Dynamical processes on complex networks

Marc Barthélemy

CEA, IPhT, France
Outline

I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
      - Diameter, degree distribution, correlations (clustering, assortativity), modularity
   2. Some empirical results
   3. Models
      - Fixed N: Erdos-Renyi random graph, Watts-Strogatz, Molloy-Reed algorithm, Hidden variables
      - Growing networks: Barabasi-Albert and variants (copy model, weighted, fitness)

III. Dynamical processes on complex networks
   1. Overview: from physics to search engines and social systems
   2. Percolation
   3. Epidemic spread: contact network and metapopulation models
References: reviews on complex networks

• Statistical mechanics of complex networks
  Reka Albert, Albert-Laszlo Barabasi
  Reviews of Modern Physics 74, 47 (2002)
  cond-mat/0106096

• The structure and function of complex networks
  cond-mat/0303516

• Evolution of networks
  cond-mat/0106144
References: books on complex networks

• Evolution and structure of the Internet: A statistical Physics Approach
  R. Pastor-Satorras, A. Vespignani

• Evolution of networks: from biological nets to the Internet and WWW
  S.N. Dorogovtsev, J.F.F. Mendes

• Scale-free networks
  G. Caldarelli
Books and reviews on processes on complex networks

• Complex networks: structure and dynamics
  S. Boccaletti et al.

• Dynamical processes on complex networks
  A. Barrat, M. Barthélemy, A. Vespignani
  Cambridge Univ. Press, 2008.
What is a network?

Network = set of nodes joined by links $G = (V, E)$

- very abstract representation
- very general
- convenient to describe many different systems: biology, infrastructures, social systems, …
Most networks of interest are:

- Complex
- Very large

Statistical tools needed! (see next chapter)
Studies on complex networks (1998-)

• 1. Empirical studies
   Typology- find the general features

• 2. Modeling
   Basic mechanisms/reproducing stylized facts

• 3. Dynamical processes
   Impact of the topology on the properties of dynamical processes: epidemic spread, robustness, …
Empirical studies: Unprecedented amount of data.....

- Transportation infrastructures (eg. BTS)
- Census data (socio-economical data)
- Social networks (eg. online communities)
- Biological networks (-omics)
Empirical studies: sampling issues

- Social networks: various samplings/networks
- Transportation network: reliable data
- Biological networks: incomplete samplings
- Internet: various (incomplete) mapping processes
- WWW: regular crawls
- ...

possibility of introducing biases in the measured network characteristics
Networks characteristics

Networks: of very different origins

- The abstract character of the graph representation and graph theory allow to give some answers…
- Important ingredients for the modeling

Do they have anything in common? Possibility to find common properties?
Modeling complex networks

Microscopical processes

• many interacting elements
• dynamical evolution
• self-organisation

Statistical physics

Properties at the macroscopic level

• Non-trivial structure
• Emergent properties, cooperative phenomena
Networks in Information technology
Networks in Information technology: processes

Importance of Internet and the web

- Congestion
- Virus propagation
- Cooperative/social phenomena (online communities, etc.)
- Random walks, search (pagerank algo, ...
--- Internet ---

- Nodes=routers
- Links= physical connections

different granularities
Internet mapping

- continuously evolving and growing
- intrinsic heterogeneity
- self-organizing

Largely unknown topology/properties

Many mapping projects (topology and performance):
  - CAIDA, NLANR, RIPE, …

(http://www.caida.org)
Internet backbone

Nodes: Computers, routers

Links: physical lines

Large-scale visualization
World Wide Web

Virtual network to find and share informations

Over 1 billion documents

**ROBOT:** collects all URL’s found in a document and follows them recursively

**Nodes:** WWW documents

**Links:** URL links
Social networks
Social networks: processes

Many social networks are the support of some dynamical processes

- Disease spread
- Rumor propagation
- Opinion/consensus formation
- Cooperative phenomena
- ...
Scientific collaboration network

Nodes: scientists
Links: co-authored papers

Weights: depending on
• number of co-authored papers
• number of authors of each paper
• number of citations…
Citation network

Nodes: papers

Links: citations

Hopfield J.J., PNAS 1982

Science citation index
S. Redner
**Actor’s network**

**Nodes**: actors  
**Links**: cast jointly

---

**John Carradine**

The Sentinel (1977)  
The story of Mankind (1957)

---

Ava Gardner  
Groucho Marx

**distance**(Ava, Groucho) = 2

---

N = 212,250 actors  
\( \langle k \rangle = 28.78 \)

Character network

**Nodes**: characters

**Links**: co-appearance in a scene

Les Miserables - V. Hugo
Newman & Girvan, PRE (2004)
-> Community detection problem
The web of Human sexual contacts

Liljeros et al., Nature (2001)
Transportation networks
Transportation networks

Transporting energy, goods or individuals

- formation and evolution

- congestion, optimization, robustness

- disease spread

NB: spatial networks (planar): interesting modeling problems
Transporting individuals: global scale (air travel)

**Nodes**: airports

**Links**: direct flight
**Transporting individuals: intra city**

**TRANSIMS project**

**Nodes**: locations (homes, shops, offices, …)

**Links**: flow of individuals

---

### Example Data

<table>
<thead>
<tr>
<th>Person ID</th>
<th>Location ID</th>
<th>Location type</th>
<th>Arrival Time</th>
<th>Departure Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>4356</td>
<td>Home</td>
<td>00:00</td>
<td>07:00</td>
</tr>
<tr>
<td>116</td>
<td>98135</td>
<td>Work</td>
<td>08:00</td>
<td>11:00</td>
</tr>
<tr>
<td>116</td>
<td>71457</td>
<td>Work</td>
<td>11:20</td>
<td>13:00</td>
</tr>
<tr>
<td>116</td>
<td>98135</td>
<td>Work</td>
<td>13:20</td>
<td>17:00</td>
</tr>
<tr>
<td>116</td>
<td>4356</td>
<td>Home</td>
<td>18:00</td>
<td>19:15</td>
</tr>
<tr>
<td>116</td>
<td>21343</td>
<td>Social</td>
<td>19:30</td>
<td>21:00</td>
</tr>
<tr>
<td>116</td>
<td>4356</td>
<td>Home</td>
<td>21:00</td>
<td>07:00</td>
</tr>
<tr>
<td>324</td>
<td>12679</td>
<td>Home</td>
<td>00:00</td>
<td>07:00</td>
</tr>
<tr>
<td>324</td>
<td>431</td>
<td>School</td>
<td>08:00</td>
<td>14:00</td>
</tr>
<tr>
<td>324</td>
<td>12679</td>
<td>Home</td>
<td>14:30</td>
<td>19:00</td>
</tr>
</tbody>
</table>

---

Transporting goods

State of Indiana (Bureau of Transportation statistics)
Transporting electricity

New York state power grid

**Nodes**: power plants, transformers, etc,…

**Links**: cables
Transporting electricity  US power grid
Transporting gas European pipelines

European Natural Gas Transmission System in 2002

Source: Ruhrgas AG
Transporting water

**Nodes**: intersections, auxins sources  
**Links**: veins

Example of a planar network
Networks in biology
Networks in biology

- (sub-)cellular level: Extracting useful information from the huge amount of available data (genome, etc). Link structure-function?

- Species level: Stability of ecosystems, biodiversity
Neural Network

**Nodes**: neurons

**Links**: axons
Metabolic Network

**Nodes**: metabolites
**Links**: chemical reactions

Protein Interactions

**Nodes**: proteins
**Links**: interactions
Genetic (regulatory) network

**Nodes**: genes

**Links**: interaction

Genes are colored according to their cellular roles as described in the Yeast Protein Database.

