Fabrizio Luccio

Mathematical issues in network construction and security

Dottorato 08

1. The growth of networks

random graphs, power laws, and small worlds

Basic notions on undirected graphs

$G = (V,E)$

$N = |V|, \quad M = |E|$

C: number of connected components

L: number of independent loops

k: vertex degree

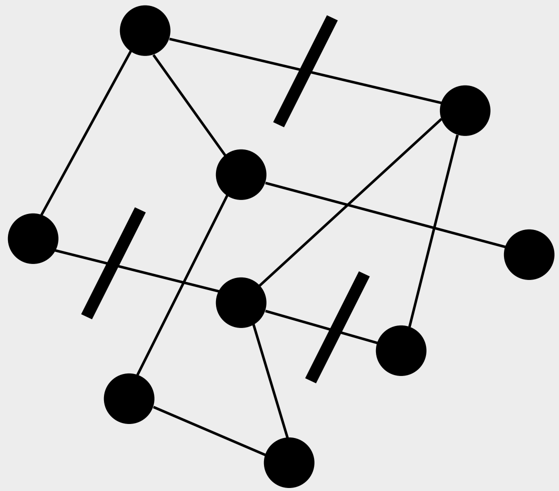
if $C = 1$ then $M \geq N - 1$

if $C = 1$ and $M = N - 1$, G is a tree ($L = 0$)

A basic formula on undirected graphs

$$N + L = M + C$$

$N = 9$
$M = 11$
$C = 1$
$L = 3$



Random networks have a disordered arrangement of edges.

A particular random network under study is only one member of a statistical ensemble of all possible realizations.

Therefore the statistical description of a random network is in fact the description of the corresponding ensemble.

We shall study networks in the form of graphs (possibly, random graphs).

Degree distribution

p(k,s) is the probability that vertex s has degree k

Total degree distribution $$P(k) = \frac{1}{N}\sum_{s=1}^{N} p(k,s)$$

Average degree (first moment) $$\bar{k} = \sum_k kP(k)$$

The number of edges is $$M = \bar{k}N/2$$

Networks with directed edges (directed graphs)

$p(k_i,s)$ and $p(k_o,s)$ are the probabilities that vertex s has in-degree k_i and out-degree k_o

The total degree distributions $P(k_i)$ and $P(k_o)$ are defined as before

The average in and out-degrees are equal:

$$\bar{k_i} = \bar{k_o} = \bar{k}/2$$

Typical degree distributions for networks, for $N \to \infty$ and fixed value of $\bar{k}$

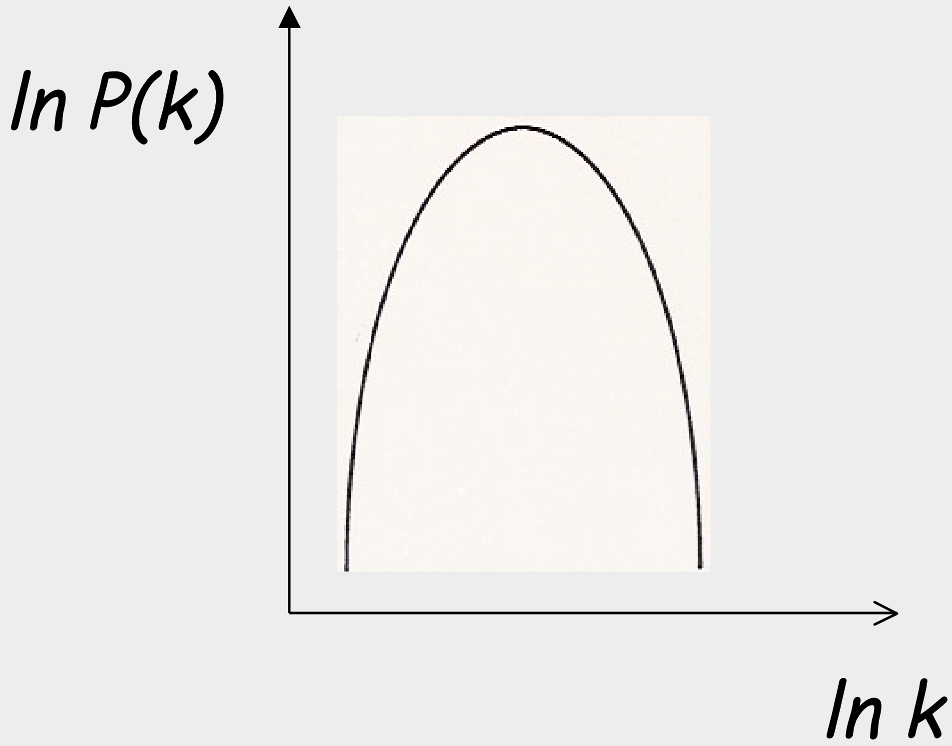
The Poisson distribution

$$P(k) = \frac{e^{-\bar{k}}\bar{k}^k}{k!}$$

where the average is computed from 0 to ∞

$$\bar{k} = \sum_{k=0}^{\infty} k P(k)$$

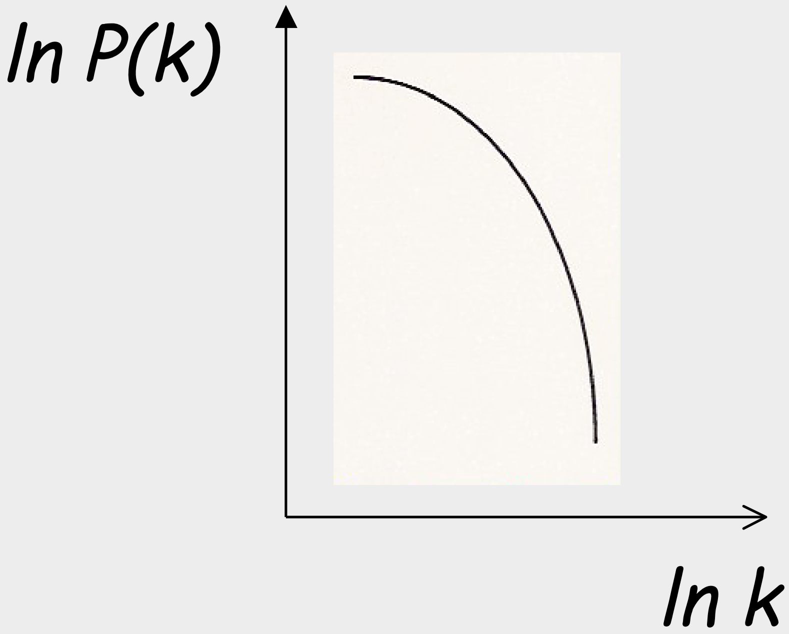
The Poisson distribution



Natural scale of the order average degree

The Exponential distribution

$$P(k) \propto e^{-k/\bar{k}}, \quad \bar{k} = \sum_{k=0}^{\infty} kP(k)$$



Natural scale of the order average degree

In the Poisson and Exponential distributions:

