Mini course on Complex Networks

Massimo Ostilli[1]

[1]UFSC, Florianopolis, Brazil

September 2017
Dep. de Fisica

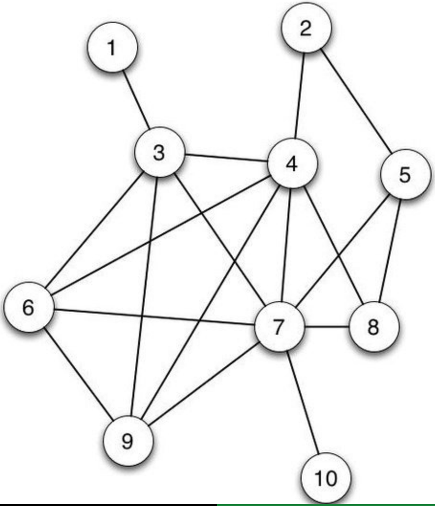
Organization of The Mini Course

- Day 1: Basic Topology of Equilibrium Networks
- Day 2: Percolation and Magnetism
- Day 3: Growing Networks

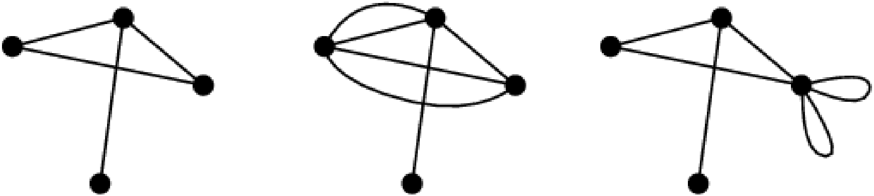
First of all

- What is a graph? Preliminary Definitions
- A historical example

Formally, a graph G is a pair of sets $G = (V, E)$, where V is a set of $N = |V|$ nodes, and E a set of $L = |E|$ edges (or links)



G is called Simple if there are no multiple links and no self-links

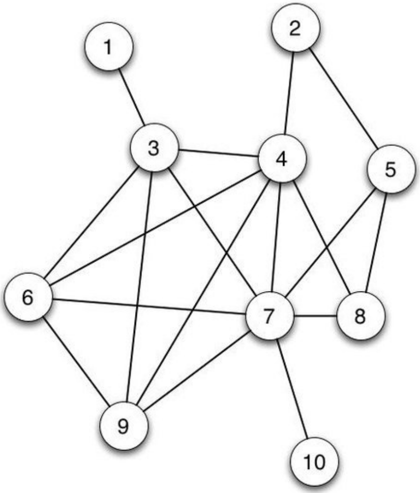


simple graph

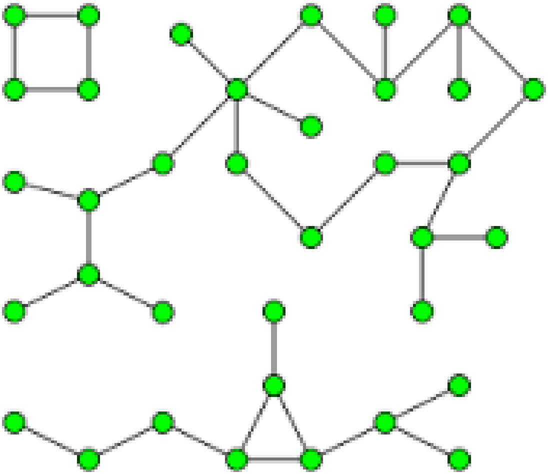
nonsimple graph with multiple edges

nonsimple graph with loops

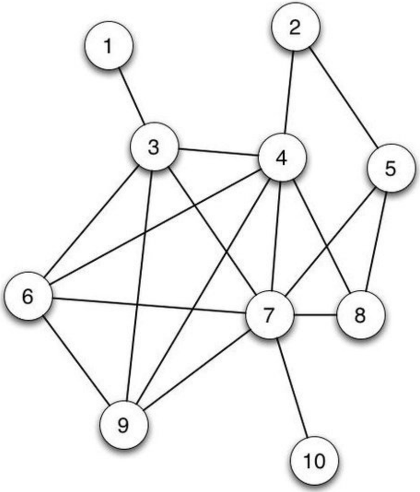
G is Connected if any node can be reached from any other node via a path of links



G is Disconnected if it is Not Connected

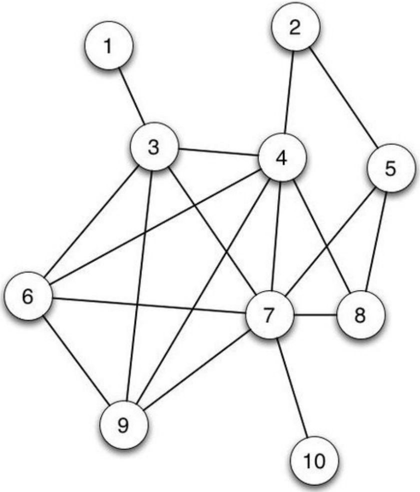


Node-degree k



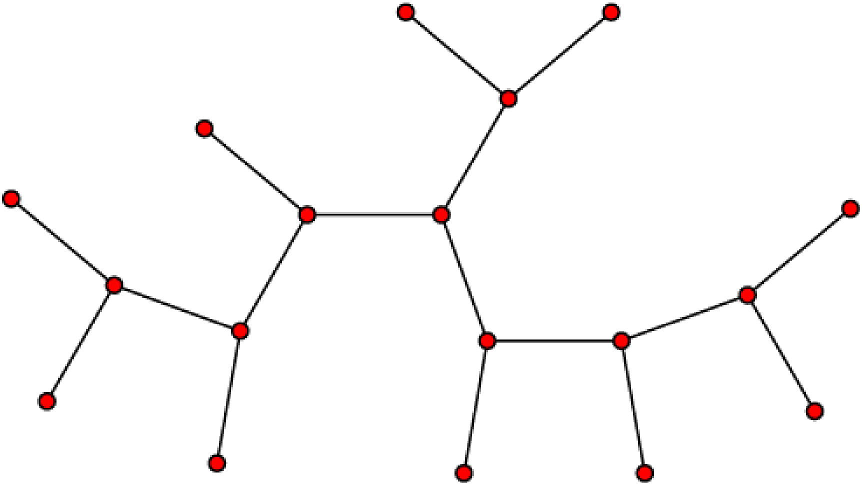
$$k_1 = 1,\ k_2 = 2,\ k_3 = 4, \ldots$$

Mean-degree (or connectivity of G) $\langle k \rangle$

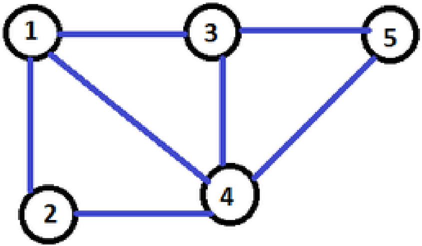


$$\langle k \rangle = \sum_i k_i / N = 2L/N \qquad N = 10,\ L = 17 \Rightarrow 2L/N = 3.4$$