Understanding these networks: one of the most important challenges in complex network studies.
Food webs

**Nodes**: species

**Links**: feeds on

N. Martinez
Dynamical processes on complex networks

Marc Barthélemy

CEA, IPhT, France
I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
      - Diameter, degree distribution, correlations (clustering, assortativity), modularity
   2. Some empirical results
   3. Models
      - Fixed N: Erdos-Renyi random graph, Watts-Strogatz, Molloy-Reed algorithm, Hidden variables
      - Growing networks: Barabasi-Albert and variants (copy model, weighted, fitness)

III. Dynamical processes on complex networks
   1. Overview: from physics to search engines and social systems
   2. Percolation
   3. Epidemic spread: contact network and metapopulation models
II. 1 Graph theory basics and statistical tools
Graph theory: basics

Graph \( G=(V,E) \)

- \( V \) = set of nodes/vertices \( i=1,\ldots,N \)
- \( E \) = set of links/edges \( (i,j) \)

**Undirected edge:**

**Directed edge:**

Bidirectional communication/interaction
**Graph theory: basics**

Maximum number of edges

- Undirected: $N(N-1)/2$
- Directed: $N(N-1)$

**Complete graph** $K_n$:

(all-to-all interaction/communication)
**Graph theory: basics**

**Planar graph**: can be embedded in the plane, i.e., it can be drawn on the plane in such a way that its edges may intersect only at their endpoints.

\[ \frac{E}{N} \leq 3 \quad (\langle k \rangle \leq 6) \]

- Planar
- Non planar
Adjacency matrix

N nodes $i=1,\ldots,N$

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E \end{cases}$$

Symmetric for undirected networks

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

0 - 1 - 2 - 3
Adjacency matrix

\[ A_{ij} = \begin{cases} 
1 & \text{if } (i,j) \in E \\
0 & \text{if } (i,j) \notin E
\end{cases} \]

N nodes i=1,…,N

Non symmetric for directed networks
Sparse graphs

Density of a graph \( D = \frac{|E|}{(N(N-1)/2)} \)

\[
D = \frac{\text{Number of edges}}{\text{Maximal number of edges}}
\]

Sparse graph: \( D << 1 \) \( \rightarrow \) Sparse adjacency matrix

Representation: lists of neighbours of each node

\( l(i, V(i)) \)

\( V(i) = \text{neighbourhood of } i \)
Paths

\[ G=(V,E) \]

Path of length \( n \) = ordered collection of
- \( n+1 \) vertices \( i_0, i_1, \ldots, i_n \in V \)
- \( n \) edges \( (i_0, i_1), (i_1, i_2), \ldots, (i_{n-1}, i_n) \in E \)

Cycle/loop = closed path \( (i_0=i_n) \)
Paths and connectedness

$G = (V, E)$ is connected if and only if there exists a path connecting any two nodes in $G$

- is connected
- is not connected
- is formed by two components
Paths and connectedness

G=(V,E)

In general: different components with different sizes

**Giant component** = component whose size scales with the number of vertices N

Existence of a giant component ↔ Macroscopic fraction of the graph is connected
Paths and connectedness: directed graphs

Paths can be *directed* (eg. WWW)
Paths and connectedness: directed graphs

“Bow tie” diagram for the WWW (Broder et al, 2000)
**Shortest paths**

**Shortest** path between i and j: minimum number of traversed edges

Distance \( l(i,j) = \text{minimum number of edges traversed on a path between } i \text{ and } j \)

**Diameter** of the graph:

\[
\max l(i, j)
\]

Average shortest path:

\[
\bar{l} = \frac{2}{N(N-1)} \sum_{i<j} l(i, j)
\]
Shortest paths

Complete graph: \( \ell(i, j) = 1 \ \forall i, j \)

Regular lattice: \( \ell(i, j) = |\vec{u}_i - \vec{u}_j| \)
\[ \bar{\ell} \sim N^{1/d} \]

“Small-world” network
\[ \bar{\ell} \ll N^{1/d} (\text{eg. } \log N) \]
Degree of a node

How many friends do you have?

$k=8$

$k>>1$: Hubs
Statistical characterization

- List of degrees $k_1, k_2, \ldots, k_N$  
  Not very useful!

- Histogram:
  $N_k =$ number of nodes with degree $k$

- Distribution:
  $P(k) = N_k / N =$ probability that a randomly chosen node has degree $k$

- Cumulative distribution: probability that a randomly chosen node has degree at least $k$ (reduced noise)

\[
P > (k) = \sum_{k' > k} P(k')
\]
Statistical characterization

- Average degree: \( \langle k \rangle = \frac{2E}{N} = \sum_k kP(k) \sim \int dk kP(k) \)

  Sparse graph: \( \langle k \rangle \ll N \)

- Second moment (variance):

  \( \langle k^2 \rangle - \langle k \rangle^2 \sim \int dk (k^2 - \langle k \rangle^2)P(k) \)

  Heterogeneous graph: \( \frac{\sqrt{\langle k^2 \rangle - \langle k \rangle^2}}{\langle k \rangle} \gg 1 \)
Degree distribution $P(k)$: probability that a node has $k$ links

- Poisson/Gaussian distribution
- Power-law distribution
Centrality measures

How to quantify the importance of a node?

- **Degree** = number of neighbours = $\sum_j a_{ij}$

  ![Diagram of a network with node i and its degree $k_i=5$](image)

- **Closeness centrality**

  $g_i = \frac{1}{\sum_j l(i,j)}$
Betweenness Centrality

\[ \sigma_{st}(ij) = \text{# of shortest paths from } s \text{ to } t \text{ via } (ij) \]

\[ g(ij) = \sum_{s,t} \frac{\sigma_{st}(ij)}{\sigma_{st}} \]

\( i j \): large centrality

\( j k \): small centrality
Multipoint degree correlations

$P(k)$: not enough to characterize a network

Large degree nodes tend to connect to large degree nodes
Ex: social networks

Large degree nodes tend to connect to small degree nodes
Ex: technological networks
Multipoint degree correlations

Measure of correlations:
\[ P(k', k'', \ldots k^{(n)}|k) : \text{conditional probability that a node of degree } k \text{ is connected to nodes of degree } k', k'', \ldots \]

Simplest case:
\[ P(k'|k) : \text{conditional probability that a node of degree } k \text{ is connected to a node of degree } k' \]
Case of random uncorrelated networks

• $P(k'|k)$ independent of $k$

• proba that an edge points to a node of degree $k'$:

\[
\frac{\text{number of edges from nodes of degree } k'}{\text{number of edges from nodes of any degree}} = \frac{k' \ N_{k'}}{\sum_{k''} k'' \ N_{k''}}
\]

\[
P_{unc}(k'|k) = \frac{k' \ P(k')}{\langle k \rangle}
\]

proportional to $k'$
2-points correlations: Assortativity

- $P(k'|k)$: difficult to handle and to represent

Are your friends similar to you?

$$k_{nn}(i) = \frac{1}{k_i} \sum_{j \in \mathcal{V}(i)} k_j$$
Assortativity

Exemple:

\[ k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j \]

\[ k_i = 4 \]

\[ k_{nn,i} = \frac{(3 + 4 + 4 + 7)}{4} = 4.5 \]
Correlation spectrum:

Putting together nodes which have the same degree:

