Lectures on Complex Networks

G. Caldarelli
CNR-INFM Centre SMC Dep. Physics University “Sapienza” Rome, Italy
http://www.guidocaldarelli.com
From Technological Systems to Social and Biological ones, similar structures appear.
1.5 The Properties of Complex Networks

- Scale Invariant Degree Distribution $P(k)$
- Distribution of distances $P(l)$ peaked around small values
- Clustering (with respect to random connections)
- Assortativity
1.5 The Properties of Complex Networks

• Scale Invariant Degree Distribution $P(k)$

Lectures on Complex Networks
• Distribution of distances $p(l)$ peaked around small values

1.5 The Properties of Complex Networks

- Clustering (with respect to random connections)

1.5 The Properties of Complex Networks

- Assortativity $C(k) \sim k^{-\phi}$ or average degree of neighbours $<K_{nn}>$

**a.** In assortative networks, well-connected nodes tend to join to other well-connected nodes, as in many social networks — here illustrated by friendship links in a school in the United States. **b.** In disassortative networks, by contrast, well-connected nodes join to a much larger number of less-well-connected nodes. This is typical of biological networks; depicted here is the web of interactions between proteins in brewer's yeast, *Saccharomyces cerevisiae*. Clauset and colleagues' hierarchical random graphs provide an easy way to categorize such networks.

A Graph is an object composed by vertices and edges

- Degree $k$ (In-degree $k_{in}$ and out-degree $k_{out}$) = number of edges (oriented) per vertex
- Distance $d$ = number of edges amongst two vertices (in the connected region !)
- Diameter $D$ = Maximum of the distances (in the connected region !)
1.1 Definition and Adjacency Matrix

\[ A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix} \]

Fig. 1.3 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).
1.1.2 Basic Quantities (degree)

Through the adjacency matrix we can write

**DEGREE**

\[ k_i = \sum_{j=1}^{n} a_{ij} \]

**WEIGHTED DEGREE (Strength)**

\[ k_i^w = \sum_{j=1}^{n} a_{ij}^w \]
One way to visualize the behaviour of the degree in a network (especially for large ones) is to check the behaviour of the degree frequency distribution $P(k)$.
1.1.3 Basic Quantities (distance)

Through the adjacency matrix we can write

\[ d_{ij} = \min \left\{ \sum_{k,l \in P_{ij}} a_{kl} \right\} \]

WEIGHTED DISTANCE \[ d_{ij} = \min \left\{ \sum_{k,l \in P_{ij}} a_{kl}^w \right\} \]

\[ d_{ij} = \min \left\{ \sum_{k,l \in P_{ij}} \frac{1}{a_{kl}} \right\} \]

DISTANCE \( \propto \) SUM OF WEIGHTS

DISTANCE \( \propto \) INVERSE OF WEIGHTS
Similarly to the degree, one usually plots the histogram of the frequency density $P(d)$ of distances $d$.
Fig. 1.5 The clustering coefficient of the central vertex is 1/3. This is because its degree is three and its neighbours can be connected each other in three different ways. Of these possibilities (dashed line) only one is actually realized (solid line) and therefore $C_i = 1/3$. The three connected vertices form the coloured triangle. For that reason, sometimes the clustering coefficient of a vertex is defined through the number of triangles it belongs to.
Through the adjacency matrix we can write

\[
C_i = \frac{1}{k_i(k_i - 1)/2} \sum_{jk} a_{ij} a_{ik} a_{jk}
\]

DIRECTED (or even worse WEIGHTED) CLUSTERING

Fig. 1.7 Two examples of special vertices in a tree. On the left (as in a real tree) nutrients flow from the root (dark) to reach the leaves (light). Root and leaves are defined through their in-degree. On the right the case of river networks. Here light-coloured vertices represent the highest zone in the basin (no points uphill). The dark vertex is the outlet of the whole basin. Root and leaves are defined through the out-degree.
1.2 Trees

Trees are particularly important, since they appear in

- Girvan-Newman Algorithm for communities
- Taxonomy of species
- Taxonomy of information
1.2 Trees

Trees can also be viewed as a way to filter information. In a stock exchange the prices of all stocks are correlated, but restricting only to a Minimum Spanning Tree is a way to visualize the stronger correlations.

