Structure and dynamics of Complex Networks

A.Vespignani
A network is a system that allows its abstract/mathematical representation as a graph.

- Vertices (nodes) = elements of the system
- Edges (links) = interactions/relations among the elements of the system
Many complex systems (from the molecular level to the scale of large communications infrastructures) can be regarded as a collection of inhomogeneously and generically interacting units.
A protein interaction network is the set of binary interactions among the proteins of a given proteome.

- **Nodes**: proteins
- **Links**: physical interactions (binding)
Social networks

- **Nodes**
  - Individuals

- **Edges**
  - Relationship
  - Communications
  - Interactions

Social network analysis
Many complex systems (from the molecular level to the scale of large communications infrastructures) can be regarded as a collection of inhomogeneously and generically interacting units.
Airport network

- Each edge is characterized by weight $wij$ defined as the number of passengers in the year.
Physical Internet

- Computers (routers)
- Satellites
- Modems (??)
- Phone cables
- Optic fibers
- Wireless
The World-Wide Web

- Virtual network to find and share information
  - Web pages
  - Hyperlinks
What’s new ??

- Euler circa 1736 (Koningsberg problem)
- Moreno ’34 (sociogram) – social
- Erdos ’60-70 (random graph theory) – math

.........
Size does matter !!!

- Starting in ’96 large scale internet and web measurements
- High throughput experiments in Biology
- Electronic databases and indexing
Networks sizes...

- **Biology**
  - Genome(s), regulatory networks, metabolic networks, protein interaction networks ($10^3$-$10^4$ nodes)

- **Social network**
  - Co-authorship, citations, patents, grants, e-mails, P2P, instant messaging $\ldots \ldots (10^3$ – $10^7$ nodes)

- **Physical Internet**
  - ISP level ($10^4$ nodes), Router level ($10^5$-$10^6$), Host level ($10^7$ nodes)

- **WWW**
  - Web pages (Url address $10^8$ nodes)
What’s new....

- Data size shifts ($10^2 \rightarrow 10^8$ elements)
  (Complexity??)

- Different domains (biology, info-structures, infrastructures, social, scientometrics)
  (universality ??)

- Large scale longitudinal studies (time series)
  (dynamical modeling??)
Complexity = very complicated

Complications at all scales (compatible with the finite world)

- Scale invariance \rightarrow \text{geometric Fractals}
- Infinite susceptibility/response
- Diverging fluctuations

\textbf{Complex Features}

(symptoms)
Heavy tailed statistical distribution (degree etc.)

\[ P(k) = \text{probability that a node has } k \text{ links} \]

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

- \( \langle k \rangle = \text{const} \)
- \( \langle k^2 \rangle \to \infty \)

Scale-free properties ↔ Diverging fluctuations
Ubiquity/universality (weak)

- Metabolic
- Airport
- E-mail
(more) Complex features

- Small-world + clustering
- Assortativity/correlations
- Community structure
- Motifs
- ..............
The Internet growth

- 1997
  - 3112 AS
- 2000
  - 9107 AS

- In 1999:
  - 3410 new AS
  - 1713 lost AS

Self-organizing and evolving system
Main ingredients for complex systems

- Many interacting units
- Dynamical evolution
- Self-organization

Supervising entity

Project/blueprint

Non-trivial architecture
Unexpected emergent properties
Cooperative phenomena

Complexity features
Undirected graph $G(V, \varepsilon)$

$|V| = N = \# \text{ vertices}$

$|\varepsilon| = \# \text{ of edges} = E = \text{undirected pairs of vertices } (i,j)$

Max $E \rightarrow \frac{N(N-1)}{2}$

**Complete $N$-graph**
Graph

Directed graph (digraph)
- Given a single graph
  - Vertices/nodes/elements/individuals...
  - Edges/link/relations/connections
- **Adjacency matrix** (sociomatrix/connectivity matrix)
  \[ X = \{ x_{ij} \} \]
Network observables

- Size (order) $N$
- Degree $k_i$
- Average degree $K = \frac{1}{N} \sum_i k_i \equiv \frac{2E}{N}$
- Degree sequence $\{k_1, k_2, \ldots, k_N\}$
- Connected triples…

$z_i (X) = \text{statistical observables}$
Shortest paths = minimum (# hops) between two nodes

Regular lattice with $N = 10^4$ $d \sim 10^2$

Small world with $N = 10^4$ $d \sim \ln N$
Small world properties

Distribution of Shortest paths (# hops) between two nodes

Average fraction of nodes within a shortest path of length $d$
Statistical distributions for centrality measures

- A statistical look at the hierarchies in terms of connections, traffic flows etc.