all the moments

$$\mathcal{M}_m = \sum_{k=0}^{\infty} k^m P(k)$$ are finite

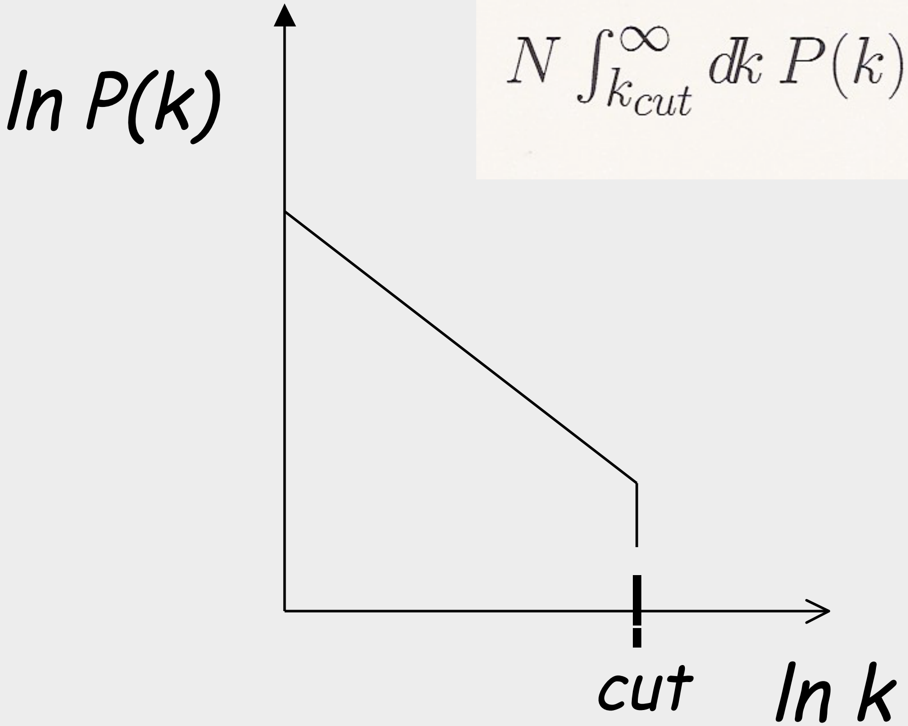
The Power-law distribution

$$P(k) \propto k^{-\gamma}, \quad k \geq k_0 > 0$$

The Power-law distribution

Real networks have a cut point: the number of vertices of degree $> k_{cut}$ is of order 1

$$N \int_{k_{cut}}^{\infty} dk\, P(k) \sim 1 \;\Rightarrow\; k_{cut} \sim k_0 N^{1/(\gamma-1)}$$



In an infinite Power-law distribution

$$P(k) \propto k^{-\gamma}, \quad k \geq k_0 > 0$$

all higher moments of order $m > \gamma - 1$ diverge.

If the first-order moment (average degree) is finite, we have $\gamma > 2$.

In a growing network, M may grow faster than a linear function of N. In this case the average degree diverges and we have $1 < \gamma \leq 2$.

Infinite power-laws are self-similar

Self-similarity means that an infinite structure S and a part of it appear to be the same. This entails the possibility of scaling, i.e., for $S = S(x)$ we have $S(cx) = c^{\gamma} S(x)$ where c is a constant and γ is the scaling exponent.

The only functions obeying this relationship are the power-laws.

In the Euclidean space a volume V scales with exponent +3 in the linear length L: a cube with $V = L^3$ is still a cube if the edge is doubled, $L = 2L$ and $V = 2^3 L^3$.

Fractals scale according to their non integer dimensions.

The Erdös-Rényi *graph process*

The network has N fixed vertices.

- $M \leq N(N-1)/2$ edges are added one by one. After all insertions, each two vertices are connected by an edge with probability p
- One of the $2^{N(N-1)/2}$ possible graphs is generated

Degree analysis

In a vertex of degree k, the edges can reach any k of the other N-1 vertices. Standard combinatorics yields the binomial distribution:

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

Together with the binomial distribution we have:

the average degree is $\bar{k} = p(N-1)$

and the network has $M = pN(N-1)/2$ edges on average

For $N \to \infty$ and $\bar{k}$ fixed, the degree distribution takes the Poisson form:

$$P(k) = \frac{e^{-\bar{k}}\bar{k}^k}{k!}$$

The *one vertex at a time* graph growing

The network has N vertices, inserted one at a time. At step s, vertex s is added.

- $M = N-1$ edges are added randomly between existing vertices, one at each step. $k(s,t)$ is the degree of vertex s at time $t \geq s$.
- Upon birth, each vertex is not connected: $k(s=t,t) = 0$.

Degree analysis

At step t, each vertex may increase its degree by $2/t$ (two vertices are connected by an edge) with same probability. On average:

$$\frac{\partial \bar{k}(s,t)}{\partial t} = \frac{2}{t} \quad \Rightarrow \quad \bar{k}(s,t) = 2\ln t + C(s)$$

One vertex at a time analysis (continued)

For determining C(s) in $\bar{k}(s,t) = 2\ln t + C(s)$

use the boundary condition $\bar{k}(s,s) = 2\ln s + C(s) = 0$

to finally have: $\bar{k}(s,t) = 2\ln(t/s)$

Then, for any fixed vertex s, the degree grows moderately with t.

The degree distribution P(k,t) is proportional to the number of vertices with degree k in a small interval around k. Then:

$$P(k,t) = -\frac{1}{t}\frac{\partial s(k,t)}{\partial k} = \frac{1}{2}e^{-k/2}$$

Exponential distribution

The *preferential linking* growing

- The network has N vertices, inserted one at a time. At step s, vertex s is added.

- $M = N-1$ edges are added, one at each step. $k(s,t)$ is the degree of vertex s at time $t \geq s$. The new edge is attached at vertices x, y with probabilities proportional to $k(x,t)+A$, $k(y,t)+A$.

- Note that $A>0$ otherwise a new vertex never gets an attached edge.

- A greater value of A (from 0 to ∞) indicates a smaller "preference".

Preferential linking analysis

Degree analysis

On average, at step t a vertex s increases its degree by:

$$2[k(s,t)+A]/\sum_{u=0}^{t}[k(u,t)+A] = 2[k(s,t)+A]/[(2+A)t]$$

then the average degree of s is determined by:

$$\frac{\partial \bar{k}(s,t)}{\partial t} = \frac{2(\bar{k}(s,t)+A)}{(2+A)t}$$

with boundary condition $\bar{k}(s=t,t) = 0$ (upon birth each vertex is not connected).

This yields:

$$\bar{k}(s,t)+A = (t/s)^{\frac{1}{1+A/2}}$$

The degree of s grows sharply with t

Preferential linking analysis (continued)

As for the exponential case, the degree distribution P(k,t) is proportional to the number of vertices with degree k in a small interval around k. Then:

$$P(k,t) = -\frac{1}{t}\frac{\partial s(k,t)}{\partial k}$$

that yields:

$$P(k) \propto k^{-(2+A/2)} \equiv k^{-\gamma}$$

Power-law distribution

The simple Barabási-Albert model (SBAM) 1999

- The network has N vertices, inserted one at a time. At step s, vertex s is added.
- M = N-1 edges are added, one at each step. k(s,t) is the degree of vertex s at time t ≥ s. The new edge is attached to s, and to vertex x with probability proportional to k(x,t).

