it receives.
- The outdegree is the number of arcs it sends.

Weighted Graphs

In the above definitions all the edges count the same. This is not the case for weighted graphs.

Weighted Graphs

Edges can have a weight attached

The elements of the adjacency matrix are then real numbers

Cycle and Trees

Two vertices can be connected in different ways.

Cycle

a cycle is a closed path that visits a set of vertices only once (apart from the end-vertices that coincide)

From that quantity we have now that

Tree

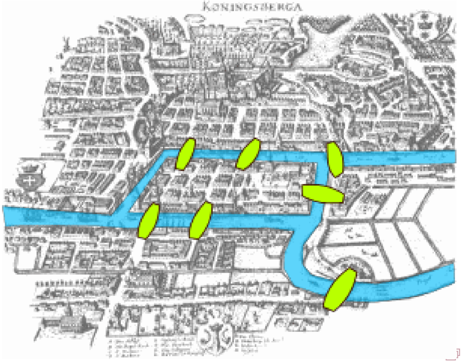
- a set of vertices connected to each other without cycles is a tree;
- a set of disconnected trees is a forest.

The degree

Degree

The degree is the number of edges owned by a vertex

Despite the simplicity is one powerful quantity to describe a network. In case of arrows we have **in/out** degree
The most famous example is the Königsberg bridge problem

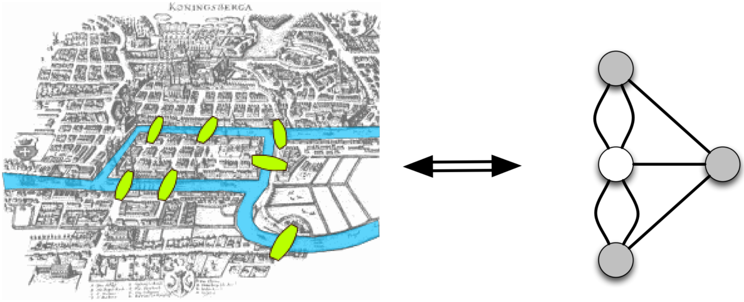


Bridging

The problem was to find a walk through the city that would cross each bridge once and only once. Euler proved that the problem has no solution.

Parity of degree

The solution comes when considering degree



Solution by degree

- introduce graphs as a model of the problem
- check the degrees
- whenever one enters a vertex by an edge, one leaves the vertex by another edge.
- passage points must have even degree
- odd degree only allows for starting and ending points!!!!
- 4 vertices have odd degree
- no path is possible

Degree from Adjacency Matrix

Degree

The degree k_i of a vertex i can be computed as $k_i = \sum_{j=1,n} a_{ij}$

Oriented Degree

$$k_i^{in} = \sum_{j=1,n} a_{ji}$$

$$k_i^{out} = \sum_{j=1,n} a_{ij}$$

Weigted Degree

The weighted degree k_i^w of a vertex i is then defined as

Strength

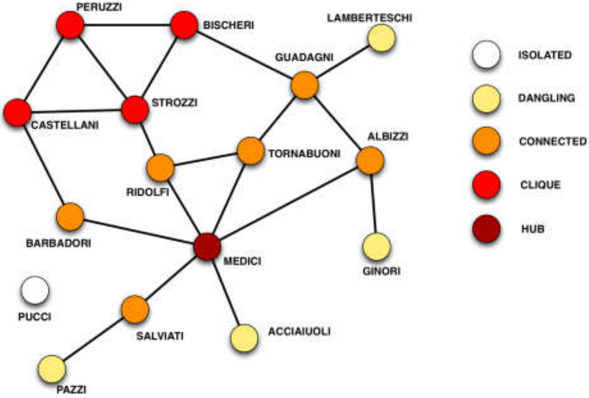
$$k_i^w = \sum_{j=1,n} a_{ij}^w$$

Several studies show that the weighted degree is empirically related to the 'topological' degree k_i by means of a simple relation

$$k_i^w \propto (k_i)^\zeta$$

Medici Family

Graphs allow to visualize many situations where **degree** clarifies the role of vertices
Consider the relationships of Florence families in the Renaissance



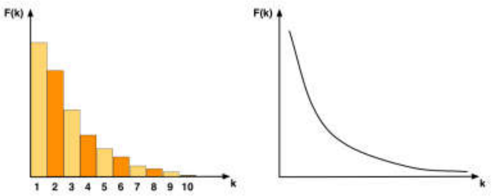
Statistics

Seldom, we have to deal with such small graphs.

Statistical Properties

When dealing with real world networks we have often systems by hundreds of thousand of nodes.

The only possibility is to pass to a Statistical Descritpion. We plot the frequency distribution of the values of the degrees. For large system we treat it as a continuous function and use as a Probability Density.