Degree centrality
Heavy tailed statistical distribution (degree etc.)

\( P(k) = \) probability that a node has \( k \) links

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

- \( \langle k \rangle = \text{const} \)
- \( \langle k^2 \rangle \to \infty \)

Scale-free properties \quad \leftrightarrow \quad \text{Diverging fluctuations}
Non-local measure of centrality

Betweenness centrality $= \#$ of shortest paths traversing a vertex or edge (flow of information) if each individuals send a message to all other individuals
Betweeenness Probability distribution

Heavy-tailed and highly heterogeneous
- Skewed
- Heterogeneity and high variability
- Very large fluctuations (variance >> average)
- Various fits: power-law + cut-off; Weibull etc.
Multi-point correlations

\( P(k,k') \) ??

Average nearest neighbors degree

\[
k_{nn}(i) = \frac{1}{k_i} \sum_j k_j
\]

Correlation spectrum:

Average over degree classes

\(< k_{nn}(k) >\)
\( \langle k_{nn}(k) \rangle = \sum_{k'} k' p(k'|k) \) 

Degree correlation function

- **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
Degree correlation function

\[ \langle k_{nn}(k) \rangle = \sum_{k', k} k' p(k' | k) \]

Highly degree ASs connect to low degree ASs
Low degree ASs connect to high degree ASs

No hierarchy for the router map
Average connectivity of a neighbor of a node with connectivity $k$

Highly degree proteins connect to low degree proteins

Low degree proteins connect to high degree proteins

Clustering coefficient = connected peers will likely know each other

\[ C = \frac{\text{# of links between } 1, 2, \ldots, k \text{ neighbors}}{k(k-1)/2} \]

Higher probability to be connected
$a_{ij}$: Adjacency matrix

\[ c_i = \frac{1}{k_i(k_i - 1)} \sum_{j \neq k} a_{ij}a_{ik}a_{jk} \]

\[ C = \langle c_i \rangle \quad \text{Average clustering} \]

\[ C(k) = \frac{\sum_i \delta(k_i - k)c_i}{\sum_i \delta(k_i - k)} \quad \text{Clustering spectrum} \]
Clustering Spectrum in the Internet

Clustering coefficient as a function of the vertex degree

$$C(k) = \frac{\sum_i \delta(k_i - k)c_i}{\sum_i \delta(k_i - k)}$$

Highly degree ASs bridge not connected regions of the Internet
Low degree ASs have links with highly interconnected regions of the Internet

No hierarchy at the router level
K-core decomposition

1-core

2-core

3-core

K-shell

K-shell

K-shell
Network statistics

- Random processes defined by statistical ensambles: To each graph realization $X$ corresponds probability $P(X)$

\[
X = \begin{pmatrix}
0 & x_{12} \\
x_{21} & 0
\end{pmatrix}
\]
Logit models (exponential random graphs)

- Probability of a given realization $X$

$$P(X) = \frac{\exp(\sum_i \theta_i z_i(X))}{\kappa(\theta_i)}$$

- $\theta_i =$ set of model parameters
- $\kappa(\theta_i) =$ normalization factor

Parameters $\theta_i$ to be estimated from the real data

(vast traditions in the social literature/ large amount of techniques)
Equivalence with the statistical mechanics of networks (Gibbs/Boltzmann)

\[ P(X) = \frac{\exp(\sum_i \theta_i z_i(X))}{\kappa(\theta_i)} \]

\[ P_{eq} = \frac{\exp(-\beta H(X))}{Z} \]

Statistical mechanics follows from the maximum entropy principle. The exponential family can be obtained in the very same way.
Maximum entropy principle

- $P(X)$ is given by the distribution that maximizes the Gibbs entropy:

$$S[P] = - \sum_x P(X) \ln P(X)$$

Entropy is a measure of the disorder encoded in the probability distribution, we want to maximize the statistical disorder.