SBAM analysis

On average, at step t the degree of a vertex s is increased by:

$$k(s,t)/\sum_{u=0}^{t} k(u,t) = k(s,t)/2t$$

then the average degree of s is determined by:

$$\frac{\partial \bar{k}(s,t)}{\partial t} = \frac{\bar{k}(s,t)}{2t}$$

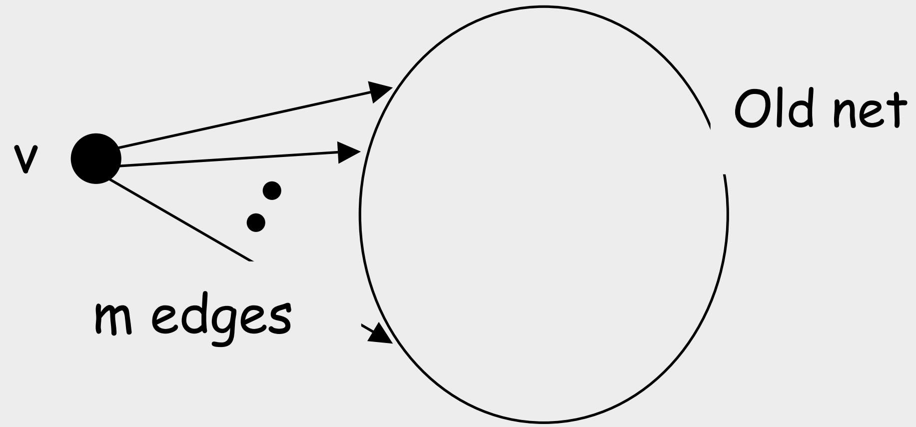
with boundary condition $\bar{k}(s=t,t) = 1$ (each new vertex is connected to the new edge).

This yields: $\bar{k}(s,t) = (t/s)^{\frac{1}{2}}$ and then $P(k) \propto k^{-3}$

Power-law

The Barabási-Albert model with directed edges (DBAM)

At each step a new vertex v and m edges are added, directed from v to existing vertices chosen with preferential linking on the in-degree.



DBAM growing

- The network has N vertices, inserted one at a time. At step s, vertex s is added. $k_i(s,t)$ is the in-degree of vertex s at time $t \geq s$
- $M = m(N-1)$ edges are added, m of them for each time step. The m edges are directed from the new vertex v to vertices x with probability proportional to $k_i(x,t) + \alpha m$ (the original DBAM had $\alpha = 1$). At each time step the total degree of s is $k_i(s,t) + m$, as the out-degree is always m.

DBAM analysis

On average, at step t the in-degree of a vertex s is increased by:

$$\frac{m(k_i(s,t)+\alpha m)}{\sum_{u=0}^{t}(k_i(u,t)+\alpha m)} = \frac{k_i(s,t)+\alpha m}{(1+\alpha)t}$$

then the average degree of s is determined by:

$$\frac{\partial \bar{k}_i(s,t)}{\partial t} = \frac{\bar{k}_i(s,t)+\alpha m}{(1+\alpha)t}$$

with boundary condition $\bar{k}_i(s=t,t) = 0$.

This yields

$$\bar{k}_i(s,t) = \alpha m[(t/s)^{\frac{1}{1+\alpha}} - 1]$$

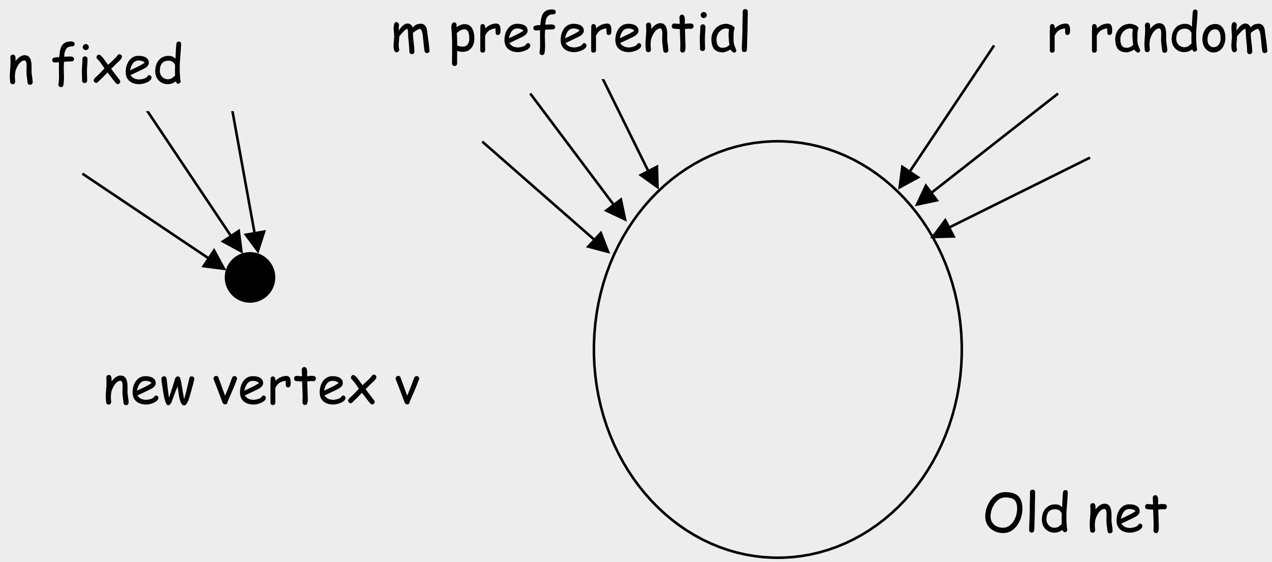
and

$$P(k_i) \propto k_i^{-\gamma},\ with\, \gamma = 2+\alpha$$

Power-law

The *preferential-and-random* model of Dorogovtsev-Mendes (PRDMM) 2003

This is a more realistic model of growth with directed edges. In fact, is a minimal model that captures the effect of both preferential and random linking.



PRDMM growing

The network has N vertices, inserted one at a time. At step s, vertex s is added. $k_i(s,t)$ is the in-degree of vertex s at time $t \geq s$

• $n + m + r$ edges are added at each time step. n of these edges are directed to the new vertex v. m are directed to vertices x with probability proportional to $k_i(x,t) + A$. r are directed to randomly chosen vertices.

• The source vertices of the new edges are immaterial for the analysis.