In order to build a Minimum Spanning Tree
1) Compute the correlation between the N(n-1)/2 vertices
2) Define a distance out of correlation.
3) Rank the distances
4) Draw the vertices of the shortest distance
5) Run on the ranking, whenever you find a new (or two) vertex(-ices), draw it (them) if you do not close a cycle.
6) Stop whenever all the vertices have been drawn
One way to visualize the behaviour of a tree (on top of the degree) is to check the behaviour of the basin size frequency distribution $P(n)$. 
A way to check the Conditioned Probability $P(k|k')$ that a vertex whose degree is $k$ is connected with another vertex with degree $k'$ is given by the measure of the average degree of the neighbours.
Similarly, one can define a measure of the assortativity by starting from the correlation function for degrees of vertices $i,j$

If we introduce the variance

We have the ASSORTATIVITY COEFFICIENT $r$ given by
# GRAPH THEORY

## 1.3 Vertex Correlation: Assortativity

<table>
<thead>
<tr>
<th>Network</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics co-authorship</td>
<td>52,909</td>
<td>0.363</td>
</tr>
<tr>
<td>Biology co-authorship</td>
<td>1,520,251</td>
<td>0.127</td>
</tr>
<tr>
<td>Mathematics co-authorship</td>
<td>253,339</td>
<td>0.120</td>
</tr>
<tr>
<td>Film actors collaboration</td>
<td>449,913</td>
<td>0.208</td>
</tr>
<tr>
<td>Company directors</td>
<td>7,673</td>
<td>0.276</td>
</tr>
<tr>
<td>Internet</td>
<td>10,697</td>
<td>-0.189</td>
</tr>
<tr>
<td>World Wide Web</td>
<td>269,504</td>
<td>-0.065</td>
</tr>
<tr>
<td>Protein interactions</td>
<td>2,115</td>
<td>-0.156</td>
</tr>
<tr>
<td>Neural network</td>
<td>307</td>
<td>-0.163</td>
</tr>
<tr>
<td>Little Rock Lake</td>
<td>92</td>
<td>-0.276</td>
</tr>
</tbody>
</table>

2.1 Introduction

Fig. 2.1 A division in communities in a highly clustered graph.

The presence of communities in a graph is one of the most important features.
- Communities are important for Amazon, to run their businesses
- For biologists to detect proteins with the same function
- For physicians to detect related diseases
Motifs are the smallest version of communities, it is currently under debate if their presence is important or not in the area of Complex networks.

Some communities must be defined as made by vertices with similar properties (they could share no edge at all). More often the communities are made by vertices sharing many edges.
The concept of centrality is at the basis of the Newman-Girvan method for the analysis of communities.

2.4 Centrality Measures Betweenness and Robustness

Starting from a graph we iteratively:
• Compute betweenness
• Cut the edges with the largest value of it

For the graph above the result is given by the series of deletions:
- E-F, E-I, B-D, A-D, C-F, C-L,
- D-H, G-L, D-I, B-E, A-B, H-I,
- F-L, C-G.
2.4 Centrality Measures Betweenness and Robustness

Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., and Arenas, A.

Lectures on Complex Networks
Protein interact in various ways during cell life.


Most of the reactions are not reversible and the network is oriented.

M.L. Gifford et al. Plant Physiology on-line Essay 12.2
6.3 Gene Regulatory Networks

A particular class of reactions in which there is the expression of a gene is at the basis of GRN.


Geophysical Networks

Trees and Basin distributions

The case of river networks presents a scale invariance in the basin size distributions.


Most of the activity on ecological networks is related to the statistical properties of food webs.

PEaCE LAB http://www.foodwebs.org

Small-world but not scale-free


Traditionally the analysis of the Internet structure is made by means of *traceroutes*. That is to say, by exploring all the paths from a given point to all the possible destinations.

**Internet Mapping Project**
http://www.cheswick.com/ches/map/
http://www.caida.org
http://www.cybergeography.org


**9.1 Internet**

TABLE I. Total number of new ($N_{\text{new}}$) and deleted ($N_{\text{del}}$) nodes in the years 1997, 1998, and 1999. We also report the number of deleted nodes with connectivity $k > 10$.

<table>
<thead>
<tr>
<th>Year</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{\text{new}}$</td>
<td>309</td>
<td>1990</td>
<td>3410</td>
</tr>
<tr>
<td>$N_{\text{del}}$</td>
<td>129</td>
<td>887</td>
<td>1713</td>
</tr>
<tr>
<td>$N_{\text{del}}(k &gt; 10)$</td>
<td>0</td>
<td>14</td>
<td>68</td>
</tr>
</tbody>
</table>

TABLE II. Average properties of the Internet for three different years. $N$, number of nodes; $E$, number of connections; $\langle k \rangle$, average connectivity; $\langle c \rangle$, average clustering coefficient; $\langle \ell \rangle$, average path length; $\langle b \rangle$, average betweenness. Figures in parentheses indicate the statistical uncertainty from averaging the values of the corresponding months in each year.