According to the constraints imposed by the statistical observables:

$$\langle z_i \rangle = \sum_x P(x) z_i(x),$$
Lagrange multipliers

- The solution is given by the solution of

$$\frac{\partial}{\partial P(x)} \left[ S[P] + \alpha_0 \left( 1 - \sum_x P(x) \right) + \sum \alpha_i \left( \langle z_i \rangle - \sum_x P(x)z_i(x) \right) \right] = 0$$

Where $\alpha_i$ are the lagrange multipliers
Gibbs solution

\[
P(X) = \frac{\exp\left(-\sum_i \alpha_i z_i(X)\right)}{Z(\alpha_i)},
\]

With normalization condition

\[
Z(\alpha_i) = e^{\alpha_0 + 1} = \sum_x e^{-\sum_i \alpha_i z_i(x)}.
\]

It is the same of the exponential family

\[
P(X) = \frac{\exp\left(\sum_i \theta_i z_i(X)\right)}{\kappa(\theta_i)}
\]
Exponential models

= Equilibrium Stat. mechanics

- Model parameters = potentials

- Statistical Ensemble
  (microcanonical/canonical/grandcanonical) = edge and vertices constraints

- Partition function => Free Energy
Maximum entropy approach

- Solid statistical foundation
- Rigorous formalism
- Clear set of assumptions (statistical observables)
- Long traditions (social, statistics, physics..)
A different perspective: dynamics

Dynamical evolution governing the evolution of the system from one Configuration to the other

\[ \mathbf{X} = \begin{pmatrix} 0 & x_{12} \\ x_{21} & 0 \end{pmatrix} \]

Ultimately we look for the stationary state: \( t \to \infty \)
The focus shifts on \[ W_{(x\rightarrow y)} \]

i.e. the dynamical rules governing the evolution from one configuration
to the other (transition rates/probability)
“Dynamical approach”

- The dynamical evolution is followed.

- In “regular” situation the stationary state
  \[ P_s(x) = \lim_{t \to \infty} P(x,t) \]
  is the equilibrium one.
Detailed balance condition

\[ P(y)w(y \rightarrow x) = P(x)w(x \rightarrow x) \forall y, x. \]

\[
\frac{w(y \rightarrow x)}{w(x \rightarrow y)} = \frac{P(x)}{P(y)} = \exp(\sum \theta_i[z_i(x) - z_i(y)]),
\]
Nonequilibrium

- Non-equilibrium systems

- Enlarging phase space (growing networks), non-stationary states, currents and fields.
Dynamical approach (con’s)

- Less rigorous

- Guesses and approximations on the dynamics and the solutions

- “Somehow” statistical observable are used a posteriori
Dynamical approach (pro’s)

- Very intuitive/technically easier
- Extremely suitable for large scale simulations/monte-carlo approaches
- In non-equilibrium cases is the only viable approaches.
- Asymptotic and universality
Universality

- Some common symmetries or principles in the dynamical rules generate “large-scale” properties shared by large classes of systems (in some case very different one from each other)

Example: Phase transitions
Universality 2

- In network science it is not clear the level of universality

- Many does not believe to that and predicate the importance of the precise details of the structure.

N.B. “Universality” does not implies similarity. Only large scale properties are universal (other properties may be very relevant and different e.g. liquid/gas vs ferromagnetic material.)
Erdös-Rényi model
(1960)

With probability $p$ an edge is established among couple of vertices

$$\langle k \rangle = p(N-1)$$

Poisson distribution
Degree Distribution

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

\[ N \to \infty; \quad p \to 0; \quad pN \to \langle k \rangle \]

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

Probability that

\[ \langle c \rangle = p = \frac{\langle k \rangle}{N} \]