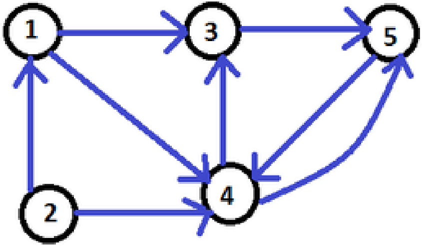
G is a Tree if there are No Loops



Undirected and Directed Graphs



Undirected Graph

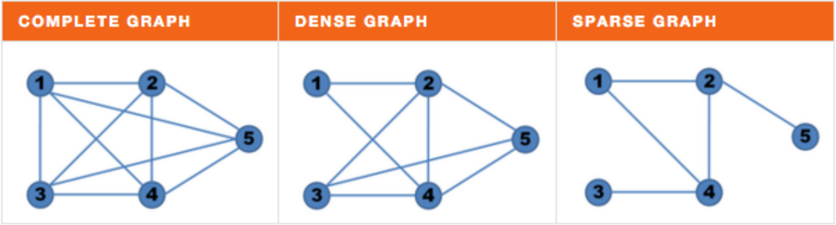


Directed Graph

$k_4 = 4$

$k_4^{IN} = 3, \quad k_4^{OUT} = 2$

G is sparse if $L = O(N)$
G is dense if $L = O(N^\alpha)$ with $\alpha > 1$



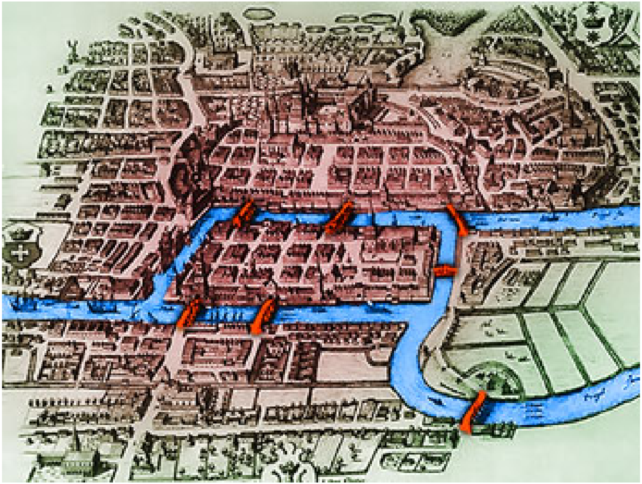
To summarize

- G is a pair of sets $G = (V, E)$, where V is a set of $N = |V|$ nodes, and E a set of $L = |E|$ edges (links)
- G is called simple if there are no multiple links and no self-links
- G is called sparse if $L = O(N)$
- G is called dense if $L = O(N^\alpha)$ with $\alpha > 1$
- The degree k_i (or connectivity) of a node i, is the number of links emanating from the node $\Rightarrow$

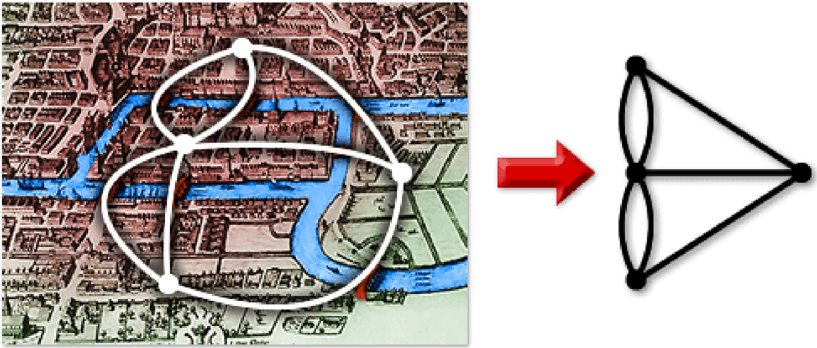
$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \frac{2L}{N},$$

for the moment $\langle \cdot \rangle$ refers to the mean over a single G

Why studying graphs: An historical example



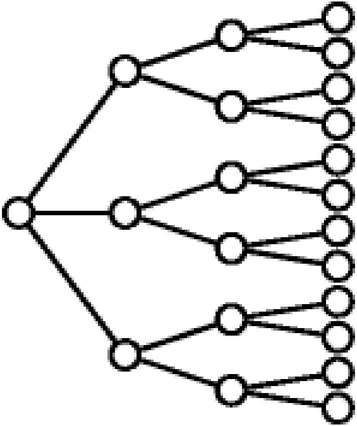
Why studying graphs: An historical example



Day 1: Basic Topology of Equilibrium Networks

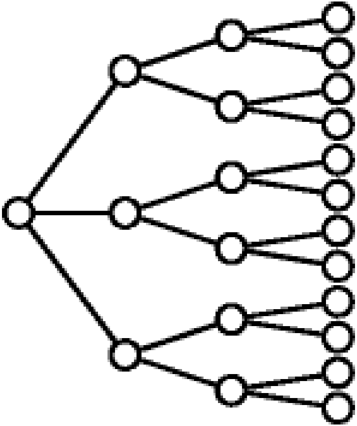
- Cayley Trees and Bethe Lattices
- Adjacency Matrix
- Main Graph Metrics: $P(k)$; $\langle C \rangle$; $\langle \ell \rangle$
- The Random Graph Model
- The Configuration Model
- Main Graph Metrics: Simple Evaluations for locally Tree-like nets

Cayley Tree = Finite “Regular” Tree



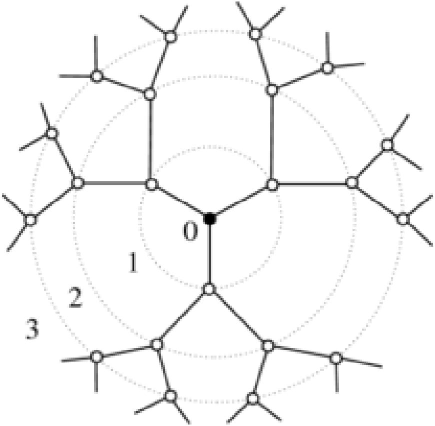
Here $q = 3$ and $\langle k \rangle = ?$

Cayley Tree = Finite "Regular" Tree



Here $q = 3$ and $\langle k \rangle = 2 - 2/N$ (holds for any finite tree)

Bethe Lattice = Infinite Regular Tree



Here $q = 3$ and $\langle k \rangle = q$

Adjacency Matrix $\boldsymbol{a}$

Any graph G can be encoded via the Adjacency Matrix $\boldsymbol{a}$.
We label the nodes by an index $i = 1, \ldots, N$

$$a_{i,j} = \begin{cases} 1, & \text{if there a link between } i \text{ and } j, \\ 0, & \text{otherwise} \end{cases}$$

In other words $G = (V, E) \equiv \boldsymbol{a}$

Adjacency Matrix $\boldsymbol{a}$

The degree of the vertex i

$$k_i = \sum_j a_{i,j}, \quad L(G) = \frac{1}{2}\sum_{i,j} a_{i,j} = \frac{N\langle k\rangle}{2},$$

The number of triangles passing through the vertex i

$$N_T(i) = \sum_{j,k} a_{i,j}a_{j,k}a_{k,i}, \quad N_T(G) = \frac{1}{3}\mathrm{Tr}(\boldsymbol{a}^3)$$

The number of non self overlapping paths of length ℓ passing between i and j

$$N_{Paths}(i,j;\ell) \sim (\boldsymbol{a}^\ell)_{i,j} + O((\boldsymbol{a}^{\ell-1}))$$

Random Matrix Theory approach...