<table>
<thead>
<tr>
<th>Year</th>
<th>1997</th>
<th>1998</th>
<th>1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>3112</td>
<td>3834</td>
<td>5287</td>
</tr>
<tr>
<td>$E$</td>
<td>5450</td>
<td>6990</td>
<td>10100</td>
</tr>
<tr>
<td>$\langle k \rangle$</td>
<td>3.5(1)</td>
<td>3.6(1)</td>
<td>3.8(1)</td>
</tr>
<tr>
<td>$\langle c \rangle$</td>
<td>0.18(3)</td>
<td>0.21(3)</td>
<td>0.24(3)</td>
</tr>
<tr>
<td>$\langle \ell \rangle$</td>
<td>3.8(1)</td>
<td>3.8(1)</td>
<td>3.7(1)</td>
</tr>
<tr>
<td>$\langle b \rangle/N$</td>
<td>2.4(1)</td>
<td>2.3(1)</td>
<td>2.2(1)</td>
</tr>
</tbody>
</table>

---

A. Vazquez, R. Pastor-Satorras and A. Vespignani

*PRE* 65 066130 (2002)

R. Pastor-Satorras and A. Vespignani

_Evolution and Structure of Internet: A Statistical Physics Approach._

WWW is probably the largest network available of the order of billions of elements. Different centrality properties arise in such structure.


Social Networks

Wikipedia

Complexity

From Wikipedia, the free encyclopedia
Revision history

To view a previous version, click the date for that version.

Legend: (run) = difference with current version, (last) = difference with preceding version, m = minor edit

- [run] (last) 20:26, 5 April 2006 CX
- [run] (last) 18:52, 4 April 2006 Amalas m
- [run] (last) 21:55, 3 April 2006 80.218.158.176
- [run] (last) 22:14, 1 April 2006 Tripper234
- [run] (last) 18:58, 1 April 2006 81.151.50.58
- [run] (last) 18:58, 1 April 2006 81.151.50.58
- [run] (last) 06:46, 25 March 2006

Lectures on Complex Networks
Social Networks

Wikipedia

The degree shows fat tails that can be approximated by a power-law function of the kind $P(k) \sim k^{-\gamma}$ Where the exponent is the same both for in-degree and out-degree.


in–degree(empty) and out–degree(filled) Occurrency distributions for the Wikigraph in English (o) and Portuguese (Δ).

| DB | $|V|$ | $|E|$ |
|----|-----|-----|
| PT | 8,645 | 51,231 |
| IT | 13,132 | 159,965 |
| ES | 27,262 | 288,766 |
| FR | 42,987 | 660,401 |
| DE | 116,251 | 2,163,405 |
| EN | 339,834 | 5,278,037 |

Lectures on Complex Networks
Correlation based minimal spanning trees of real data from daily stock returns of 1071 stocks for the 12-year period 1987-1998 (3030 trading days). The node colour is based on Standard Industrial Classification system. The correspondence is:

red for mining  
green for transportation, electric, gas and sanitary services  
black for retail trade  
cyane for construction  
light blue for public administration  
purple for finance and insurance  
yellow for manufacturing  
magenta wholesale trade  
orange for service industries


Lectures on Complex Networks
The network shows a rather peculiar architecture. The banks form a disassortative network where large banks interact mostly with small ones.

INTRODUCTION

Free Resources on the web

• R. Albert, A.-L. Barabasi STATISTICAL MECHANICS OF COMPLEX NETWORKS
  http://arxiv.org/abs/cond-mat/0106096

• M.E.J. Newman THE STRUCTURE AND FUNCTION OF COMPLEX NETWORKS,
  http://arxiv.org/abs/cond-mat/03030516

• R. Diestel GRAPH THEORY, Springer-Verlag (2005)
  http://www.math.ubc.ca/~solymosi/443/GraphTheoryIII.pdf

• Networks Visualization and Analysis PAJEK, Springer-Verlag (2005)
  http://pajek.imfm.si/doku.php

• Bibliography I collected © http://www.citeulike.org/user/gcalda/tag/book
INTRODUCTION

Dissemination on Networks


INTRODUCTION

Resources in your library


Lectures on Complex Networks
INTRODUCTION

Resources in your library