Small World

\[ \langle \ell \rangle \approx \frac{\log N}{\log \langle k \rangle} \]
<table>
<thead>
<tr>
<th>Code of the network</th>
<th>$N$</th>
<th>$E$</th>
<th>$&lt;k&gt;$</th>
<th>$&lt;l&gt;$</th>
<th>$\gamma$</th>
<th>$&lt;C&gt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMCN</td>
<td>375</td>
<td>16,248</td>
<td>43.33</td>
<td>2.05</td>
<td>3.46</td>
<td>0.260</td>
</tr>
<tr>
<td>WAN</td>
<td>3,880</td>
<td>18,810</td>
<td>9.70</td>
<td>4.37</td>
<td>2.00</td>
<td>0.53</td>
</tr>
<tr>
<td>SCN</td>
<td>12,722</td>
<td>39967</td>
<td>6.28</td>
<td>6.8</td>
<td>---</td>
<td>0.65</td>
</tr>
<tr>
<td>WWW Altavista</td>
<td>203,549,046</td>
<td>2,130,000,000</td>
<td>10.46</td>
<td>16.18</td>
<td>2.10-2.70</td>
<td>---</td>
</tr>
<tr>
<td>Internet</td>
<td>10,697</td>
<td>31,992</td>
<td>5.98</td>
<td>3.31</td>
<td>2.50</td>
<td>0.035</td>
</tr>
<tr>
<td>Film actors</td>
<td>449,913</td>
<td>25,516,482</td>
<td>113.43</td>
<td>3.48</td>
<td>2.30</td>
<td>0.200</td>
</tr>
<tr>
<td>Metabolic network</td>
<td>765</td>
<td>3686</td>
<td>9.64</td>
<td>2.56</td>
<td>2.20</td>
<td>0.090</td>
</tr>
</tbody>
</table>
Small-world model

\[ <c> = \frac{3m(m-1)}{2m(2m-1)} (1-p)^3 \]

Watts & Strogatz 98
# shortcuts = $pN$

# of vertices for each shortcut $\cong \frac{N}{pN} \cong p^{-1}$

Size of regions

If $S \ll N$ many far away regions are connected

$N^{-1} \ll p \ll 1$

Barthelemy, Amaral 1999
What about the degree distribution?

Heavy tails?

Static construction

Molloy-reed
Position model
Hidden variables
Etc.

Generalized random graphs with pre-assigned degree distribution
Classical topology generators

- Waxman generator
- Structural generators
  - Transit-stub
  - Tiers

Scale-free topology generators

- INET (Jin, Chen, Jamin)
- BRITE (Medina & Matta)

Modeling of the Network structure with ad-hoc algorithms tailored on the properties we consider more relevant
Shift of focus:

Static construction  $\rightarrow$  Dynamical evolution

Direct problem

Evolution rules  $\rightarrow$  Emerging topology

Inverse problem

Given topology  $\rightarrow$  Evolution rules
The rich-get-richer mechanism

(Barabasi & Albert 1999)

**Growth:** The network starts with a small core of \( m_0 \) connected vertices. Every time step we add a new vertex, with \( m \) edges (\( m < m_0 \)) connected to old vertices in the system.

**Preferential attachment:** The new edges are connected to the old \( s \)-th vertex with a probability proportional to its degree \( k_s \).

\[
\Pi[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)}
\]

\[
P(k) \sim k^{-3}
\]
Theory of evolving networks

Master equations and its projections……..

Probability

$p(k, s, t)$ that a vertex introduced at time $s$ has degree $k$ at the time $t \geq s$. Once known the probability $p(k, s, t)$, we can obtain the average degree $k_s(t)$ of the $s$-th vertex at time $t$ as

$$k_s(t) = \sum_{k=0}^{\infty} kp(k, s, t),$$

and the degree distribution at time $t$ (i.e. for a network of size $N = t + m_0$) using the expression

$$P(k, t) = \frac{1}{t + m_0} \sum_{s=0}^{t} p(k, s, t).$$

Stationary State

$$P(k) = \lim_{t \to \infty} P(k, t)$$
Continuous approximations

Average degree value \( k_s(t) \) that the node born at time \( s \) has a time \( t \)

\( \{ p(k, s, t) = \delta(k - k_s(t)) \} \)

**Evolution equation**

\[
\frac{\partial k_s(t)}{\partial t} = m \Pi [k_s(t)]
\]

**Degree distribution**

\[
P(k, t) = \frac{1}{t + m_0} \int_0^t \delta(k - k_s(t)) \, ds \equiv - \frac{1}{t + m_0} \left( \frac{\partial k_s(t)}{\partial s} \right)^{-1} \bigg|_{s=s(k,t)}
\]
Preferential Attachment construction

Attachment kernel

\[ \Pi[k_s(t)] = \frac{k_s(t)}{\sum_j k_j(t)} \]

\[ k_s(t) \simeq m \left( \frac{t}{s} \right)^{1/2} \]

\[ k_{s'}(t)/k_s(t) \simeq (s/s')^{1/2} \]

- Degree distribution

\[ P(k, t) = 2m^2 \frac{t + \frac{m_0}{m} \langle k \rangle_0}{t + m_0} k^{-3} \quad t \to \infty \]

\[ P(k, t) = 2m^2 k^{-3} \]
Illustration of the rules of the copying model. A prototype vertex (black dot surrounded by a circle) is selected and a new vertex (hollow dot) is created with virtual edges pointing to the neighbors of the prototype. With probability $1 - \alpha$ the virtual edges are kept; with probability $\alpha$ they are rewired to a randomly chosen vertex.

