Ανάλυση Κοινωνικών Δικτύων
(Social Network Analysis)

Τοπολογίες Σύνθετων Δικτύων και Εφαρμογές

Συμεών Παπαβασιλείου (papavass@mail.ntua.gr)
Βασίλειος Καρυώτης (vassilis@netmode.ntua.gr)

08 Νοεμβρίου, 2019

Today's Menu

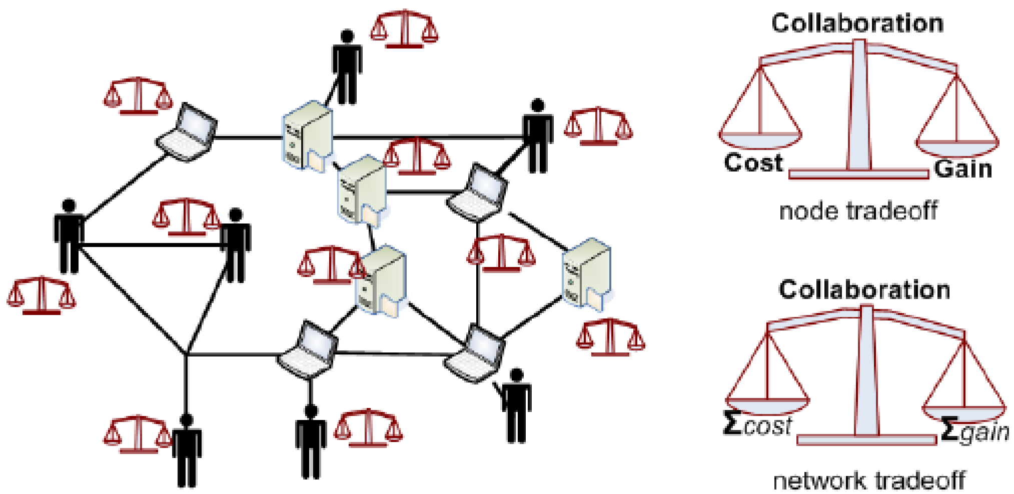
- Network manifest
  - Network formation
- Network classification
- Overview of emerging complex network topologies and their applications
  - Regular
  - Random (Gilbert, Erdos-Renyi models)
    - Random Regular
    - Generalized Random
  - Random Geometric
  - Scale-free (Barabasi-Albert model)
  - Small-world (Watts-Strogatz model)

Network Fundamentals

- Set of interacting entities
  - Collaborating actors → coalitions
  - Competing actors
- Emerging tradeoff:

gain vs. **cost** of collaboration

- Regards:
  - Gain obtained by collaboration/selfish behavior
  - Cost incurred by collaboration/selfishness
    - synchronization issues
    - message complexity



Complex Network Taxonomy

Communication, infrastructure, technological networks	Social and economical networks	Biological networks
Designed and/or engineered	Human initiated, Spontaneous growth	Spontaneous evolution

Static – Evolving Networks

Static: Topology does not change with time (closed)

- mesh
- sensor (rule of thumb)
- nodes remain fixed
- connections can vary
- growth maybe observed
- focus is on network optimization

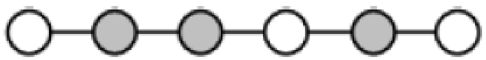
Evolving: Topology is time-dependent (open)

- scale-free (Internet)
- small-world
- random geometric (ad hoc)
- indexing of nodes is important
- closer to reality
- more challenging – open
- focus on network modification

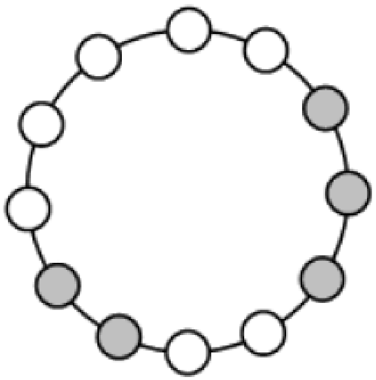
Regular (Lattice) Networks

- Finite – infinite (reference network models)
- **Regularity** $\rightarrow$ all vertices the **same** node **degree**
- *K*-regular graph $\rightarrow$ node degree is *k*
- Terminology: <u>Regular</u> (theory) – <u>Lattice</u> (applications)
- Emerging in:
  - Crystals (physics and chemistry)
  - Electrical power systems
  - Sensor networks (monitoring, computation)
  - Image processing
  - Optical networks
  - Material science (natural and composite materials)
  - Mobile cellular networks