Also given by:

\[ k_{nn}(k) = \frac{1}{N_k} \sum_{i/k_i=k} k_{nn,i} \]

class of degree \( k \)

\[ k_{nn}(k) = \sum_{k'} k' P(k' \mid k) \]
Assortativity

- **Assortative behaviour**: growing $k_{nn}(k)$
  - Example: social networks
    - Large sites are connected with large sites

- **Disassortative behaviour**: decreasing $k_{nn}(k)$
  - Example: internet
    - Large sites connected with small sites, hierarchical structure
3-points: Clustering coefficient

- $P(k',k''|k)$: cumbersome, difficult to estimate from data

Do your friends know each other?

$$C(i) = \frac{\text{# of links between neighbors}}{k(k-1)/2}$$
Correlations: Clustering spectrum

• Average clustering coefficient

\[ \overline{C} = \frac{1}{N} \sum_i C(i) \]

= average over nodes with very different characteristics

• Clustering spectrum:

\[ C(k) = \frac{1}{N_k} \sum_{i/k_i = k} C(i) \]

class of degree \( k \)

putting together nodes which have the same degree

(link with hierarchical structures)
More on correlations: communities and modularity

- Real networks are fragmented into groups or modules.

Modularity vs. Fonctionality?
More on correlations: communities and modularity

- Extract relevant and useful information in large complex network

- Mesoscopic objects: communities (or modules)

More links “inside” than “outside”
Applications

- Biological networks: function?
- Social networks: social groups
- Geography/regional science: urban communities
- Neurophysiology (?)
How can we compare different partitions?
Modularity [Newman & Girvan PRE 2004]

- Modularity: quality function $Q$
- $P_1$ is better than $P_2$ if $Q(P_1) > Q(P_2)$

$$Q(P) = \sum_{s=1}^{n_M} \frac{l_s}{L} - \left( \frac{d_s}{2L} \right)^2$$

- $n_M$: Number of modules in $P$
- $l_s$: Number of links in module $s$
- $L$: Number of links in the network
- $d_s$: Total degree of nodes in module $s$
Modularity [Newman & Girvan PRE 2004]

\[ Q(\mathcal{P}) = \frac{1}{L} \sum_{s=1}^{n_M} \left( \ell_s - \frac{d_s^2}{4L} \right) \]

\[ l_s = \# \text{ links in module } s \]

\[ d_s = \text{Total degree of nodes in module } s \]

\[ \frac{d_s^2}{4L} = \frac{d_s}{2L} \times \frac{d_s}{2L} \times L = \text{Average number of links in module } s \text{ for a random network} \]
Modularity \cite{NewmanGirvanPRE2004}

- probability that a half-link randomly selected belongs in module $s$: \( \frac{d_s}{2L} \)

- probability that a link begins in $s$ and ends in $s'$: \( \frac{d_s}{2L} \times \frac{d_{s'}}{2L} \)

- average number of links internal to $s$: \( \frac{d_s^2}{4L} \)
Motifs

• Find n-node subgraphs in real graph.

• Find all n-node subgraphs in a set of randomized graphs with the same distribution of incoming and outgoing arrows.

• Assign Z-score for each subgraph.

• Subgraphs with high Z-scores are denoted as **Network Motifs**.

Motifs:

\[ Z = \frac{N_{\text{real}} - \langle N_{\text{rand}} \rangle}{\sigma_{\text{rand}}} \]

Summary: characterization of large networks

Statistical indicators:
- degree distribution $P(k)$
- diameter $\ell = \max_{i,j} d(i, j)$
- clustering $C(i) = 2 \sum_{j,l} a_{ij} a_{il} a_{jl} / k_i (k_i - 1)$
- assortativity $k_{nn}(i) = \sum_j a_{ij} k_j / k_i$
- modularity $Q = \frac{1}{L} \sum_{s=1}^{nM} \left[ l_s - \left( \frac{d_s}{2L} \right)^2 \right]$
Beyond topology: weights

- Internet, Web, Emails: importance of traffic
- Ecosystems: prey-predator interaction
- Airport network: number of passengers
- Scientific collaboration: number of papers
- Metabolic networks: fluxes heterogeneous

Are:
  - Weighted networks
  - With broad distributions of weights
Weighted networks

General description: weights

\[ a_{ij}: 0 \text{ or } 1 \]

\[ w_{ij}: \text{continuous variable} \]

usually \( w_{ii} = 0 \)

symmetric: \( w_{ij} = w_{ji} \)
Weighted networks

Weights: on the links

**Strength** of a node:

\[ s_i = \sum_{j \in V(i)} w_{ij} \]

Naturally generalizes the degree to weighted networks

Quantifies for example the total traffic at a node
Weighted clustering coefficient

\[ C(i) = \frac{1}{k_i(k_i - 1)} \sum_{j \neq h} a_{ij}a_{jh}a_{ih} \]

\[ C^w(i) = \frac{1}{s_i(k_i - 1)} \sum_{j \neq h} a_{ij}a_{jh}a_{ih} \frac{w_{ij} + w_{ih}}{2} \]

- \( s_i = 16 \)
- \( c_i = 0.625 > c_i \)
- \( k_i = 4 \)
- \( c_i = 0.5 \)

- \( s_i = 8 \)
- \( c_i = 0.25 < c_i \)
Weighted clustering coefficient

Average clustering coefficient
\[ C = \frac{\sum_i C(i)}{N} \]
\[ C^w = \frac{\sum_i C^w(i)}{N} \]

Random(ized) weights: \( C = C_w \)
\( C < C_w \) : more weights on cliques
\( C > C_w \) : less weights on cliques

Clustering spectra
\[
C(k) = \frac{1}{N_k} \sum_{i/k_i=k} C(i)
\]
\[
C^w(k) = \frac{1}{N_k} \sum_{i/k_i=k} C^w(i)
\]
Weighted assortativity

\[ k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_{j \in V(i)} a_{ij} k_j \]

\[ k_i = 5; \quad k_{nn,i} = 1.8 \]
Weighted assortativity

\[ k_{nn,i} = \frac{1}{k_i} \sum_{j \in V(i)} k_j = \frac{1}{k_i} \sum_{j \in V(i)} a_{ij} k_j \]

\[ k_i = 5; \ k_{nn,i} = 1.8 \]
Weighted assortativity

\[ k_{nn,i}^w = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j \]

\[ k_i = 5; \ s_i = 21; \ k_{nn,i} = 1.8; \ k_{nn,i}^w = 1.2: \ k_{nn,i} > k_{nn,i}^w \]
Weighted assortativity

\[ k_{nn,i}^w = \frac{1}{s_i} \sum_{j \in V(i)} w_{ij} k_j \]

\[ k_i = 5; \quad s_i = 9; \quad k_{nn,i} = 1.8 \quad ; \quad k_{nn,i}^w = 3.2: \quad k_{nn,i} < k_{nn,i}^w \]
Participation ratio "disparity"

\[ Y_2(i) = \sum_{j \in V(i)} \left( \frac{w_{ij}}{s_i} \right)^2 \]

\[ \frac{1}{k_i} \text{ if all weights equal} \]

\[ \text{close to 1 if few weights dominate} \]
Participation ratio “disparity”

\[ Y_2(k) = \frac{1}{N_k} \sum_{i/k_i = k} Y_2(i) \]

In general: \( Y_2(k) \sim k^{-\theta} \)

\( \theta = 0 \)  Heterogeneous

\( \theta = 1 \)  Homogeneous (all equal)
Dynamical processes on complex networks