Degree Distribution $P(k)$

$N(k)$ is the number of nodes with degree k

$$P(k) = \frac{N(k)}{N},$$

In a Regular d-dimensional Lattice

$$P(k) = \delta_{k,2d},$$

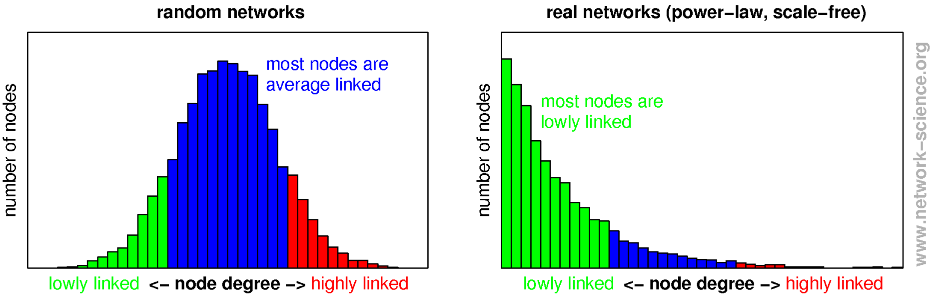
In the "Random Graph" (classical)

$$P(k) = \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle},$$

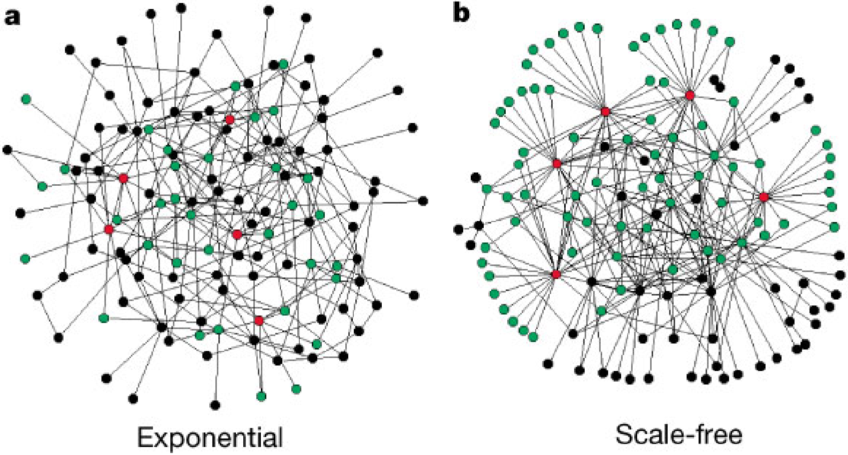
In "Complex Networks" (typically, for large k)

$$P(k) \sim k^{-\gamma}, \gamma > 2$$

Compare Random and Scale-Free Complex Network



Compare Random and Scale-Free Complex Network



Note that, if $P(k) \sim k^{-\gamma} \Rightarrow$

$$\sum_{k=1}^{N} P(k)k^{\alpha} \sim \int_{1}^{N} dk\, P(k)k^{\alpha},$$

$\Rightarrow$

$$\begin{array}{lll}
\lim_{N\to\infty} \langle k \rangle < \infty \quad \text{and} & \lim_{N\to\infty} \langle k^2 \rangle - \langle k \rangle^2 < \infty & \gamma > 3, \\
\lim_{N\to\infty} \langle k \rangle < \infty \quad \text{and} & \lim_{N\to\infty} \langle k^2 \rangle - \langle k \rangle^2 = \infty & 2 < \gamma < 3, \\
\lim_{N\to\infty} \langle k \rangle = \infty \quad \text{and} & \lim_{N\to\infty} \langle k^2 \rangle - \langle k \rangle^2 = \infty & \gamma < 2.
\end{array}$$

Most of the complex networks observed in nature and technology have $\gamma \leq 3$ and very often $\gamma \simeq 2$.

Link-Degree Distribution (or Excess Distribution) $P_L(k)$

$P_L(k) = Prob.\,(\text{that the end of a link points to a node of degree } k)$

It is simple to see that

$$P_L(k) = \frac{P(k)k}{\langle k \rangle}$$

We define also

$$P_L(k,k') = Prob.\,(\text{that the two ends of a link point to a node of degree } k \text{ and to a node of degree } k', \text{ resp.})$$

A graph G is called Uncorrelated if $P_L(k,k') = P_L(k)P_L(k')$.

Clustering Coefficient $\langle C \rangle$

$$C(i) = Prob.\ \text{(that between two neighbors of node } i \text{ there is a link)}$$

$$\langle C \rangle = \frac{1}{N} \sum_i C(i) \quad \text{Average Clustering Coefficient}$$

$$\langle C \rangle = 0, \quad \text{Trees and Bethe Lattices}$$

$$\langle C \rangle = \mathrm{O}\left(N^{-1}\right), \quad \text{Random Graph}$$

$$\langle C \rangle = \mathrm{O}\left(N^{-\alpha}\right), \alpha < 1 \quad \text{Uncorrelated Complex Networks}$$

$$\langle C \rangle = \mathrm{O}\left(1\right), \quad \text{Lattices and Strongly Correlated Complex Networks}$$

$$\langle C \rangle = 1, \quad \text{Complete Graph}$$

Most of the complex networks observed in nature and technology have a small but not negligible $\langle C \rangle$.