PRDMM analysis

As in the analyses done for the previous models, we have:

$$\frac{\partial \bar{k}_i(s,t)}{\partial t} = r\frac{1}{t} + m\frac{\bar{k}_i(s,t) + A}{\sum_{u=0}^{t}(\bar{k}_i(u,t) + A)}$$

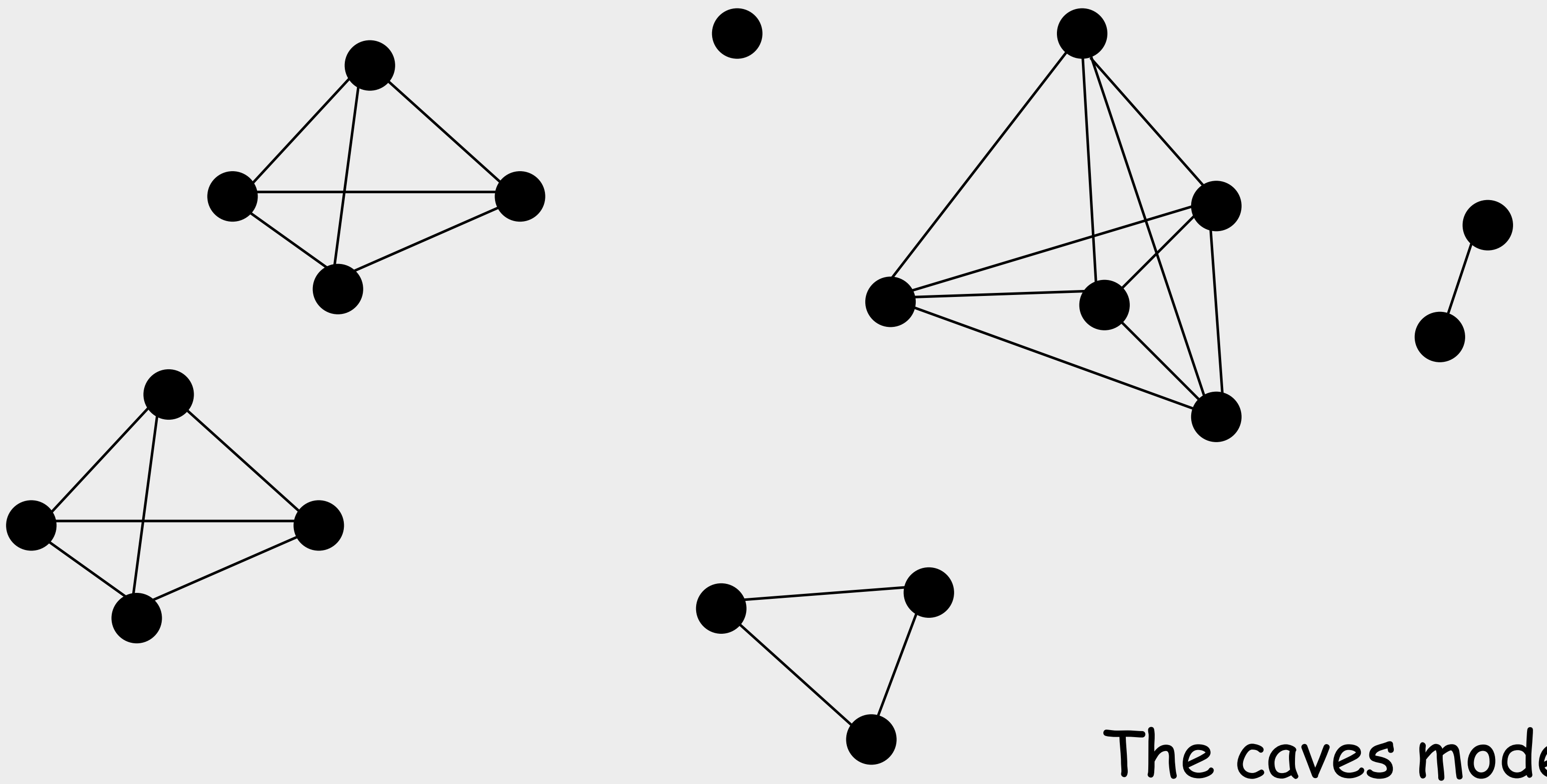
with boundary condition $\bar{k}_i(s=t,t) = n$.

This is the superimposition of two effects, ending in the power-law distribution:

$$P(k_i) \propto k_i^{-\gamma}, \;\; with\; \gamma = 2 + \frac{r+n+A}{m}$$

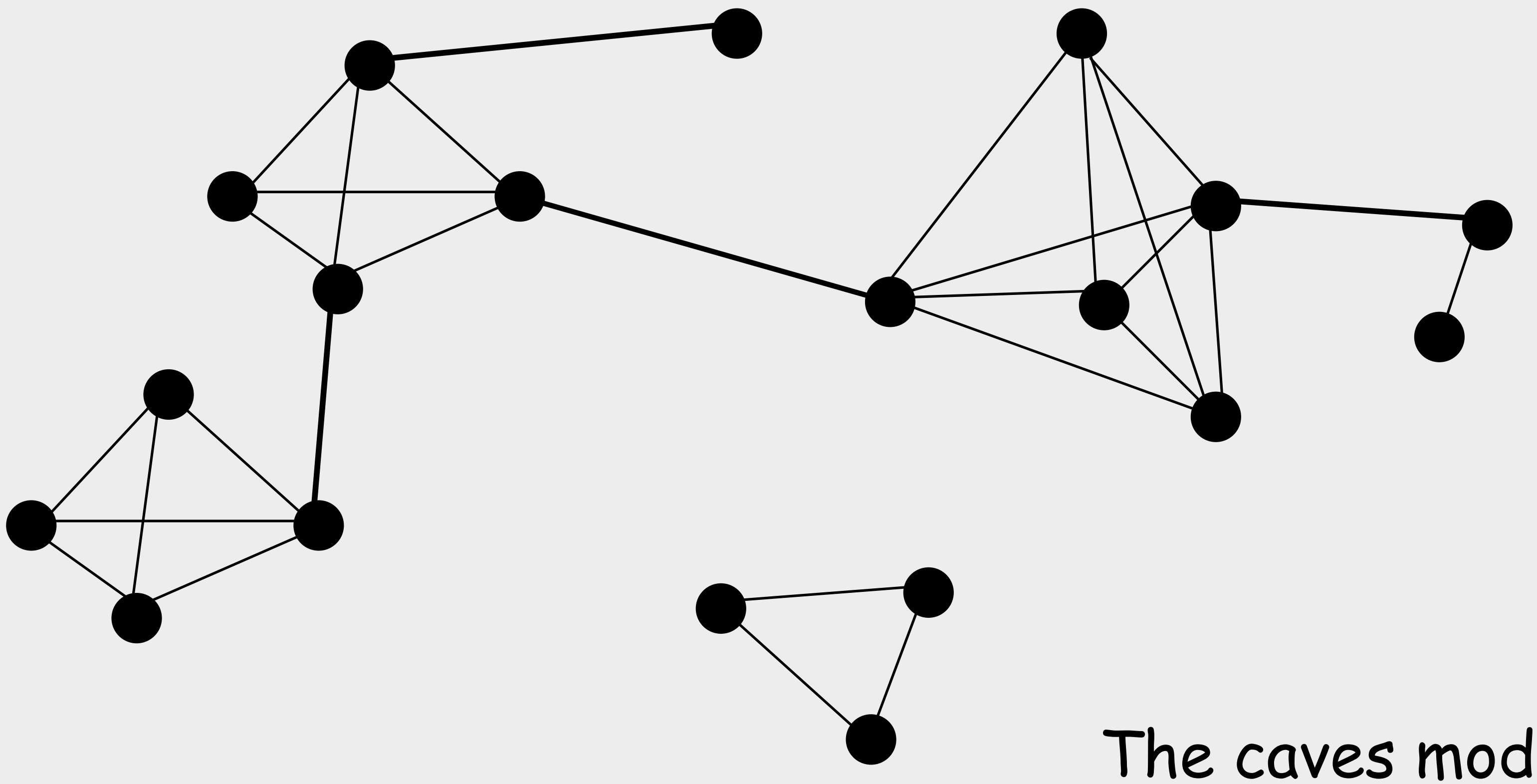
The value of γ shows that the factors of random attachment, incoming connections to v, and attractiveness A, have comparable effects. A may be negative, but its reasonable values are greater than $-(r+n)$.