Marc Barthélemy

CEA, IPhT, France
Outline

I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
      - Diameter, degree distribution, correlations (clustering, assortativity), modularity
   2. Some empirical results
   3. Models
      - Fixed N: Erdos-Renyi random graph, Watts-Strogatz, Molloy-Reed algorithm, Hidden variables
      - Growing networks: Barabasi-Albert and variants (copy model, weighted, fitness)

III. Dynamical processes on complex networks
   1. Overview: from physics to search engines and social systems
   2. Percolation
   3. Epidemic spread: contact network and metapopulation models
II. 3 Models of complex networks
Some models of complex networks

- **Static networks:** N nodes from the beginning
  - The archetype: Erdos-Renyi random graph (60’s)
  - Fitness models (hidden variables)

- **Dynamic:** the network is growing
  - Scale-free network: Barabasi-Albert (1999)
  - Copy model
  - Weighted network model

- **Optimal networks**
Simplest model of random graphs: Erdös-Renyi (1959)

N nodes, connected with probability $p$

Paul Erdős and Alfréd Rényi (1959)
Erdös-Renyi (1959): recap

Some properties:

- Average number of edges
  \[ \langle E \rangle = p \frac{N(N-1)}{2} \]

- Average degree
  \[ \langle k \rangle = p(N-1) \]

Finite average degree \( \Rightarrow p \propto \frac{1}{N} \)
Erdös-Renyi model: recap

Proba to have a node of degree $k=\cdot$ connected to $k$ vertices,
• not connected to the other $N-k-1$

\[ P(k) = \binom{N}{k} p^k (1-p)^{N-1-k} \]

Large $N$, fixed $pN = \langle k \rangle$: Poisson distribution

\[ P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \]

Exponential decay at large $k$
Erdös-Renyi model: clustering and average shortest path

• N points, links with proba p: $C = p \sim \frac{1}{N}$

• Neglecting loops, $N(l)$ nodes at distance l:

$$N(l) = \langle k \rangle^l$$

For $\ell = \bar{l}$, $N(\bar{l}) = N$ ⇒

$$\bar{l} = \frac{\log N}{\log k}$$
Erdös-Renyi model: components

$\langle k \rangle < 1$ : many small subgraphs

$\langle k \rangle > 1$ : giant component + small subgraphs
Erdős-Renyi model: summary

- Poisson degree distribution
- Small clustering
- Small world
Generalized random graphs

Desired degree distribution: $P(k)$

- Extract a sequence $k_i$ of degrees taken from $P(k)$
- Assign them to the nodes $i=1,\ldots,N$
- Connect randomly the nodes together, according to their given degree
Generalized random graphs

- Average clustering coefficient
  \[ C = \frac{1}{N} \frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3} \]

- Average shortest path
  \[ \bar{\ell} \approx 1 + \frac{\log \left[ \frac{N}{\langle k \rangle} \right]}{\log \left[ \left( \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) / \langle k \rangle \right]} \]

Small-world and randomness
<table>
<thead>
<tr>
<th>Lattice</th>
<th>Real-World networks*</th>
<th>Random graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell \sim N^{1/d}$</td>
<td>$\ell \sim \log N$</td>
<td>$\ell \sim \log N$</td>
</tr>
<tr>
<td>$C \sim \text{const.}$</td>
<td>$C \sim \text{const.}$</td>
<td>$C \sim 1/N$</td>
</tr>
</tbody>
</table>

* Power grid, actors, C. Elegans

N nodes forms a regular lattice. With probability $p$, each edge is rewired randomly.

$=>$ Shortcuts

- Large clustering coeff.
- Short typical path
Watts-Strogatz (1998)

\[ \bar{\ell} = N^* F \left( \frac{N}{N^*} \right), \quad N^* \sim \frac{1}{p} \]

\[ F(x \ll 1) \sim x \]

\[ F(x \gg 1) \sim \log x \]

\[ C \sim (1 - p)^3 \]

MB and Amaral, PRL 1999

Barrat and Weight EPJB 2000
Fitness model (hidden variables)

Erdos-Renyi: p independent from the nodes

• For each node, a fitness \( x_i \sim \rho(x) \)

• Connect (i,j) with probability

\[
p_{ij} = f(x_i, x_j)
\]

• Erdos-Renyi: f=const

Soderberg 2002
Caldarelli et al 2002
Fitness model (hidden variables)

• Degree

\[ k(x) = N \int_{0}^{\infty} f(x, y) \rho(y) \, dy = NF(x) \]

• Degree distribution

\[ P(k) = \int dx \rho(x) \delta [k - k(x)] \]

\[ = \rho \left[ F^{-1} \left( \frac{k}{N} \right) \right] \frac{d}{dk} F^{-1} \left( \frac{k}{N} \right) \]
Fitness model (hidden variables)

• If \( \rho \) power law \( \rightarrow \) scale free network

• If \( \rho(x) \sim e^{-x} \) and \( f(x_i, x_j) = \theta [x_i + x_j - z(N)] \)

\[ \Rightarrow P(k) \sim k^{-2} \]

Generates a SF network!
Barabasi-Albert (1999)

Everything’s fine?

- Small-world network
- Large clustering
- Poisson-like degree distribution

Except that for

- Internet, Web
- Biological networks
- ...

Power-law distribution: Diverging fluctuations!
Internet growth

Moreover - dynamics!
(1) The number of nodes (N) is NOT fixed.

Networks continuously expand by the addition of new nodes

Examples:
WWW: addition of new documents
Citation: publication of new papers

(2) The attachment is NOT uniform.

A node is linked with higher probability to a node that already has a large number of links: “Rich get richer”

Examples:
WWW: new documents link to well known sites (google, CNN, etc)
Citation: well cited papers are more likely to be cited again
Barabasi-Albert (1999)

(1) **GROWTH** : At every time step we add a new node with $m$ edges (connected to the nodes already present in the system).

(2) **PREFERENTIAL ATTACHMENT** :

The probability $\Pi$ that a new node will be connected to node $i$ depends on the connectivity $k_i$ of that node

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

$P(k) \sim k^{-3}$

\[ \Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad \Rightarrow \quad \frac{dk_i(t)}{dt} = m \frac{k_i(t)}{\sum_{j=1}^{t} k_j(t)} \]

\[ \sum_{j=1}^{t} k_j = 2mt \quad \Rightarrow \quad k_i(t) = m \sqrt{\frac{t}{i}} \]
Barabasi-Albert (1999)

\[ k_i(t) = m \sqrt{\frac{t}{i}} \]

\[ P(k, t) = \frac{1}{t} \int_0^t di \delta(k - k_i(t)) \]

\[ u = k_i(t), \quad di = \frac{2m^2t}{u^3} du \]

\[ P(k, t) \sim \frac{2m^2}{k^3} \]
Barabasi-Albert (1999)

Clustering coefficient

\[ C \sim \frac{(\log N)^2}{N} \]

Average shortest path

\[ \bar{\ell} \sim \frac{\log N}{\log \log N} \]
Copy model

Growing network:

a. Selection of a vertex

b. Introduction of a new vertex

c. The new vertex copies $m$ links of the selected one

d. Each new link is kept with proba $1-\alpha$, rewired at random with proba $\alpha$
Copy model

Probability for a vertex to receive a new link at time $t$ ($N=t$):

- Due to random rewiring: $\alpha/t$
- Because it is neighbour of the selected vertex: $k_{in}/(mt)$

\[
\frac{\partial k_{in,s}(t)}{\partial t} = m \left[ \frac{\alpha}{t} + (1 - \alpha) \frac{k_{in,s}}{mt} \right]
\]

effective preferential attachment, without a priori knowledge of degrees!
Copy model

Degree distribution:

\[ P(k_{in}) \sim (k_0 + k_{in})^{-\frac{2-\alpha}{1-\alpha}} \]

\[ \Rightarrow \text{Heavy-tails} \]

model for WWW and evolution of genetic networks
Preferential attachment: generalization

Who is richer?