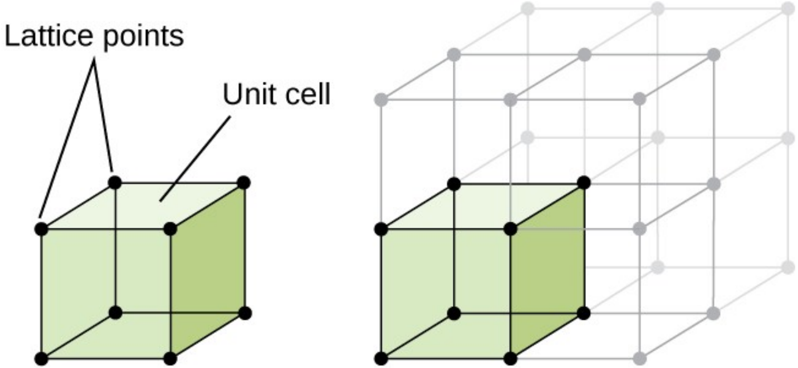
Average Path Length $\langle \ell \rangle$

Given G, let $\ell_{i,j}$ the length (*i.e.*, the number of links), of the shortest path between i and j. Their mean is

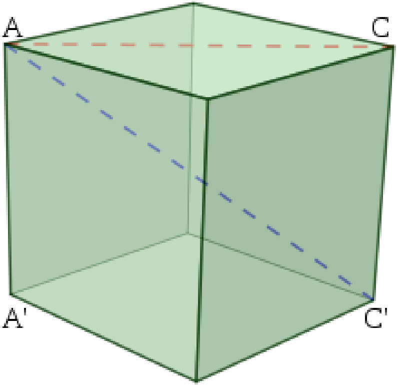
$$\langle \ell \rangle = \frac{2}{N(N-1)} \sum_{i<j} \ell_{i,j}$$

In most cases of interest $\langle \ell \rangle$ scales very slowly with N (Small-World) and, furthermore the distribution of the $\ell_{i,j}$ is quite picked around $\langle \ell \rangle$, and $\langle \ell \rangle \sim Diameter(G) = \max_{i,j} \ell_{i,j}$.

Regular *d*-dimensional Lattice

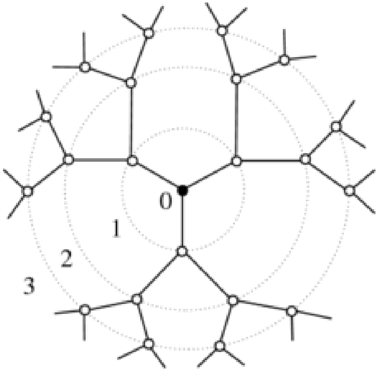


Regular d-dimensional Lattice



In a cube of side R, $N \propto R^d$ and $\langle \ell \rangle \sim D \sim R \Rightarrow \langle \ell \rangle \sim N^{1/d}$.

Bethe Lattice = Infinite Regular Tree



R=distance between the central node, chosen as reference and the nodes on the boundary of this sub-graph having N nodes. We have $N = 3 \times 2^{R-1} \Rightarrow R - 1 = \log(N/3)/\log(2) \Rightarrow$ the maximal distance between two randomly chosen nodes in the sub-graph will be $D = 2R = 2 + 2\log(N/3)/\log(2)$, from which we guess also $\langle \ell \rangle = \mathrm{O}(\log(N)/\log(2))$.

The Random Graph (A Finite “Random Bethe Lattice”)

Given N and a parameter $0 \leq p \leq 1$, for each pair of nodes put a link with probability p. In other words the $a_{i,j}$'s are i.i.d random variables with

$$Prob.\left(a_{i,j} = 1\right) = p\ , \quad Prob.\left(a_{i,j} = 0\right) = 1 - p.$$

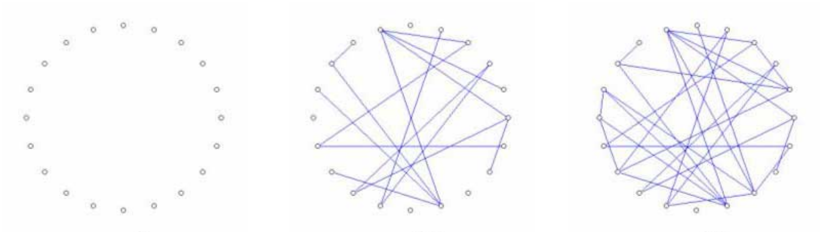
We have

$$\langle k \rangle = \frac{\langle 2L \rangle}{N},$$

and from

$$\langle L \rangle = \frac{N(N-1)}{2} p \quad \Rightarrow \quad \langle k \rangle = (N-1)p$$

A Random Graph with N=20



$p = 0$ (a) $p = 0.1$ (b) $p = 0.2$ (c)

Sparse Random Graph

Since

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

if we choose

$$p = \frac{c}{N-1}, \quad c = O(1) \quad \Rightarrow \quad \langle k \rangle = c = O(1) \quad \text{(Sparse} - \text{Graph)}$$

It is immediate to see that

$$\langle C \rangle = O\left(c\, N^{-1}\right) \quad \text{“Locally Tree} - \text{like...''} \text{ to be discussed later}$$

$\Rightarrow$ we can use the analogy with the Bethe Lattice and find-out the Small-World property:

$$\langle \ell \rangle \sim \frac{\log(N)}{\log(\langle k \rangle)} \quad \text{“Six Degrees of Separation...''} \text{ to be discussed later}$$

We have

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

$\Rightarrow$ in the sparse case, $p = \frac{c}{N-1}$, we have

$$\lim_{N\to\infty} P(k) = \frac{c^k}{k!} e^{-c}$$

$\Rightarrow$ This Degree Distribution is not representative of real-world networks!

How Real Nets look like: Flickr-User

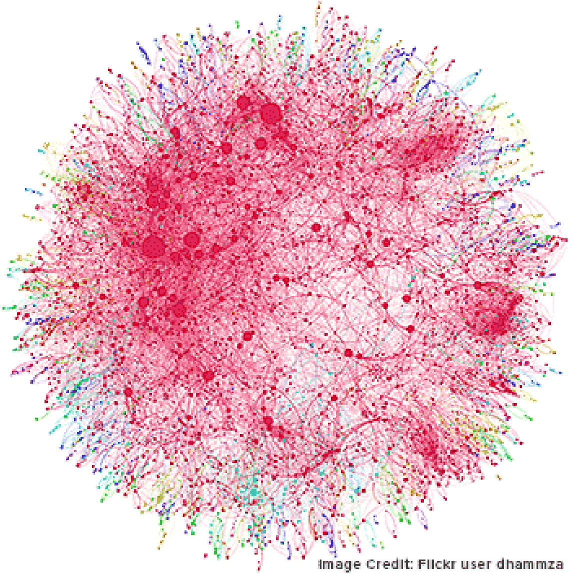
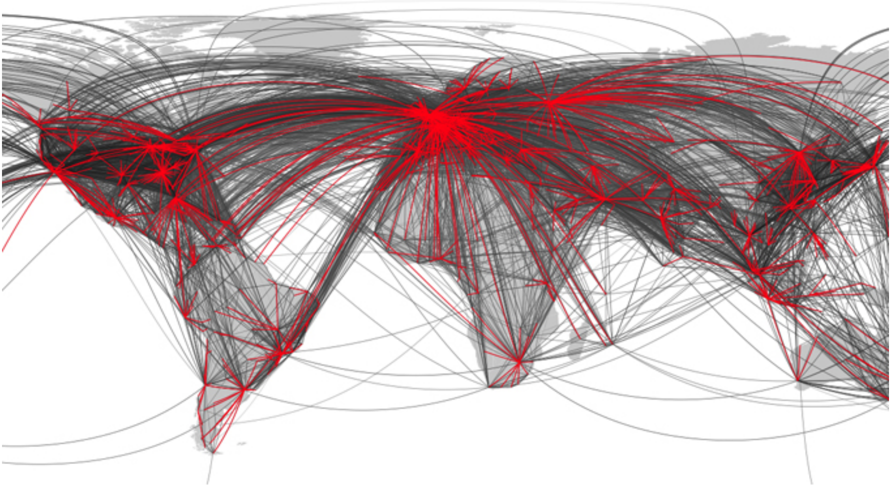
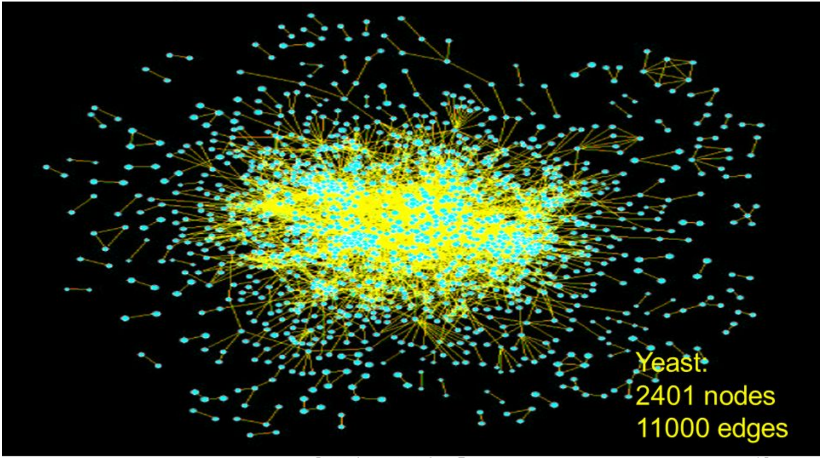