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

Degree distribution

\[ P(k_{in}) \sim k_{in}^{-(2-\alpha)/(1-\alpha)} \]
• Single genes, sometime the entire genome, can be passed to the next generation in two copies.

• After duplication, the two copies of the same gene can follow different evolutionary paths. They both can undergo mutations.

• If the mutation is beneficial, it can be evolutionary preserved and further transmitted to future generations.

IS THIS CONSISTENT WITH WHAT WE OBSERVE ON PIN?
At time step $t$ a random node is selected for duplication...

Now, any neighbor of “mother” node has also a link with the “son” node.

The link with the “mother” (or the “son”) node is lost with prob. $q$, with prob. $1-q$ both are kept.

A new link “mother-son” is established with prob. $p$
Iterate the above until the wanted number of nodes is achieved.
Model validation......

- Correlations
- Clustering
- Hierarchies (k-cores, modularity etc.)
- ............
Degree hierarchy

Clustering hierarchy
E-R model
More models

• Generalized BA model

  Non-linear preferential attachment: $\Pi(k) \sim k^\alpha$

  (Redner et al. 2000)

  Initial attractiveness: $\Pi(k) \sim A + k^\alpha$

  (Mendes & Dorogovstev 2000)

  Rewiring

  (Albert et al. 2000)

  • Highly clustered

  (Eguiluz & Klemm 2002)

  • Fitness Model

  (Bianconi et al. 2001)

  • Multiplicative noise

  (Huberman & Adamic 1999)
Heuristically Optimized Trade-offs (HOT)

Papadimitriou et al. (2002)

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
Effect of complex network topologies on physical processes

- Epidemic models
- Resilience & robustness
- Avalanche and failure cascades
- Search and diffusion
Stages of an epidemic outbreak

Infected individuals => prevalence/incidence
Compartmental models

Homogeneous mixing assumption
The Susceptible-Infected-Susceptible (SIS) model

- Each node is infected with rate $\nu$ if connected to one or more infected nodes.
- Infected nodes are recovered (cured) with rate $\delta$ without loss of generality $\delta = 1$ (sets the time scale).
- Definition of an effective spreading rate $\lambda = \nu / \delta$.

Dynamical Mean-Field equation for the order parameter $\rho = \text{density of infected nodes}$

\[ \partial_t \rho(t) = -\rho(t) + \lambda \langle k \rangle \rho(t) [1 - \rho(t)] + h.o. \]

In the stationary state $\partial_t \rho = 0$, we have

\[ \rho [-1 + \lambda \langle k \rangle (1 - \rho)] = 0 \]

Definition of the epidemic threshold $\lambda_c = \langle k \rangle^{-1}$

\[
\begin{align*}
\rho &= 0 & \text{if } \lambda < \lambda_c, \\
\rho &\sim \lambda - \lambda_c & \text{if } \lambda > \lambda_c,
\end{align*}
\]
the reproduction rate $R = \lambda \langle k \rangle \geq 1$ defines the threshold

Similar models with immunity and death (removal) can be defined (SIR etc)

The epidemic threshold is a general result

The question of thresholds in epidemics is central
Immunization

- Random immunization: 
  \[ g = \text{density of immune nodes} \]

Regular or random networks

\[ \lambda \rightarrow \lambda (1-g) \]

Epidemic dies if 
\[ \lambda (1-g) \leq \lambda_c \]

\[ \rho_g = \rho_0 (g_c - g) / g_c \]

Immunization threshold

\[ g_c = (\lambda - \lambda_c) / \lambda \]
Natural computer virus
  • DNS-cache computer viruses
  • Routing tables corruption

Data carried viruses
  • ftp, file exchange, etc.