Rank known but not the absolute value

Fortunato et al, PRL (2006)
Preferential attachment: generalization

Rank known but not the absolute value

\[ \Pi(s) = \frac{R(s)^{-\alpha}}{\sum_j R(j)^{-\alpha}} \]

\[ \Rightarrow \gamma = 1 + \frac{1}{\alpha} \]

Scale free network even in the absence of the value of the nodes' attributes

Fortunato et al, PRL (2006)
Weighted networks

- Topology and weights uncorrelated

\[ s(k) = \langle w \rangle k \quad (s(k) = \frac{1}{N_k} \sum_{i/k_i=k} s_i) \]

- (2) Model with correlations?

\[ s(k) \sim k^\beta \]

\[ s(k) = Ak \quad \text{with} \quad A \neq \langle w \rangle \]
Weighted growing network

• **Growth**: at each time step a new node is added with $m$ links to be connected with previous nodes

• **Preferential attachment**: the probability that a new link is connected to a given node is proportional to the node’s strength

$$\Pi_{n\rightarrow i} = \frac{s_i}{\sum_j s_j}$$

Barrat, Barthelemy, Vespignani, PRL 2004
Redistribution of weights

New node: n, attached to i
New weight $w_{ni}=w_0=1$
Weights between i and its other neighbours:

\begin{align*}
  w_{ij} &\rightarrow w_{ij} + \Delta w_{ij} \\
  \Delta w_{ij} &= \delta \frac{w_{ij}}{s_i}
\end{align*}

\begin{align*}
  s_i &\rightarrow s_i + w_0 + \delta
\end{align*}

The new traffic n-i increases the traffic i-j
Mean-field evolution equations

\[ \frac{dk_i}{dt} = m \frac{s_i}{\sum_l s_l} \]

\[ \frac{ds_i}{dt} = m \frac{s_i}{\sum_l s_l} + \sum_{j \in V(i)} m \frac{s_j}{\sum_l s_l} \delta \frac{w_{ij}}{s_j} \]

\[ \sum_{i=1}^{t} k_i(t) \approx 2mt \]

\[ \sum_{i=1}^{t} s_i(t) \approx 2m(1 + \delta)t \]
Mean-field evolution equations

\[ \frac{ds_i}{dt} = \frac{2\delta + 1}{2\delta + 2} \frac{s_i}{t} \]

\[ \frac{dk_i}{dt} = \frac{s_i}{2(1 + \delta)t} \]

\[ \Rightarrow P(k) \sim k^{-\gamma}, \quad P(s) \sim s^{-\gamma} \quad \gamma = \frac{4\delta + 3}{2\delta + 1} \]

\[ P(w) \sim w^{-\alpha}, \quad \alpha = 2 + \frac{1}{\delta} \]

Correlations topology/weights:

\[ s_i \approx (2\delta + 1)k_i \not\equiv \langle w \rangle k_i \]
Numerical results: $P(w)$, $P(s)$

$(N=10^5)$
Another mechanism: Heuristically Optimized Trade-offs (HOT)

New vertex $i$ connects to vertex $j$ by minimizing the function

$$Y(i,j) = \alpha d(i,j) + V(j)$$

$d =$ euclidean distance
$V(j) =$ measure of centrality

Optimization of conflicting objectives
Dynamical processes on complex networks

Marc Barthélemy
CEA, IPhT, France
I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
   2. Some empirical results
   3. Models
      - Erdos-Renyi random graph
      - Watts-Strogatz small-world
      - Barabasi-Albert and variants

III. Dynamical processes on complex networks
   1. Percolation
   2. Epidemic spread
   3. Other processes: from physics to social systems
III. 1 Percolation, random failures, targeted attacks
Consequences of the topological heterogeneity

- Robustness and vulnerability
Robustness

Complex systems maintain their basic functions even under errors and failures (cell: mutations; Internet: router breakdowns)

$S$: fraction of giant component

Fraction of removed nodes, $f$
Case of Scale-free Networks

Random failure $f_c = 1$ \quad (2 < \gamma \leq 3)

Attack = progressive failure of the most connected nodes $f_c < 1$

Random failures: percolation

f = fraction of nodes removed because of failure

\[ p = 1 - f \]

p = probability of a node to be present in a percolation problem

Question: existence or not of a giant/percolating cluster, i.e. of a connected cluster of nodes of size \( O(N) \)
Percolation in complex networks

$q =$ probability that a randomly chosen link does not lead to a giant percolating cluster

$$q = \sum_{k} \frac{kP(k)}{\langle k \rangle} q^{k-1}$$

- Proba that none of the outgoing $k-1$ links leads to a giant cluster
- Average over degrees
- Proba that the link leads to a node of degree $k$
Percolation in complex networks

$q =$ probability that a randomly chosen link does not lead to a giant percolating cluster

$$q = F(q), \quad \text{with} \quad F(q) = \frac{1}{\langle k \rangle} \sum_k kP(k)q^{k-1}$$

$q = 1$, always solution

Other solution iff $F''(1) > 1$
q = probability that a randomly chosen link does not lead to a giant percolating cluster

\[ q = F(q), \quad \text{with} \quad F(q) = \frac{1}{\langle k \rangle} \sum_k k P(k) q^{k-1} \]

\[ F'(1) \geq 1 \quad \iff \quad \langle k^2 \rangle \geq 2 \langle k \rangle \]

“Molloy-Reed” criterion for the existence of a giant cluster in a random uncorrelated network
Random failures

Initial network: $P_0(k), <k>_0, <k^2>_0$
After removal of fraction $f$ of nodes: $P_f(k), <k>_f, <k^2>_f$

Node of degree $k_0$ becomes of degree $k$ with proba

$$C_{k_0}^k (1 - f)^k f^{k_0 - k}$$

$$P_f(k) = \sum_{k_0} P_0(k_0) C_{k_0}^k (1 - f)^k f^{k_0 - k}$$
Random failures

Initial network: $P_0(k), \langle k \rangle_0, \langle k^2 \rangle_0$

After removal of fraction $f$ of nodes: $P_f(k), \langle k \rangle_f, \langle k^2 \rangle_f$

\[
\begin{align*}
\langle k \rangle_f &= (1 - f)\langle k \rangle_0 \\
\langle k^2 \rangle_f &= (1 - f)^2\langle k^2 \rangle_0 + f(1 - f)\langle k \rangle_0
\end{align*}
\]

Molloy-Reed criterion:
existence of a giant cluster iff \( \langle k^2 \rangle_f \geq 2\langle k \rangle_f \)

\[f \geq f_c, \text{ with } f_c = 1 - \frac{\langle k \rangle_0}{\langle k^2 \rangle_0 - \langle k \rangle_0}\]

\[\langle k^2 \rangle_0 \to \infty \quad \implies \quad f_c \to 1 \quad \text{Robustness!!!}\]
Attacks: various strategies

- Most connected nodes
- Nodes with largest betweenness
- Removal of links linked to nodes with large $k$
- Removal of links with largest betweenness
- Cascades
- ...
Failure cascade