Image Credit: Flickr user dhammza

How Real Nets look like: Air-Traffic

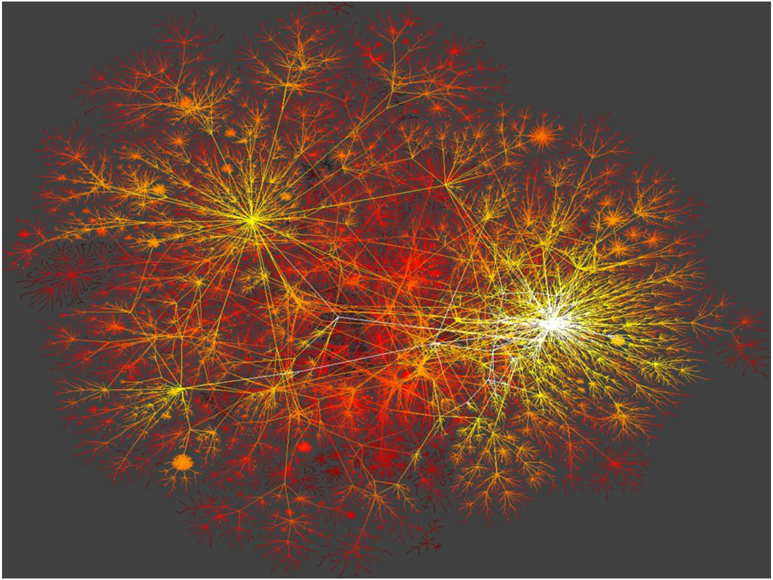


Biological networks: proteomics

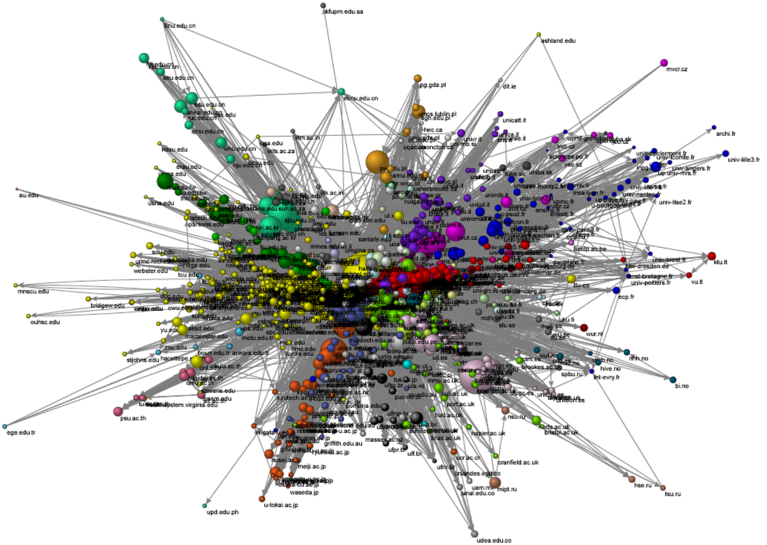


Complex networks - Bonato 12

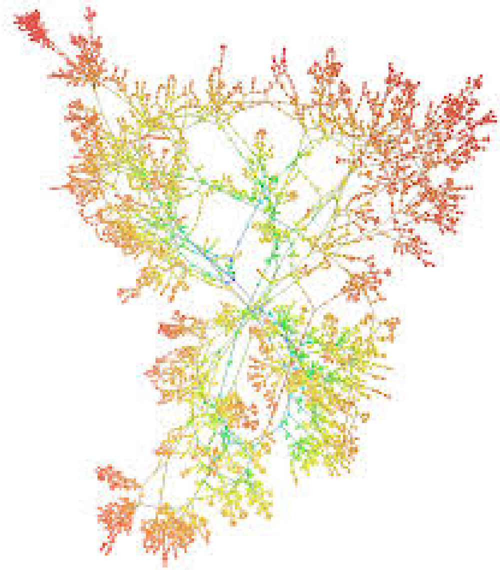
How Real Nets look like: Internet



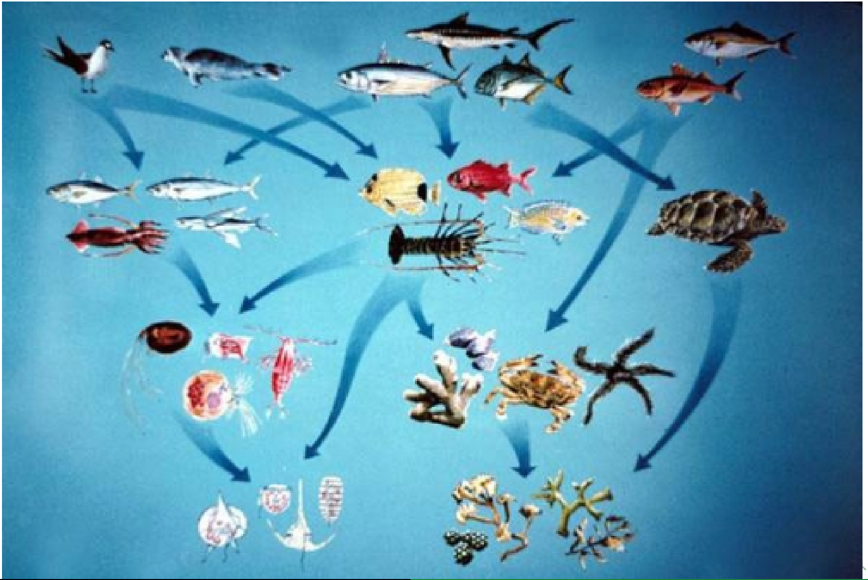
How Real Nets look like: WWW of Universities