The emergence of small worlds:
form the caves to the present society



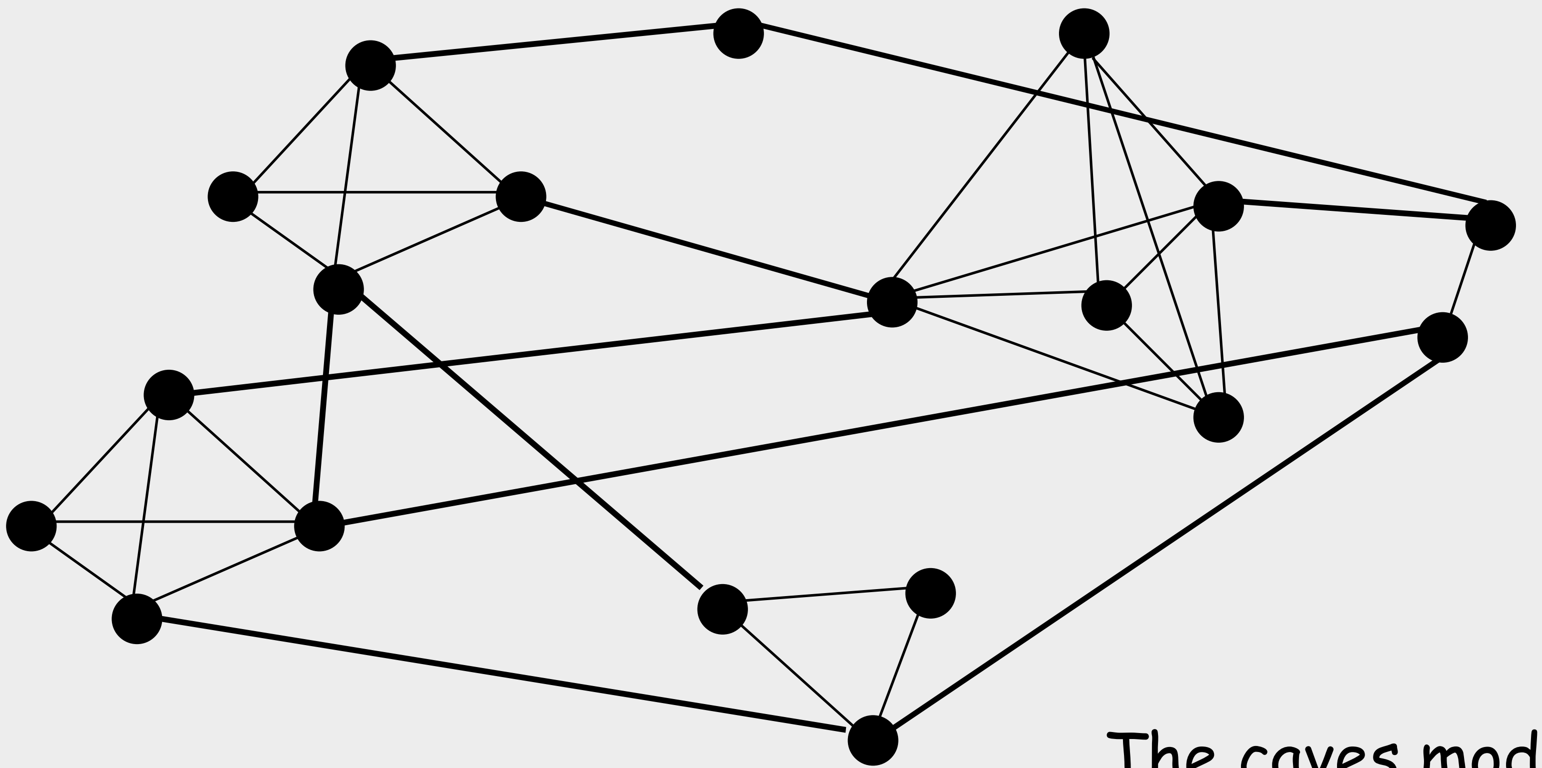
The caves model

Small worlds



The caves model
with messengers

Small worlds



The caves model
with telephones

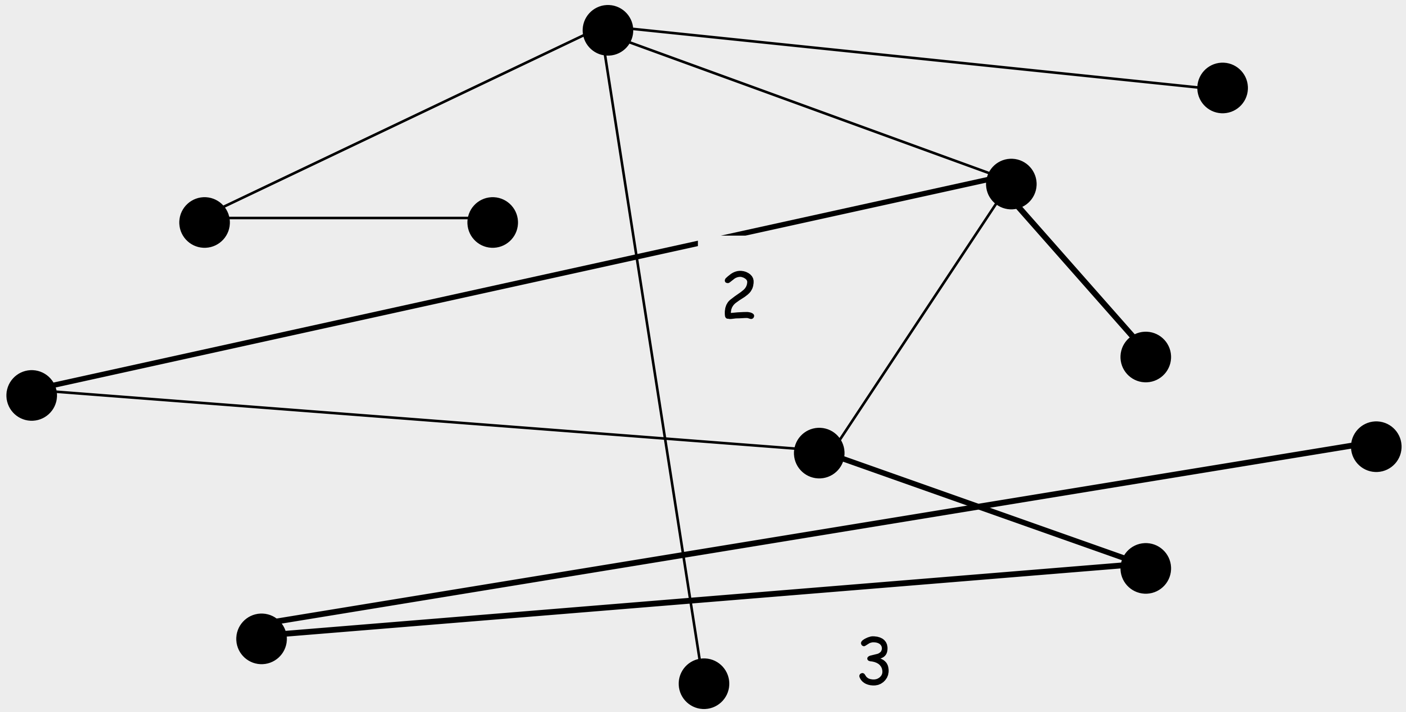
Small worlds: vertex distance

A key concept is the distance between any two vertices x, y, i.e. the number of edges in the shortest path between x and y