(a) and (b): random
(c): deterministic and dissipative

Dynamical processes on complex networks

Marc Barthélemy

CEA, IPhT, France
I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
   2. Some empirical results
   3. Models
      - Erdos-Renyi random graph
      - Watts-Strogatz small-world
      - Barabasi-Albert and variants

III. Dynamical processes on complex networks
   1. Overview: from physics to social systems
   2. Percolation
   3. Epidemic spread
III. 3 Epidemic spread

a. Contact network
b. Metapopulation model
General references

Pre(Re-)prints archive:

Books (epidemiology):


Epidemiology vs. Etiology: Two levels

- Microscopic level (bacteria, viruses): compartments
  - Understanding and killing off new viruses
  - Quest for new vaccines and medicines

- Macroscopic level (communities, species)
  - Integrating biology, movements and interactions
  - Vaccination campaigns and immunization strategies
Epidemiology

- One population:
  - nodes=individuals
  - links=possibility of disease transmission

- Metapopulation model: many populations coupled together
  - Nodes= cities
  - Links=transportation (air travel, etc.)
Modeling in Epidemiology

R.M. May Stability and Complexity in Model Ecosystems 1973
Epidemics spread on a ‘contact network’: Different scales, different networks:

- Individual: Social networks (STDs on sexual contact network)
  NB: Sometimes no network!

- Intra-urban: Location network (office, home, shops, …)

- Inter-urban: Railways, highways

- Global: Airlines (SARS, etc)
Networks and epidemiology

High School dating
www.umich.edu/~mejn

**Nodes**: kids (boys & girls)

**Links**: date each other
The web of Human sexual contacts
(Liljeros et al., Nature, 2001)

Broad degree distributions
Broad degree distributions

E-mail network

Ebel et al., PRE (2002)
Simple Models of Epidemics

**Topology of the system**: the pattern of contacts along which infections spread in population is identified by a network

- Each node represents an individual
- Each link is a connection along which the virus can spread
Simple Models of Epidemics

Stochastic model:

• SIS model:

• SIR model:

• SI model:

\[ S \xrightarrow{\lambda} I \xrightarrow{\mu} S \]

\[ S \xrightarrow{\lambda} I \xrightarrow{\mu} R \]

\[ S \xrightarrow{\lambda} I \]

\( \lambda \): proba. per unit time of transmitting the infection

\( \mu \): proba. per unit time of recovering
Numerical simulation:
For all nodes:
   (i) Check all neighbors and if infected ->(ii). If no infected => Next neighbor

(ii) Node infected with proba $\lambda dt$

(iii) proceed to next node

Infection proba is then:

$$[1 - (1 - \lambda dt)^{(ki)}] = \lambda ki dt$$
SIS on Random Graphs

- Random graph: \( k \approx \langle k \rangle \)
- Equation for density of infected \( i(t) = \frac{l}{N} \):

\[
\partial_t i(t) = -\mu i(t) + \lambda [1 - i(t)] \langle k \rangle i(t)
\]

- \( \mu^{-1} \) is the average lifetime of the infected state
- \([1 - i(t)]\) = prob. of not being infected
- \( \langle k \rangle i(t) \) is the prob. of at least one infected neighbor
- \( \lambda \) = prob(\( S \to I \))
- \( s(t) + i(t) = 1 \) \( \partial_t s = -\partial_t i \)
SIS on Random Graphs

- In the stationary state $\partial_t i=0$, we obtain:

$$\mu i = \lambda \langle k \rangle i [1 - i]$$

$$i = 0 \text{ if } \lambda < \lambda_c$$

Or:

$$i = 1 - \frac{\mu}{\lambda \langle k \rangle} = \frac{\lambda - \lambda_c}{\lambda}$$

With:

$$\lambda_c = \frac{\mu}{\langle k \rangle}$$
SIR on Random Graphs

- Equation for $s(t)$, $i(t)$ and $r(t)$:

\[
\frac{\partial}{\partial t} s(t) = -\lambda \langle k \rangle i [1-i] \quad (S \rightarrow I)
\]

\[
\frac{\partial}{\partial t} i(t) = -\mu i + \lambda \langle k \rangle i [1-i] \quad (S \rightarrow I \text{ and } I \rightarrow R)
\]

\[
\frac{\partial}{\partial t} r(t) = +\mu i \quad (I \rightarrow R)
\]

- $\mu^{-1}$ is the average lifetime of the infected state
- $s(t)+i(t)+r(t) = 1$
stationary state: \( \partial_t x = 0 \)

\[ \Rightarrow \quad s=i=0, \ r=1 \]

outbreak condition: \( \partial_t i(t=0) > 0 \)

\[ \Rightarrow \quad \lambda > \lambda_c = \mu / \langle k \rangle \]

\[ \Rightarrow \quad \mu >> 1 \text{ then } \lambda_c >> 1 \text{ (eg. hemo. fever)} \]
Epidemic Threshold $\lambda_c$

The epidemic threshold is a general result (SIS, SIR)

- Epilemic threshold = critical point
- Prevalence $i$ = order parameter

The question of thresholds in epidemics is central
Epidemic spreading on heterogeneous networks

Number of contacts (degree) can vary a lot huge fluctuations

Mean-field approx by degree classes: necessary to introduce the density of infected \( i_k \) having degree \( k \), \( i_k \).
SIS model on SF networks

• Scale-free network: $\langle k^2 \rangle \gg \langle k \rangle^2$

• Mean-field equation for the relative density of infected nodes of degree $k$ $i_k(t)$:

$$\partial_t i_k(t) = -\mu i_k(t) + \lambda k [1 - i_k(t)] \Theta(i_k)$$

where $\Theta$ is the probability that any given link points to an infected node.
SIS model on uncorrelated SF networks

• Estimate of $\Theta$: proba of finding
  - (i) a neighbor of degree $k'$
  - (ii) and infected

$$
\Theta(i_k) = \sum_{k'} i_{k'} \frac{k' N(k')}{\sum_k k N(k)} = \frac{\langle k i_k \rangle}{\langle k \rangle} 
$$

Proba. that the neighbor is infected

Proba. that degree(neighbor) = $k'$
We finally get:

\[ \partial_t i_k(t) = -\mu i_k(t) + \lambda k[1 - i_k] \frac{\langle ki_k \rangle}{\langle k \rangle} \]

Stationary state \( \partial_t i_k = 0 \):

\[ i_k = \frac{\lambda k \Theta}{1 + \lambda k \Theta} \]

\[ \Theta = \frac{\lambda}{\langle k \rangle} \left( \frac{k^2 \Theta}{1 + \lambda k \Theta} \right) \equiv F(\Theta) \]
SIS model on SF networks

Solution of $\Theta = F(\Theta)$ non zero if $F'(\Theta=0)>1$:

$$\lambda > \lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$$

[ Pastor Satorras & Vespignani, PRL 86, 320 (2001)]
Effect of topology: Epidemic threshold

- Effect of (strong) heterogeneities of the contact network

\[ \lambda_c \sim \frac{\langle k \rangle}{\langle k^2 \rangle} \rightarrow 0 \]

(Pastor-Satorras & Vespignani, PRL '01)
Consequence: Random immunization

g = density of immune nodes

Epidemic dies if \( \lambda (1-g) < \lambda_c \)

Regular or random networks

Immunization threshold:
\[
g_c = \frac{\lambda - \lambda_c}{\lambda}
\]
Consequence: Random immunization

- $g$: fraction of immunized nodes
- Epidemic dies if $\lambda(1 - g) < \lambda_c$
- Scale-free network $\lambda_c = 0$

Immunization threshold $g_c = 1$

- Random immunization is totally ineffective
- Different immunization specifically devised for highly heterogeneous systems
Targeted immunization strategies

Progressive immunization of crucial nodes

Epidemic threshold is reintroduced

\[ g_c = \exp\left(-\frac{2}{m\lambda}\right) \]

[ Pastor Satorras & Vespignani, PRE 65, 036104 (2002)]
[ Dezso & Barabasi cond-mat/0107420; Havlin et al. Preprint (2002)]
Acquaintances immunization strategies: Finding hubs

- Proba of finding a hub: \( \frac{1}{N} \ll 1 ! \)

- Average connectivity of neighbor:

\[
\sum_{k'} k' \frac{N(k')}{\sum_k k N(k)} = \frac{\langle k^2 \rangle}{\langle k \rangle} \gg 1!
\]

Vaccinate acquaintances!