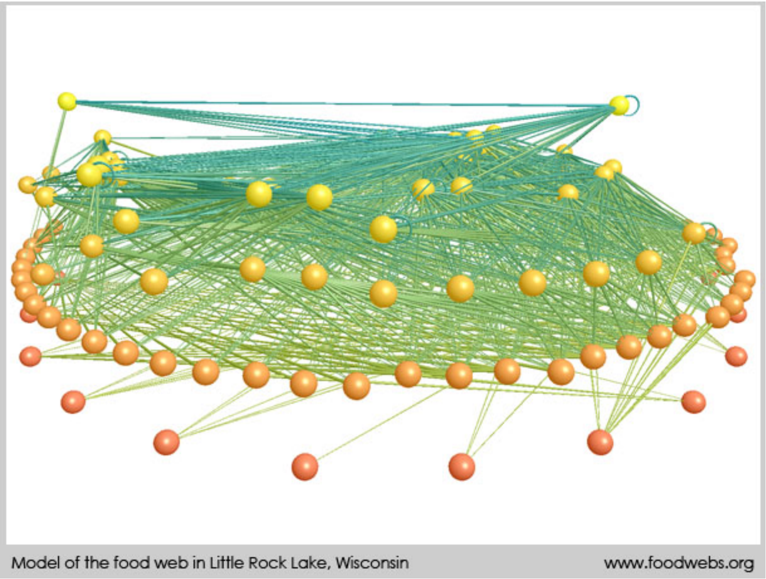
How Real Nets look like: Neural Network



How Real Nets look like: Food Web

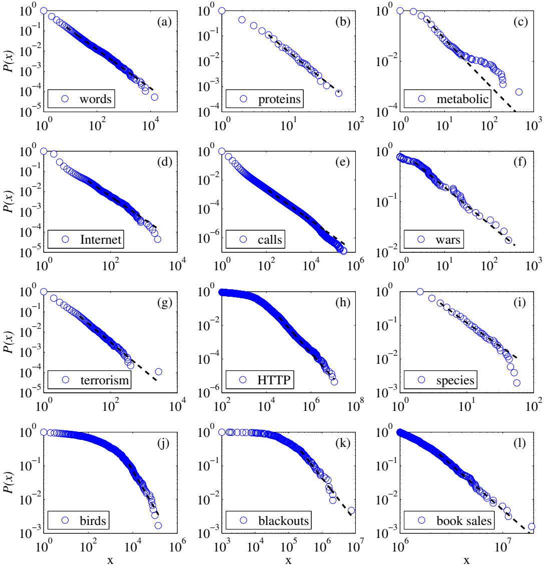


How Real Nets look like: Food Web



Power Law of Real Nets (A. Clauset *et.al.* SIAM 2009)

24 A. Clauset, C. R. Shalizi and M. E. J. Newman



Power Law of Real Nets (A. Clauset *et.al.* SIAM 2009)

quantity	n	$\langle x \rangle$	σ	$x_{\max}$	$\hat{x}_{\min}$	$\hat{\alpha}$
count of word use	18 855	11.14	148.33	14 086	7 ± 2	1.95(2)
protein interaction degree	1846	2.34	3.05	56	5 ± 2	3.1(3)
metabolic degree	1641	5.68	17.81	468	4 ± 1	2.8(1)
Internet degree	22 688	5.63	37.83	2583	21 ± 9	2.12(9)
telephone calls received	51 360 423	3.88	179.09	375 746	120 ± 49	2.09(1)
intensity of wars	115	15.70	49.97	382	2.1 ± 3.5	1.7(2)
terrorist attack severity	9101	4.35	31.58	2749	12 ± 4	2.4(2)
HTTP size (kilobytes)	226 386	7.36	57.94	10 971	36.25 ± 22.74	2.48(5)
species per genus	509	5.59	6.94	56	4 ± 2	2.4(2)
bird species sightings	591	3384.36	10 952.34	138 705	6679 ± 2463	2.1(2)
blackouts ($\times 10^3$)	211	253.87	610.31	7500	230 ± 90	2.3(3)
sales of books ($\times 10^3$)	633	1986.67	1396.60	19 077	2400 ± 430	3.7(3)
population of cities ($\times 10^3$)	19 447	9.00	77.83	8 009	52.46 ± 11.88	2.37(8)
email address books size	4581	12.45	21.49	333	57 ± 21	3.5(6)
forest fire size (acres)	203 785	0.90	20.99	4121	6324 ± 3487	2.2(3)
solar flare intensity	12 773	689.41	6520.59	231 300	323 ± 89	1.79(2)
quake intensity ($\times 10^3$)	19 302	24.54	563.83	63 096	0.794 ± 80.198	1.64(4)
religious followers ($\times 10^6$)	103	27.36	136.64	1050	3.85 ± 1.60	1.8(1)
freq. of surnames ($\times 10^3$)	2753	50.59	113.99	2502	111.92 ± 40.67	2.5(2)
net worth (mil. USD)	400	2388.69	4 167.35	46 000	900 ± 364	2.3(1)
citations to papers	415 229	16.17	44.02	8904	160 ± 35	3.16(6)
papers authored	401 445	7.21	16.52	1416	133 ± 13	4.3(1)
hits to web sites	119 724	9.83	392.52	129 641	2 ± 13	1.81(8)
links to web sites	241 428 853	9.15	106 871.65	1 199 466	3684 ± 151	2.336(9)

TABLE 6.1

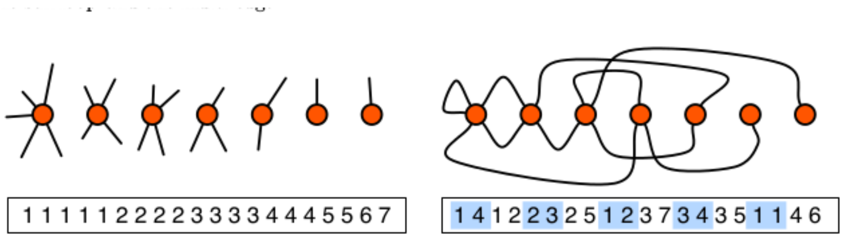
The Configuration Model

Consider a sequence of non negative integers $k_1, \ldots, k_N$ such that

$$\sum_{i=1}^{N} k_i = 2L$$

Then connect in all the possible ways the $2L$ stubs

The Configuration Model



The Configuration Model

- In principle we can build graphs with "any" given $P(k) = N(k)/N$
- Due to the random way by which we join the stubs, these graphs are approximately uncorrelated: $\mathit{Prob.}\left(a_{i,j} = 1\right) \simeq k_i k_j / 2L$
- Self-links and multiple-links exist but are negligible
- We can evaluate $\langle C \rangle$
- We can evaluate $\langle \ell \rangle$
- We can understand when correlations are important