Real World Data

Form of the Distribution

One striking feature is that despite the kind of system, the form of the distribution is qualitatively similar. It has the form of a power law $P(k) \simeq k^{-\gamma}$

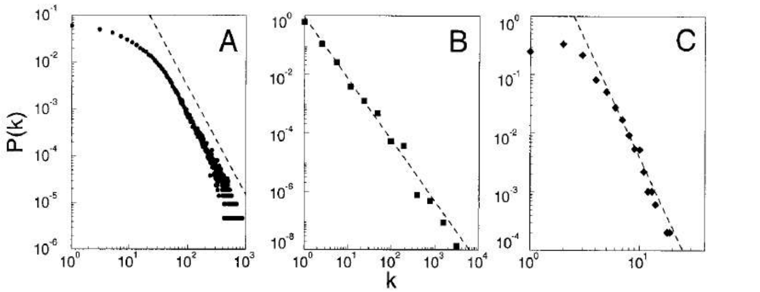
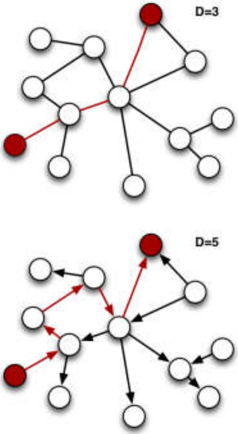


Fig. 1. The distribution function of connectivities for various large networks. **(A)** Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. **(B)** WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). **(C)** Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

The distance

Distance

Distance between two vertices is the minimum number of edges between them



Graphs can be oriented. As we all know, distances in this case can be longer

Topological and Metrical Distances

A big difference

Distance between vertices does not depend on how they are drawn

METRICAL

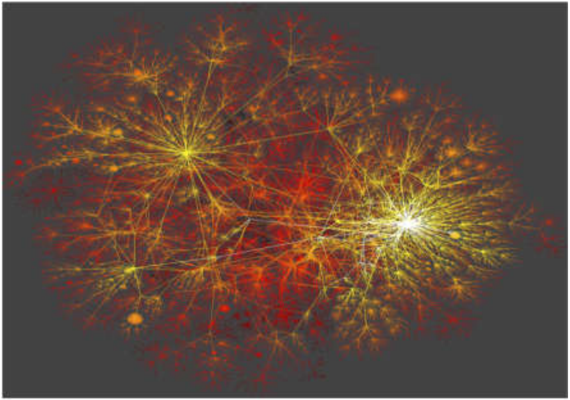
- Cities can be thousands of kilometers far away
- Space is not arranged hierarchically
- Distance grows as the number of vertices

TOPOLOGICAL

- Vertices are often only few steps away
- We can often organize vertices in hierarchies
- Distance grows as the logarithm of # vertices

Statistics

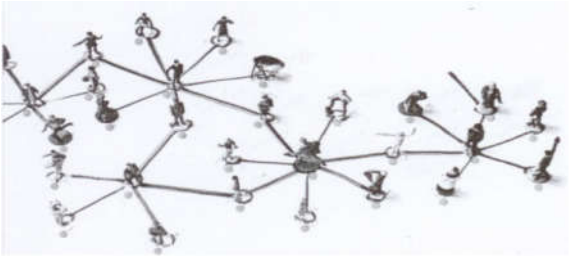
Again we often deal with million of sites in real situations



The fact that Internet has server few steps away makes possible routing of e-mails

Small world effect I

The fact that few steps are usually enough to span one system is known as **Small World Effect**



- The first evidence comes from S. Milgram
- Experiment to analyze the structure of society in USA
- Letters used as a probe of acquaintances

Small world effect II

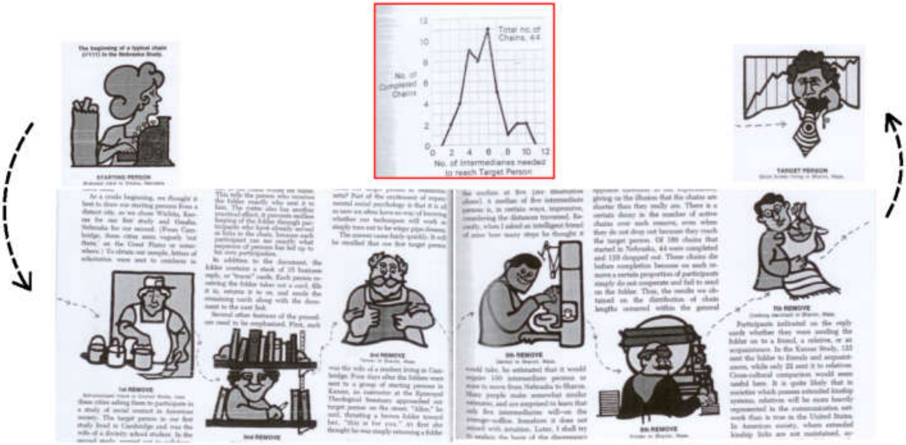
most important rule is: *"If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder . . . to a personal acquaintance who is more likely than you to know the target person . . . it must be someone you know on a first-name basis."* This rule sets the document into motion, moving it from one participant to the next, until it is sent to someone who knows the target person.

3. A roster on which each person in the chain writes his name. This tells the person who receives the folder exactly who sent it to him. The roster also has another practical effect; it prevents endless looping of the folder through participants who have already served as links in the chain, because each participant can see exactly what sequence of persons has led up to his own participation.

In addition to the document, the folder contains a stack of 15 business reply, or "tracer" cards. Each person receiving the folder takes out a card, fills it in, returns it to us, and sends the remaining cards along with the document to the next link.

- Select a target (stockbrocker) in Boston (Massachussets)
- Take random persons in Omaha (Nebraska)
- Give to these persons name and address of target
- Tell them to give the target a folder only if they know him on a personal basis
- If not: pass the folder to someone else who may know him

Small world effect III



Small world effect IV

Small World

Most of the persons are just 6 steps away.

The above results is known as *Six Degrees of Separation*

- The experiment has been recently repeated (www)
- The results confirmed have two explanations: presence of shortcuts and presence of hierarchies
- By arranging in hierarchies the various nodes, distance grows as the logarithm of the number of nodes
- in the above case by passing from 1000 to 1,000,000 distance passes from 3 to 6.