Computer worms
  • e-mail diffusing
  • self-replicating

Internet topology

E-mail network

Ebel et al. (2002)
Strain data analysis

We analyzed homogeneous groups of viruses
- effective parameters
- similar properties within the strain

- In green MACRO viruses
- In red FILE viruses
- In blue BOOT viruses
• Survival probability
\[ P_s(t) = \text{fraction of viruses still in the wild at time } t \text{ after their birth} \]

\[ P_s(t) \sim \exp \left( -\frac{t}{\tau} \right) \]

\[ \tau = \text{average lifetime (characteristic time) of the virus strain} \]

**Average lifetime extremely long compared to the virus rates time scale**

• Endemic metastable state
• Above the critical threshold ??
Prevalence from viruses in the wild

- Anti-virus software is delivered in a few hours after the first detection
- Immunization of the system should be achieved
- ILoveYou virus is still present in the wild list

In the endemic case, prevalence is always very small ($\rho<0.01$) but stationary for long period.
Long lifetime + low prevalence = computer viruses always tuned infinitesimally close to the epidemic threshold

Immunization always brings the virus just in the proximity of the threshold

According to S. White in the 90s
Epidemic spreading on Scale-Free networks

- Highly connected nodes are statistically significant $<k^2>=\infty$

- Connectivity fluctuations must be included

Relative density $\rho_k(t)$ of infected nodes with given connectivity $k$

$$\partial_t \rho_k(t) = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \Theta(\rho(t)),$$

$\Theta(\rho(t))$ is the probability that any given link points to an infected node.

- $\Theta$ is a function of the average density of infected nodes
- Links point with higher probability to highly connected nodes

[ Pastor Satorras & Vespignani, PRL 86, 320(2001)]
In the stationary state $\partial_t \rho_k(t) = 0$ we have that $\Theta(\rho) \Rightarrow \Theta(\lambda)$.

MF equations yield

$$
\rho_k = \frac{k\lambda \Theta(\lambda)}{1 + k\lambda \Theta(\lambda)}.
$$

**Stationary state**

**Simulations in a BA network**

- Higher is the node connectivity and higher is the probability to be in an infected state
- Strong heterogeneity
The average density of infected in contact with a vertex of degree $k$

\[ \Theta_k[\{\rho_{k'}(t)\}] = \sum_{k'} P(k' | k) \rho_{k'}(t) \]

\[ \frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \sum_{k'} P(k' | k) \rho_{k'}(t) \]

Uncorrelated networks

- A link is more likely connected to a node with high connectivity

- The probability that a link points to a node with $s$ links is proportional to $sP(s)$.

\[ P(k' | k) = \frac{k' P(k')}{\sum_{k'} k'P(k')} \]

Independent on $k$
Uncorrelated case

$$\Theta(\lambda) = \frac{1}{\langle k \rangle} \sum_k k P(k) \rho_k(\lambda)$$

Stationary state self-consistent equation

$$\rho_k = \frac{k \lambda \Theta(\lambda)}{1 + k \lambda \Theta(\lambda)}$$

$$\Theta = \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{\lambda k \Theta}{1 + \lambda k \Theta}$$
Vanishing epidemic threshold in the thermodynamic limit

\[
\begin{align*}
\frac{d}{d\Theta} \left( \frac{1}{\langle k \rangle} \sum_k k P(k) \frac{\lambda k \Theta}{1 + \lambda k \Theta} \right) \bigg|_{\Theta=0} & \equiv \lambda \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 1 \\
\lambda_c & = \frac{\langle k \rangle}{\langle k^2 \rangle} \\
\text{if } \langle k^2 \rangle & \to \infty \text{ then } \lambda_c \to 0
\end{align*}
\]
Absence of any epidemic threshold (critical point)

Active state for any value of $\lambda$

The infection pervades the system whatever spreading rate

In infinite systems the infection is infinitely persistent (indefinite stationary state)