The Configuration Model: $\langle C \rangle$

We use

$\langle C \rangle = \mathit{Prob.}\,$(that between two neighbors of a given node there is a link)

$$\Rightarrow \quad \langle C \rangle = \frac{1}{N} \frac{\langle k(k-1) \rangle^2}{\langle k \rangle^3}$$

The Configuration Model: $\langle \ell \rangle$

We use $\mathcal{N}_\ell =$ Number of Paths of length ℓ

$$\langle \mathcal{N}_\ell \rangle = \langle k \rangle \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)^{\ell - 1}$$

$$\Rightarrow \quad \langle \ell \rangle = \frac{\ln \left(N / \langle k \rangle \right)}{\ln \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)}$$

This makes us to understand also that the shortest loops are of length $O(\ln N)$

The Configuration Model: About Correlations

$$?? \quad Prob.\,(a_{i,j=1}) = \frac{k_i\,k_j}{N\langle k\rangle} \quad ??$$

$$\max_{i,j} \frac{k_i\,k_j}{N\langle k\rangle} = \frac{k_{\max}^2}{N\langle k\rangle}$$

If $P(k) \sim k^{-\gamma}$ we use

$$\langle k_{\max}\rangle \sim N^{\frac{1}{\gamma-1}}$$

$$\Rightarrow \quad \frac{k_{\max}^2}{N\langle k\rangle} \sim N^{\frac{3-\gamma}{\gamma-1}}$$

To summarize

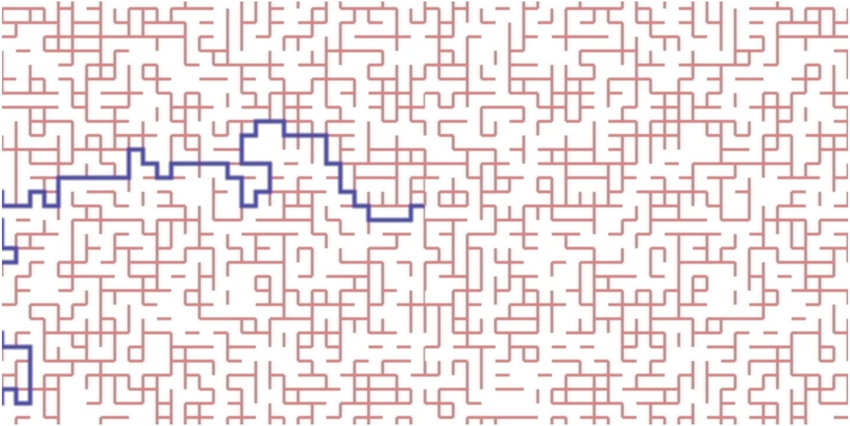
- We have static network models where $P(k)$ is the only "arbitrary" input (maximally random) and where, in particular, if $P(k) \sim k^{-\gamma}$ with $\gamma > 2$:

- $\langle C \rangle \to 0$ (Locally Tree-Like)
- $\langle \ell \rangle \sim \ln(N)$ for $\gamma > 3$ (Small-World), or
- $\langle \ell \rangle \sim \ln(\ln(N))$ for $\gamma < 3$ (Ultra-Small-World)
- Shortest Loops have lenght $\sim \ln(N)$ (Locally Tree-Like)
- Correlations do exist for $\gamma < 3$ (Degree-Degree Corr.)

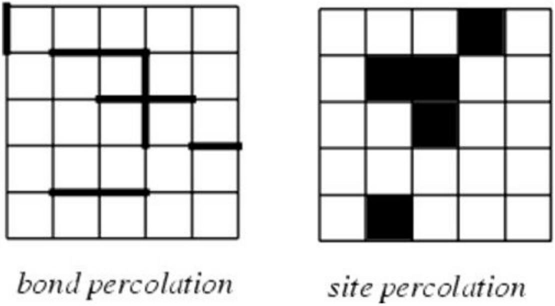
Day 2: Percolation and Magnetism

- Examples
- Node- and Link-Percolation
- Percolation in Uncorrelated Complex Networks
- Anomalous Mean-Field behavior
- Magnetism

Percolation



Link- and Node-Percolation



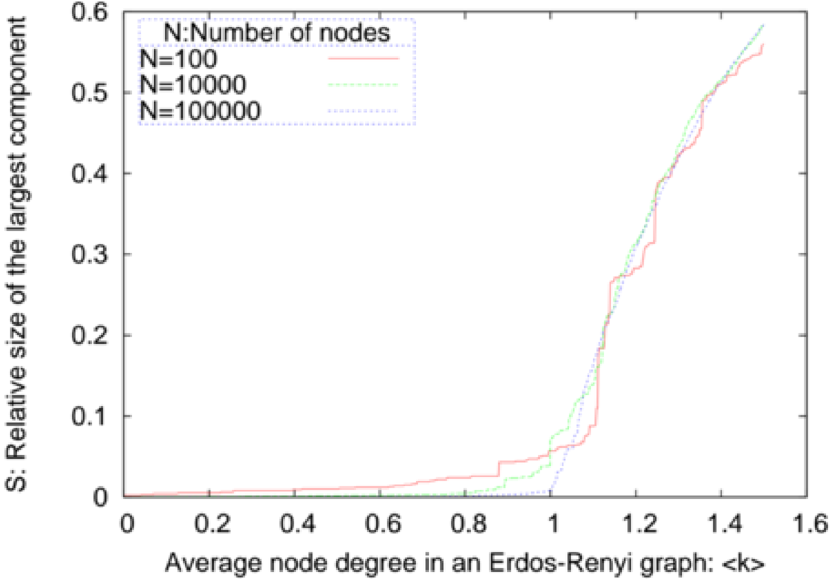
Source: http://mathworld.wolfram.com/BondPercolation.html

Bond- (or Link-) Percolation: Links are kept with probability p

Node- (or Site-) Percolation: Nodes are kept with probability p

For the moment being we can avoid making use of p.

Percolation Without Removal in the Random Graph



Percolation Without Removal in Uncorrelated Nets

F.C.C. = Finite Connected Component
G.C.C. = Giant Connected Component

$x = \mathit{Prob.}$ (that the end of a randomly chosen link points to a F.C.C.)