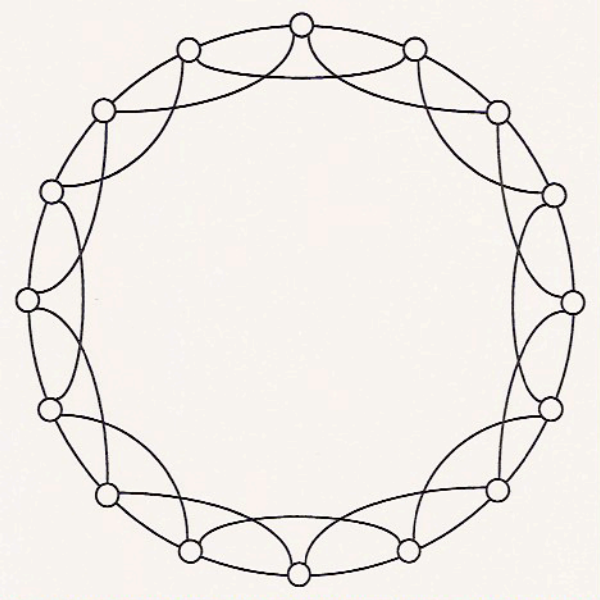
Small worlds (Milgram's experiment 1967)

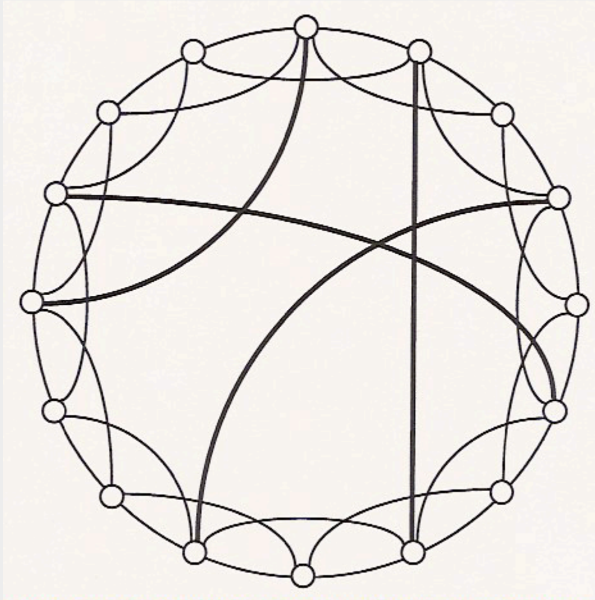
The distance between two randomly chosen vertices is probably small

TRUE in a random graph



Small worlds (Watts-Strogatz construction 1998)

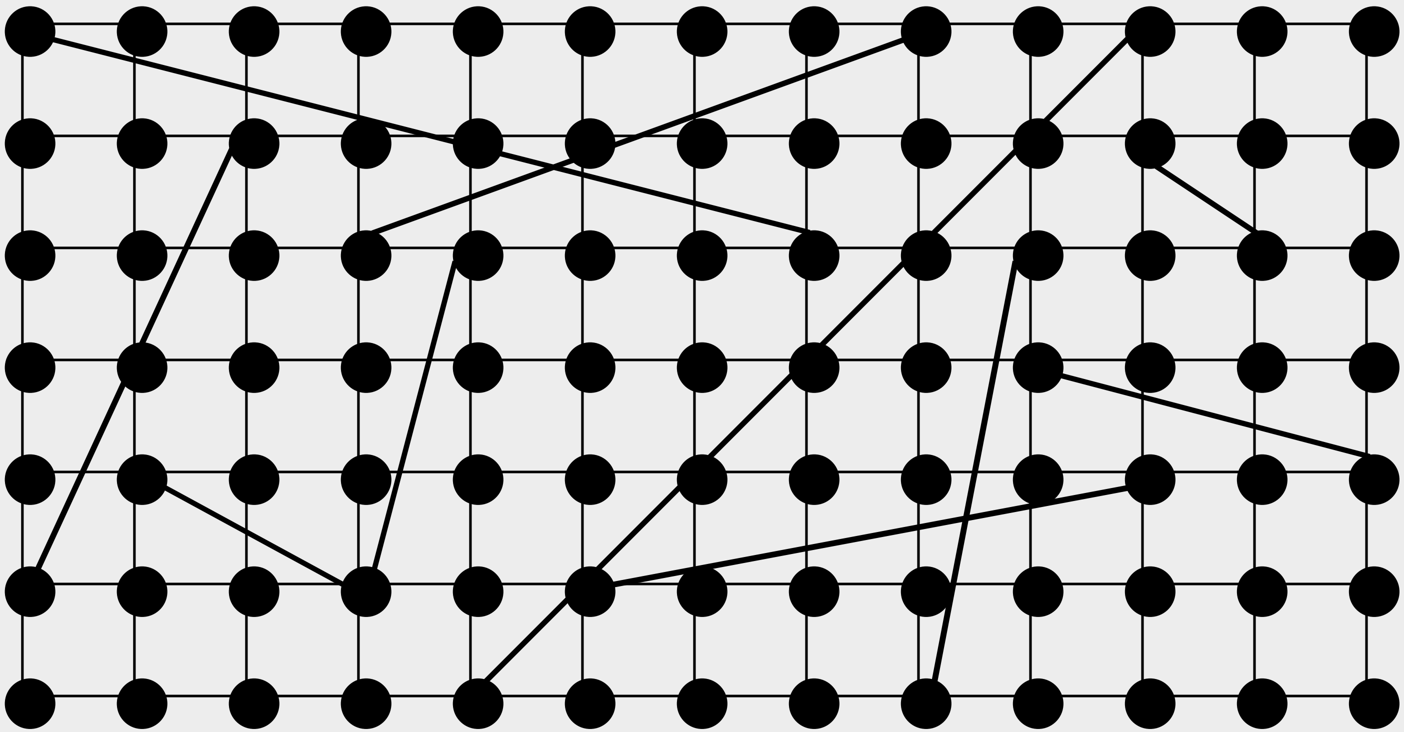




Adding random edges to a regular lattice

amounts to building a small world

Small worlds (Watts-Strogatz construction 1998)