Dynamics of Outbreaks

- Stationary state understood

- What is the dynamics of the infection?
Stages of an epidemic outbreak

Infected individuals => prevalence/incidence
Dynamics: Time scale of outbreak

At short times:

- SI model $S \rightarrow I$

\[ i(t) \approx 0 \quad s(t) \approx 1 \]

\[ \Rightarrow \lambda si \approx \lambda i \]
Dynamics: Time scale of outbreak

• Exponential networks: \( k \approx \langle k \rangle \)

\[
\frac{\partial t i(t)}{} \approx \lambda \langle k \rangle i
\]

\[
\Rightarrow \quad i(t) \approx e^{t/\tau_h}
\]

with:

\[
\tau_h = \frac{1}{\lambda \langle k \rangle}
\]
Dynamics: Time scale of outbreak (SF)

- Scale-free networks: \( \langle k^2 \rangle \gg \langle k \rangle^2 \)

\[
\partial_t i_k = \lambda k [1 - i_k] \Theta
\]

\[
\Theta = \frac{\langle ki_k \rangle}{\langle k \rangle}
\]

At first order in \( i(t) \)

\[
\partial_t i_k \approx \lambda k \Theta(t)
\]

\[
\Rightarrow \partial_t \Theta \approx \lambda \frac{\langle k^2 \rangle}{\langle k \rangle} \Theta
\]
Dynamics: Time scale of outbreak (SF)

• Scale-free networks:

\[ i(t) = \langle i_k \rangle \approx e^{t/\tau_{sf}} \]

with:

\[ \tau_{sf} = \frac{\langle k \rangle}{\lambda \langle k^2 \rangle} \]

[ MB, A. Barrat, R. Pastor-Satorras & A. Vespignani, PRL (2004)]
Dynamics: Cascade

Which nodes are infected?

What is the infection scenario?
Epidemics: Summary

1. Absence of an epidemic threshold
\[ \lambda_c \sim \frac{1}{\langle k^2 \rangle} \sim 0 \]

2. Random immunization is totally ineffective (target !)

3. Outbreak time scale controlled by fluctuations
\[ \tau_{sf} \sim \frac{1}{\langle k^2 \rangle} \sim 0 \]

4. Cascade process:
   Seeds → Hubs → Intermediate → Small k

**Dramatic consequences of the heterogeneous topology !**
Rumor spreading

• Spread of an information among connected individuals
  - protocols for data dissemination
  - marketing campaign (viral marketing)

• Difference with epidemic spread:
  - stops if surrounded by too many non-ignorant

• Categories:
  - ignorant \( i(t) \) (susceptible)
  - spreaders \( s(t) \) (infected)
  - stifflers \( r(t) \) (recovered)
Rumor spreading

- Process:
  - ignorant $\rightarrow$ spreader ($\lambda$)
  - spreader+spreader $\rightarrow$ stiffler (!?)
  - spreader+stiffler $\rightarrow$ stiffler

- Equations:

\[
\partial_t i(t) = -\lambda <k> i(t) s(t) \quad (i \rightarrow s)
\]

\[
\partial_t s(t) = +\lambda <k> i(t) s(t) - \alpha <k> s(t)[s(t)+r(t)] \quad (i \rightarrow s \text{ and } s \rightarrow r)
\]

\[
\partial_t r(t) = +\alpha i <k> s(t)[s(t)+r(t)] \quad (s \rightarrow r)
\]
Rumor spreading

- Random graph-Epidemic threshold:

\[
\frac{\lambda}{\alpha} > 0
\]

Always spreading!
Dynamical processes on complex networks

Marc Barthélemy

CEA, IPhT, France
Outline

I. Introduction: Complex networks
   1. Example of real world networks and processes

II. Characterization and models of complex networks
   1. Tool: Graph theory and characterization of large networks
   2. Some empirical results
   3. Models
      - Erdos-Renyi random graph
      - Watts-Strogatz small-world
      - Barabasi-Albert and variants

III. Dynamical processes on complex networks
   1. Overview: from physics to social systems
   2. Percolation
   3. Epidemic spread
II. 2 Empirical results

Main features of real-world networks
Plan

- Motivations
- Global scale: metapopulation model
- Theoretical results
  - Predictability
  - Epidemic threshold, arrival times
- Applications
  - SARS
  - Pandemic flu: Control strategy and Antivirals
Epidemiology: past and current

- Human movements and disease spread

- Complex movement patterns: different means, different scales (SARS): Importance of networks
Motivations

**Modeling**: SIR with spatial diffusion \( i(x,t) \):

\[
\frac{\partial i}{\partial t} = \lambda si - \mu i + D \nabla^2 i
\]

where: \( \lambda = \) transmission coefficient (fleas->rats->humans)
\( \mu = 1/\)average infectious period
\( S_0 = \)population density, \( D = \)diffusion coefficient

\[
V = 2(\lambda S_0 D)^{1/2} \left[ 1 - \frac{\mu}{\lambda S_0} \right]^{1/2}
\]

\( \approx 140 \) miles/year
Complete IATA database:
- 3100 airports worldwide
- 220 countries
- $\approx 20,000$ connections
- $w_{ij}$ #passengers on connection i-j
- $P(k)$
- $P(N)$
- $P(w)$
- $N(T)$
- $>99\%$ total traffic

Airline network and epidemic spread
Modeling in Epidemiology

<table>
<thead>
<tr>
<th>Homogeneous mixing</th>
<th>Social structure</th>
<th>Contact network models</th>
<th>Multi-scale models</th>
<th>Agent Based models</th>
</tr>
</thead>
</table>

- Hospital
- School
- Home
- Work
Metapopulation model

Baroyan et al, 1969: ≈ 40 russian cities
Rvachev & Longini, 1985: 50 airports worldwide
Grais et al, 1988: 150 airports in the US
Hufnagel et al, 2004: 500 top airports worldwide

Colizza, Barrat, Barthelemy & Vespignani, PNAS (2006)
Meta-population networks

Each node: internal structure
Links: transport/traffic
Metapopulation model

- Rvachev Longini (1985)

\[ \frac{\partial_t I_i(t)}{= K_i[I_i(t)] + \Omega_i(t)} \]

Inner city term

Travel term

- Transport operator:

\[ \Omega_i(t) = \sum_{j \in \Gamma_i} p_{ji} I_j(t) - p_{ij} I_i(t) \]

Stochastic model

compartmental model + air transportation data
Stochastic model: travel term

Travel probability from PAR to FCO:

\[ P_{PAR,FCO} = \frac{\xi_{PAR,FCO}}{N_{PAR}} \delta t \]