$S =$(Number of Nodes belonging to the G.C.C.)$/N$

$$x = \sum_{k=1} P_L(k) x^{k-1}$$

$$1 - S = \sum_{k=0} P(k) x^k$$

$$\frac{\langle k(k-1)\rangle}{\langle k\rangle} < 1 \quad \text{No Percolation}$$

$$\frac{\langle k(k-1)\rangle}{\langle k\rangle} > 1 \quad \text{Percolation}$$

Anomalous Mean-Field Behavior (No Removal)

Within a certain limit, we can choose $P(k)$ such that

$$(*) \quad \frac{\langle k(k-1)\rangle}{\langle k\rangle} = 1 \quad \text{Percolation Threshold}$$

If in particular $P(k) \sim k^{-\gamma}$, we find:

- if $\gamma > 4$, $\langle k^3\rangle < \infty$, and

  $S \sim (\langle k\rangle - \langle k\rangle_c)^\beta$, with $\beta = 1$ (classical limit)
- if $3 < \gamma < 4$, $\langle k^2\rangle < \infty$, $\langle k^3\rangle = \infty$, and

  $S \sim (\langle k\rangle - \langle k\rangle_c)^\beta$, with $\beta = \frac{1}{\gamma-3}$
- if $2 < \gamma < 3$, $\langle k^2\rangle = \infty$, to satisfy $(*)$ we need to introduce random node removal

Percolation via Node-Removal in Uncorrelated Nets

Now we remove randomly each node with probability $1 - p \Rightarrow$

$$x = 1 - p + p \sum_{k=1} P_L(k) x^{k-1}$$

$$1 - S = 1 - p + p \sum_{k=0} P(k) x^k$$

$$p \frac{\langle k(k-1) \rangle}{\langle k \rangle} < 1 \quad \text{No Percolation}$$

$$p \frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1 \quad \text{Percolation}$$

Anomalous Mean-Field Behavior (Node-Removal)

$$p_c = \frac{\langle k \rangle}{\langle k(k-1) \rangle} \quad \text{Percolation Threshold}$$

and if $P(k) \sim k^{-\gamma}$ we find

- if $\gamma > 4$, $\langle k^3 \rangle < \infty$, and

  $S \sim (p - p_c)^\beta$, with $\beta = 1$ (classical limit)
- if $3 < \gamma < 4$, $\langle k^2 \rangle < \infty$, $\langle k^3 \rangle = \infty$, and

  $S \sim (p - p_c)^\beta$, with $\beta = \frac{1}{\gamma - 3}$
- if $2 < \gamma < 3$, $\langle k^2 \rangle = \infty$, $p_c \to 0$ and $S \sim p^\beta$ with $\beta = \frac{1}{3-\gamma}$

Anomalous Ferromagnetic Mean-Field Behavior

If $P(k) \sim k^{-\gamma}$ we have:

- if $\gamma > 5$, $\langle k^4 \rangle < \infty$, and
  $m \sim (T - T_c)^{\beta}$, with $\beta = \frac{1}{2}$ (classical limit)
- if $3 < \gamma < 5$, $\langle k^2 \rangle < \infty$, $\langle k^4 \rangle = \infty$, and
  $m \sim (T - T_c)^{\beta}$, with $\beta = \frac{1}{\gamma - 3}$
- if $2 < \gamma < 3$, $\langle k^2 \rangle = \infty$, $T_c = \infty$, and
  $m \sim T^{-\beta}$ with $\beta = \frac{1}{3 - \gamma}$

In some networks $T_c = \infty$: how is it possible?

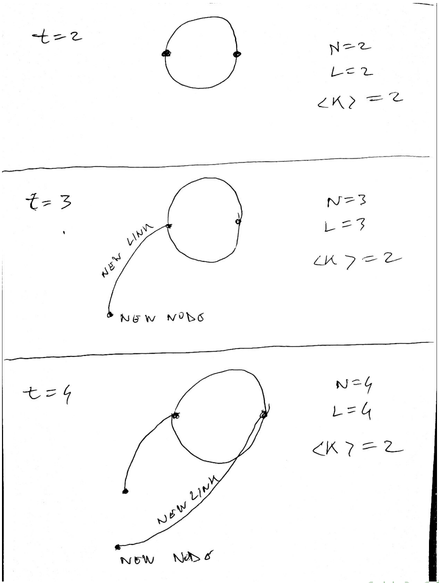
- If for k large $P(k) \sim e^{-k/\langle k \rangle}$ or $P(k) \sim k^{-\gamma}$ with $\gamma \geq 5$, we have in both cases almost homogeneous networks, $k \sim \langle k \rangle$ and $m \sim (T - T_c)^{\beta}$, with $\beta = \frac{1}{2}$ (cl. mean-field)
- If for k large $P(k) \sim k^{-\gamma}$ with $\gamma \sim 2$, we have an extremely heterogeneous network, k fluctuates a lot and $T_c = \infty$ with

  $m \sim T^{-\beta}$
- An extreme example: The Star Graph

Day 3: Growing Networks

- Examples
- Random Growing Model
- Barabasi-Albert Model
- Linear Model
- Continuum Approximation

Growing Networks



Random Growing Model (Dorogovtsev)

Prob. (that the new link goes to a node of degree k) $= \frac{1}{t}$
$\Rightarrow$ Master Equation for $P(k;t)$:

$$P(k;t+1)(t+1) - P(k;t)t = P(k-1;t) - P(k;t) + \delta_{k,1}$$

$$\Rightarrow \lim_{t\to\infty} P(k;t) = P(k) = \frac{1}{2^k}$$

Barabasi-Albert Model

$$\textit{Prob.}\,(\text{that the new link goes to a node of degree } k) = \frac{k}{\sum_{i=1}^{t} k_i}$$

$$\Rightarrow \lim_{t\to\infty} P(k;t) = P(k) \simeq \frac{1}{(k+2)(k+1)k} \simeq k^{-3}$$

$$\langle C \rangle \sim \frac{1}{N^{3/4}}$$

$$\langle \ell \rangle \sim \frac{\ln(N)}{\ln(\ln(N))}$$

Barabasi-Albert Model Generalized

Prob. (that one of the new m links goes to a node of degree k) $= \frac{k}{\sum_{i=1}^{t} k_i}$

$$\Rightarrow \lim_{t\to\infty} P(k;t) = P(k) \simeq \frac{1}{(k+2)(k+1)k} \simeq k^{-3}$$

$$\langle C \rangle \sim \frac{1}{N^{3/4}}$$

$$\langle \ell \rangle \sim \frac{\ln(N)}{\ln\left(\ln(N)\right)}$$

Linear Model

Here $k_0 \geq 0$ and k refers to IN-degree only

$$Pr.\,(\text{that one of the new } m \text{ links goes to a node of degree } k) = \frac{k+k_0}{\sum_{i=1}^{t}(k_i+k_0)}$$

$$\Rightarrow \lim_{t\to\infty} P(k;t) = P(k) \sim k^{-\gamma}, \quad \gamma = 2 + \frac{k_0}{m}$$

References:

1. R. Albert, A.L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74**, 47 (2002).
2. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of Networks* (University Press: Oxford, 2003).
3. M.E.J. Newman, *The structure and function of complex networks*, SIAM Review **45**, 167 (2003).
4. S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, *Critical phenomena in complex networks*, Rev. Mod. Phys. 80, 1275 (2008).
5. S.N. Dorogovtsev, *Lectures on complex networks* (Oxford Master Series in Statistical, Computational, and Theoretical Physics, 2010).