Adding random edges to a regular lattice

amounts to building a small world

Clustering

For a vertex v, let Z be the set of vertices at distance one (1-neighbors), $z = |Z| = k(v)$

y is the number of edges connecting the vertices in Z

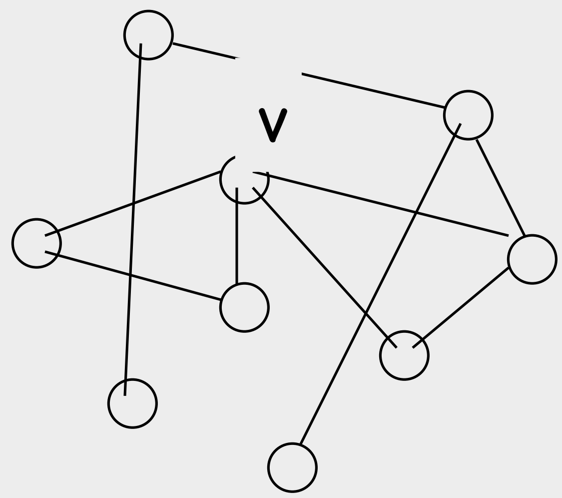
$w = z(z-1)/2$ is the maximum value of y

The clustering coefficient of v is: $C(v) = y/w$

Clustering in a random network

$N = 9,\ M = 10,\ \bar{k} = 2M/N = 2.22$

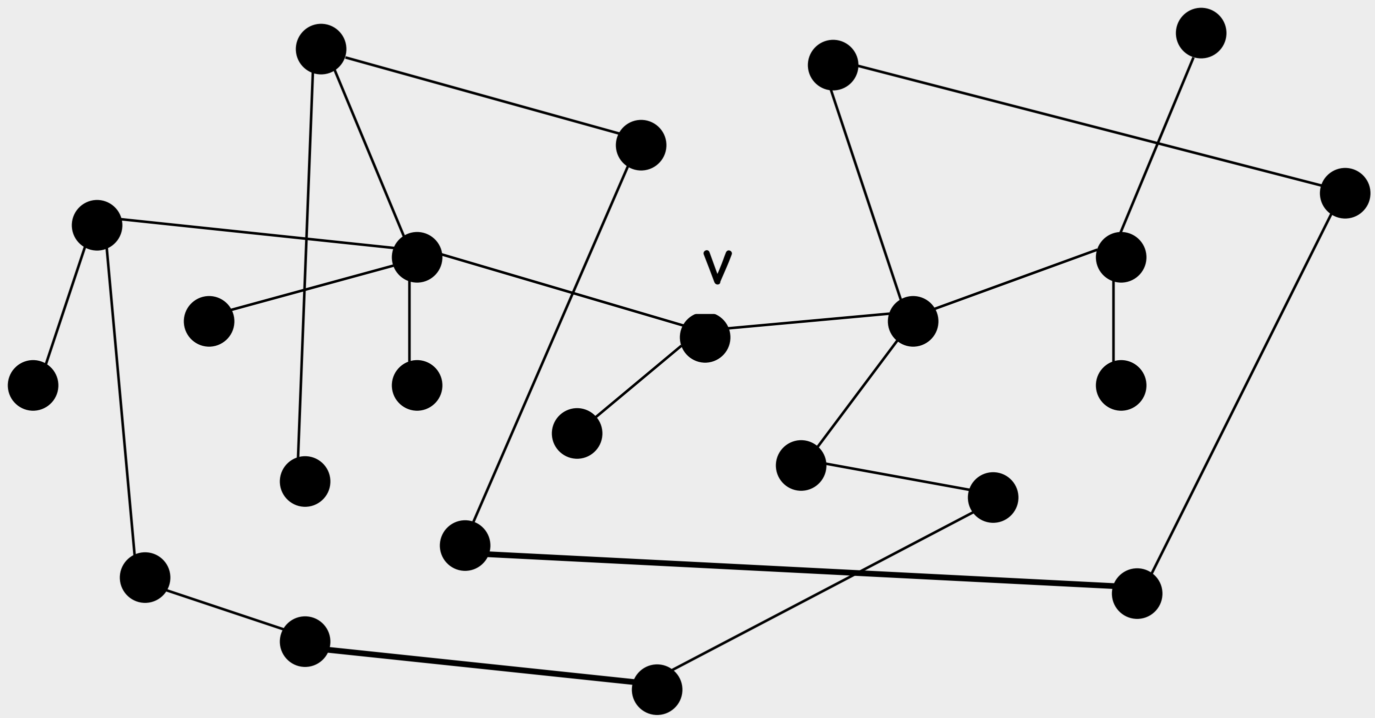
$z = 4,\ w = 6,\ y = 2,\ C(v) = 1/3$



In general $z = \bar{k}$, and $\bar{k}/N$ is the probability that two vertices are connected. We have $\bar{C} = \bar{k}/N = 2M/N^2 = 0.25$ (see later).

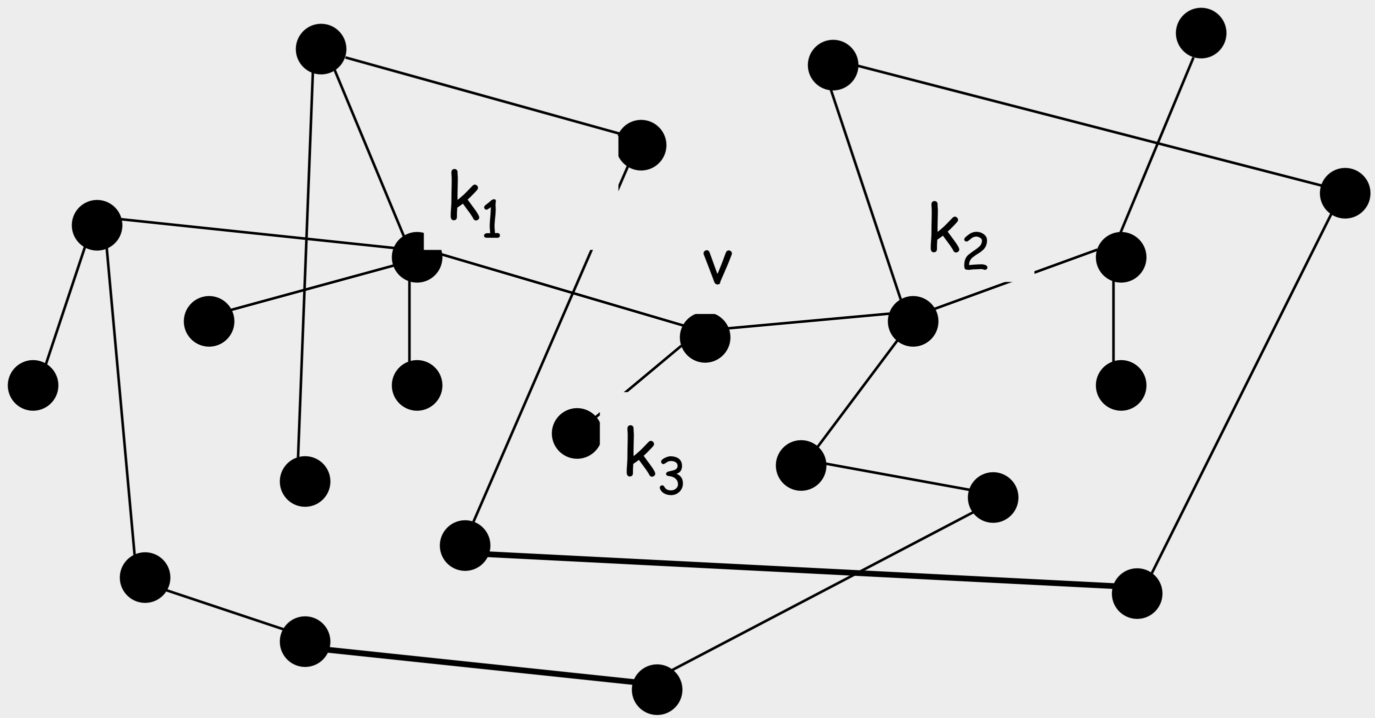
$\bar{C}$ indicates the probability that there is an edge between 1-neighbors, i.e. a loop of length 3. Random graphs with M linear in N have very few loops.