\( \xi_{PAR,FCO} \) # passengers from PAR to FCO: Stochastic variable, multinomial distr.
Stochastic model: travel term

Transport operator:

\[ \Omega_{\text{PAR}} (\{X\}) = \sum_l (\xi_{l,\text{PAR}} (\{X_l\}) - \xi_{\text{PAR},l} (\{X_{\text{PAR}}\})) \]

- **other source of noise**: \( w_{jl}^{\text{noise}} = w_{jl} [\alpha + \eta (1-\alpha)] \quad \alpha = 70\% \)

- **two-legs travel**: \( \Omega_j (\{X\}) = \Omega_j^{(1)} (\{X\}) + \Omega_j^{(2)} (\{X\}) \)
Discrete stochastic Model

\[ S_j(t + \Delta t) - S_j(t) = -\beta \frac{I_j S_j}{N_j} \Delta t + \eta_{j,1} + \Omega_j(\{S\}) \]

\[ I_j(t + \Delta t) - I_j(t) = +\beta \frac{I_j S_j}{N_j} \Delta t - \mu I_j \Delta t - \eta_{j,1} + \eta_{j,2} + \Omega_j(\{I\}) \]

\[ R_j(t + \Delta t) - R_j(t) = +\mu I_j \Delta t - \eta_{j,2} + \Omega_j(\{R\}) \]

\[ \eta_{j,1}, \eta_{j,2} \quad \text{Independent Gaussian noises} \]

\[ = \Rightarrow 3100 \times 3 \text{ differential coupled stochastic equations} \]

Metapopulation model

- Theoretical studies
  - Predictability ?
  - Epidemic threshold ?
  - Arrival times distribution ?
Propagation pattern

Epidemics starting in Hong Kong
Propagation pattern

Epidemics starting in Hong Kong

[t=14 days] [t=30 days] [t=45 days] [t=65 days] [t=145 days]
Propagation pattern

Epidemics starting in Hong Kong

[Series of maps showing propagation over time]
Predictability

One outbreak realization:

Another outbreak realization? Effect of noise?
Overlap measure

Similarity between 2 outbreak realizations:

\[ \Theta(t) \]

Overlap function

\( \Theta(t) = 1 \)

\( \Theta(t) < 1 \)
Predictability

- No degree fluctuations
- No weight fluctuations

- Degree heterogeneity
- Weight heterogeneity

Colizza, Barrat, Barthelemy & Vespignani, PNAS (2006)
Predictability

- Effect of heterogeneity:
  - degree heterogeneity: decreases predictability
  - Weight heterogeneity: increases predictability!

Good news: Existence of preferred channels!

Epidemic forecast, risk analysis of containment strategies
Theoretical result: Threshold

- Reproductive number for a population:

  $\langle k \rangle, \langle k^2 \rangle$ moments of the degree distribution

  $p$ travel probability

- Effective for reproductive number for a network of populations:

  Travel restrictions not efficient !!!

Colizza & Vespignani, PRL (2007)
Travel limitations....

Theoretical result: average arrival time

- Ansatz for the arrival time at site \( t \) (starting from \( s \))

\[
\chi(s, t) = \min_{P_{st}} \sum_{(kl)\in P_{st}} \left[ \ln \left( \frac{N_k \lambda}{w_{kl}} \right) - \gamma \right]
\]

- \( N_k \): Population of city \( k \)
- \( w_{kl} \): Traffic between \( k \) and \( l \)
- \( \lambda \): Transmissibility
- \( P_{st} \): Set of paths between \( s \) and \( t \)
Applications

- 1. SARS: test of the model
- 2. Control strategy testing: antivirals
Application: SARS

- refined compartmentalization
- parameter estimation: clinical data + local fit
- geotemporal initial conditions: available empirical data
- modeling intervention measures: standard effective modeling
SARS: predictions

Comparison forecasts/empirical data
July 11, 2003

- Red: correct prediction of outbreak
- Dotted red: correct prediction of no outbreak
- Light green: incorrect prediction of no outbreak
- Grey: no airports

North Pacific Ocean
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
Indian Ocean
SARS: predictions (2)

Colizza, Barrat, Barthelemy & Vespignani, bmc med (2007)
### SARS: predictions (3) - “False positives”

<table>
<thead>
<tr>
<th>Country</th>
<th>Median</th>
<th>90% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>30</td>
<td>[9-114]</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2</td>
<td>[1-10]</td>
</tr>
<tr>
<td>United Arab Emir.</td>
<td>2</td>
<td>[1-11]</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>2</td>
<td>[1-14]</td>
</tr>
<tr>
<td>Bahrain</td>
<td>1</td>
<td>[1-20]</td>
</tr>
<tr>
<td>Cambodia</td>
<td>1</td>
<td>[1-8]</td>
</tr>
<tr>
<td>Nepal</td>
<td>1</td>
<td>[1-20]</td>
</tr>
<tr>
<td>Brunei</td>
<td>1</td>
<td>[1-8]</td>
</tr>
<tr>
<td>Israel</td>
<td>1</td>
<td>[1-13]</td>
</tr>
<tr>
<td>Mauritius</td>
<td>1</td>
<td>[1-32]</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>1</td>
<td>[1-11]</td>
</tr>
</tbody>
</table>
More from SARS - Epidemic pathways

- For every infected country:

  where is the epidemic coming from?

  - Redo the simulation for many disorder realizations (same initial conditions)

  - Monitor the occurrence of the paths (source-infected country)
SARS - what did we learn?

- Metapopulation model, no tunable parameter:
  
good agreement with WHO data!

- Existence of pathways:
  
  confirms the possibility of epidemic forecasting!

  Useful information for control strategies
Application of the metapopulation model: effect of antivirals

- Threat: Avian Flu
- Question: use of antivirals
  - Best strategy for the countries?
- Model:
  - Etiology of the disease (compartments)
  - Metapopulation+Transportation mode (air travel)
2. Antivirals

- Flu type disease: Compartments
Pandemic forecast…

Pandemic flu with $R_0 = 1.6$ starting from Hanoi (Vietnam) in October (2006)
Baseline scenario

Effect of antivirals

- Comparison of strategies (travel restrictions not efficient)
  - Baseline: reference point (no antivirals)
  - “Uncooperative”: each country stockpiles AV
  - “Cooperative”: each country gives 10% (20%) of its own stock
Effect of antivirals: Strategy comparison

Best strategy: Cooperative!

Conclusions and perspectives

- Global scale (metapopulation model)
  - Pandemic forecasting
  - Theoretical problems (reaction-diffusion on networks)

- Smaller scales-country, city (?)
  - Global level: “simplicity” due to the dominance of air travel
  - Urban area ? Model ? What can we say about the spread of a disease ? Always more data available…
Outlook

- Prediction/Predictability vs disease parameters, initial conditions, errors etc.

- Integration of data sources: wealth, census, traveling habits, short range transportation.

- Individual/population heterogeneity

- Social behavior/response to crisis
Collaborators and links

- **Collaborators**
  - A. Barrat (LPT, Orsay)
  - V. Colizza (ISI, Turin)
  - A.-J. Valleron (Inserm, Paris)
  - A. Vespignani (IU, Bloomington)

- **PhD students**
  - P. Crepey (Inserm, Paris)
  - A. Gautreau (LPT, Orsay)

- **Collaboratory**

  [Cx-Nets](http://cxnets.googlepages.com)

  marc.barthelemy@gmail.com