Pastor-Satorras & Vespignani, PRL 86, 3200(2001)
Epidemic prevalence in uncorrelated graphs with generic Power-law tail

\[ \rho(\lambda) = \sum_k P(k) \rho_k(\lambda) \]

a) \( 2 < \gamma < 3 \)

\[ \Theta(\lambda) \approx (m\lambda)^{(\gamma-2)/(3-\gamma)} \]

\[ \rho(\lambda) \sim \lambda^{1/(3-\gamma)} \]

b) \( \gamma = 3 \)

\[ \Theta(\lambda) = \frac{e^{-1/m\lambda}}{\lambda m} (1 - e^{-1/m\lambda})^{-1} \]

\[ \rho(\lambda) \sim 2e^{-1/m\lambda} \]

c) \( 3 < \gamma < 4 \)

\[ \rho(\lambda) \sim \left( \lambda - \frac{\gamma - 3}{m(\gamma - 2)} \right)^{1/(\gamma-3)} \]

d) \( \gamma > 4 \)

\[ \rho(\lambda) \sim \lambda - \frac{\gamma - 3}{m(\gamma - 2)} \]
\[ \rho = 2e^{-1/m\lambda} \]

**Numerical simulations in a BA network**

**Zoom in lin-log scale**

**Network sizes**

\( N=10^3 \) to \( N=10^7 \)
Finite size scale-free networks

\[ P(k) \sim k^{-\gamma} \exp\left(-\frac{k}{k_c}\right) \]

Exponentially bounded

\[ P(k) \sim k^{-\gamma} \theta(k - k_c) \]

Hard cut-off

\[ \lambda_c \sim k_c^{\gamma - 3} \]

Ratio between SF and homogeneous Epidemic threshold for \( k_c, N = 10^4 \)

\[ \frac{\lambda_c}{\lambda_c^H} < 10^{-1} \]
Rationalization of computer virus data

- Wide range of spreading rate with low prevalence (no tuning)
- Lack of healthy phase = standard immunization cannot drive the system below threshold!!!
Random immunization in S-F networks

Epidemic dies if $\lambda (1 - g) \leq \lambda_c$

Scale-free networks $\Rightarrow \lambda_c = 0$

**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 PRE 2002; Cohen et al. PRL91, 247901 (2003)]
BA network

Real AS Internet map
\[ < k_{nn}(k) > = \sum_k k' p(k'|k) \]

**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
Exact result:

\[ \langle k^2 \rangle \rightarrow \infty \] is a sufficient conditions for the absence of any epidemic threshold in unstructured networks with any degree correlations

\[ \langle k_{nn} \rangle = \sum_k k_{nn}(k)P(k) \rightarrow \infty \]

Result related to the divergence of the average nearest neighbors connectivity

SIR model

\[ \frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k S_k(t) \Theta(t) \]
\[ \frac{dS_k(t)}{dt} = -\lambda k S_k(t) \Theta(t), \]
\[ \frac{dR_k(t)}{dt} = \rho_k(t). \]

\[ \Theta(t) = \frac{1}{\langle k \rangle} \sum_k (k - 1) P(k) \rho_k(t) \]

Where

\[ R_\infty = \sum_k P(k) (1 - e^{-\lambda k \phi_\infty}) \]

\[ \phi(t) = \int_0^t \Theta(\tau) \, d\tau = \frac{1}{\langle k \rangle} \sum_k (k - 1) P(k) R_k(t) \]
SIR model

\[ \phi_\infty = 1 - \frac{1}{\langle k \rangle} \frac{1}{\langle k \rangle} \sum_k (k - 1) P(k) e^{-\lambda k \phi_\infty} \]

\[ \frac{d}{d\phi_\infty} \left( 1 - \frac{1}{\langle k \rangle} - \frac{1}{\langle k \rangle} \sum_k (k - 1) P(k) e^{-\lambda k \phi_\infty} \right) \bigg|_{\phi_\infty = 0} \geq 1 \]

\[ \lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \]

if \[ \langle k^2 \rangle \rightarrow \infty \] then \[ \lambda_c \rightarrow 0 \]
Growth rate

\[ \rho(t) \propto \exp\left(\frac{t}{\tau}\right) \]

\[ \tau = \frac{\langle k \rangle}{\lambda \langle k^2 \rangle - 2 \langle k \rangle} \rightarrow 0 \]

Initial conditions

\[ \kappa_0 > k^{\text{io}} > 20 \]
(more) Implications/applications

- Spreading, Diffusion and Finding
  - Epidemic modeling
  - Behavioral patterns
  - Competition/evolution
  - Search algorithm
  - Ranking algorithms

- Resilience and robustness
  - Avalanche
  - Congestions
  - Adaptive control
Books (scientific monographs)

Reviews