Large random networks have a tree-like local structure



Loops appear on the 4-th "shell" of v

Large random networks have a tree-like local structure



Probability that the 1-neighbours are directly connected:

$$(k_1 - 1)\,(k_2 - 1) \,/\, (N\,\bar{k})$$

Averaging this probability we compute the clustering coefficient:

$$C = \sum_{k_1,k_2} \frac{k_1 P(k_1)}{\bar{k}} \frac{k_2 P(k_2)}{\bar{k}} \frac{(k_1-1)(k_2-1)}{n\bar{k}} \sim \frac{\bar{k}}{N}$$

where the approximation derives from a property of Poisson distribution.

Since the clustering coefficient of random graphs is $\bar{k}$ / N, the edges between 1-neighbours are practically inexistent.

Generalizing the computation to d-neighbours, we conclude that the network has a tree-like structure around vertex v.

So, we derive a well-known relation for random graphs:

Recalling that z is the number of 1-neighbors, $\bar{z}^d$ is the approximate number of vertices at a distance d or less from any given vertex, for “large” $\bar{z}$.

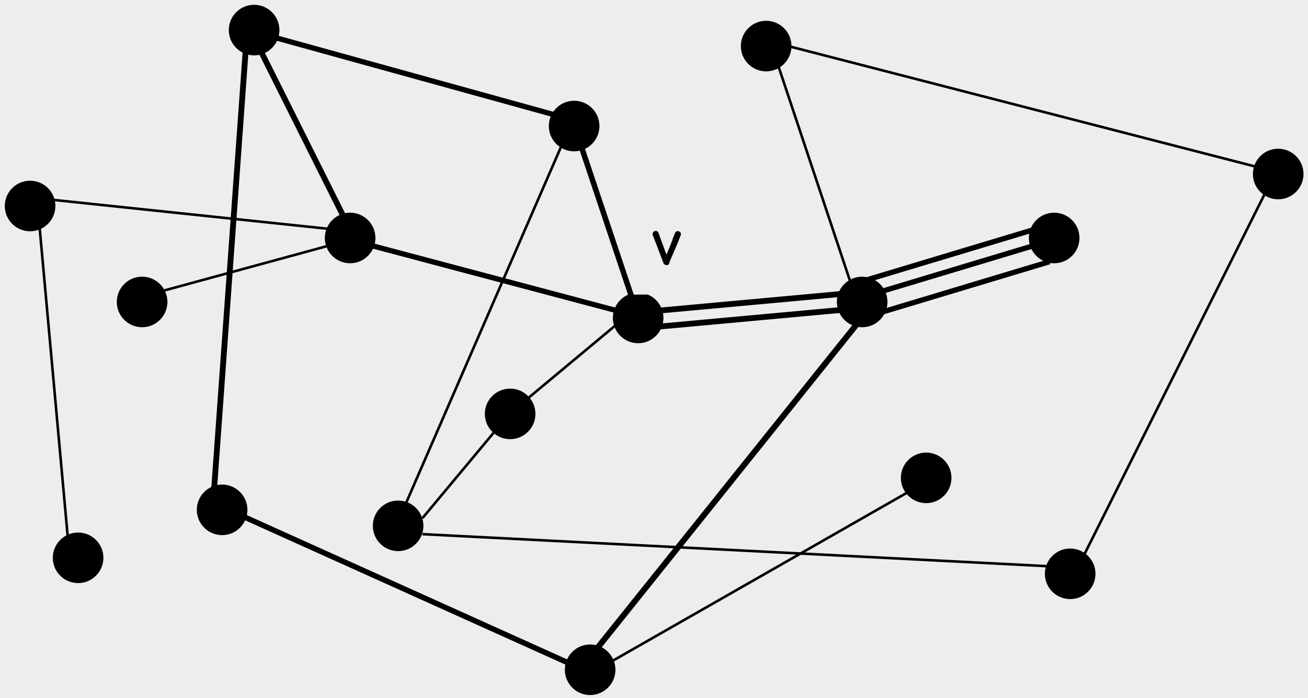
Then we can compute the average length $\bar{d}$ of the shortest path as:

$$N \sim \bar{z}^{\bar{d}} \quad \Rightarrow \quad \bar{d} \sim \frac{\ln N}{\ln \bar{z}}$$

This is the small world effect. Compare with the extreme values $N^{1/r}$ for an r-dimensional grid, or 1 for a complete graph.

Betweenness σ (also called "load")

For a vertex v, $\sigma(v)$ is a weighted measure of the number of shortest paths passing through v.



The concept of betweenness was introduced in sociology to indicate the "centrality" of a vertex

If the number of shortest paths between vertices i, j is B(i,j) > 0, and B(i,m,j) pass through v, we have:

$$\sigma(v) \equiv \sum_{i \neq j} \frac{B(i,v,j)}{B(i,j)}$$

The paths of the previous example give a contribution of 2/3 to σ(v)

Summarizing on network construction

Classical random graphs in the Erdös-Rényi model

• equilibrium graphs with Poisson degree distribution, with all finite moments

• average shortest path length of order *ln N*

• tree-like local structure with loops observable at a large scale

• clustering disappears with *N* going to infinity

Random graphs with consecutive addition of vertices

• non-equilibrium graphs with Exponential degree distribution, with all finite moments

• general properties as before, as *N* goes to infinity

Summarizing on network construction

Watts-Strogats small-world networks

•lattice local structure with superimposition of random edges, high clustering

•equilibrium graphs with Poisson-like degree distribution

• average shortest path length tends to a constant for increasing density of random edges

Barabási-Albert preferential linking

• non-equilibrium graphs with Power-law degree distribution

• finite first-order moment for the law exponent > 2; all other moments diverge

• average shortest path length of order *ln N*, low clustering

A fundamental book for starting

S.R. Dogorovtsev, J.F.F. Mendes. Evolution of Networks. Oxford University Press 2003.

many formulae in this section are taken from it