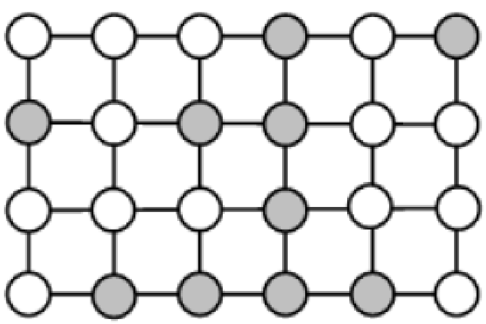
Examples of Lattice Networks



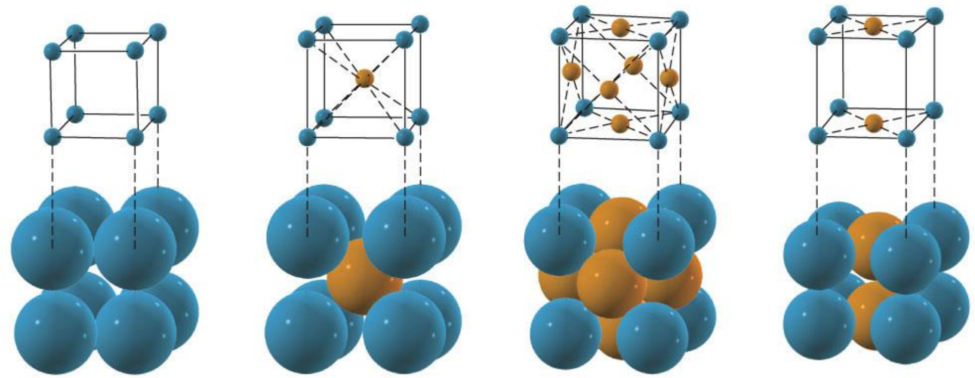
(a) chain (line) network



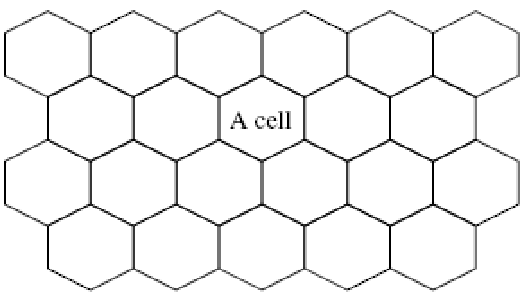
(b) ring network



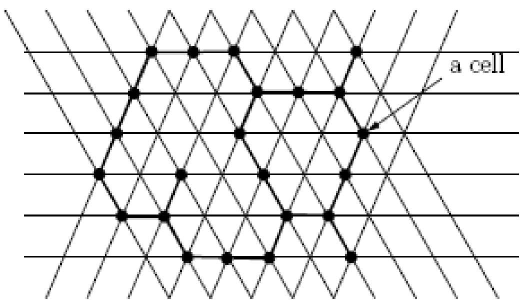
(c) 2D lattice network



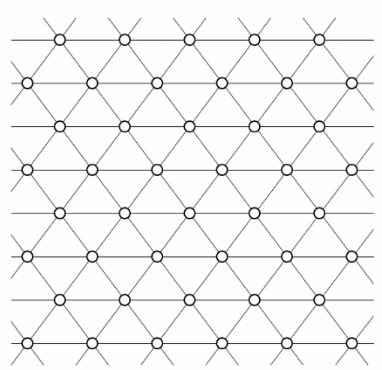
Primitive   Body-centered   Face-centered   Side-centered

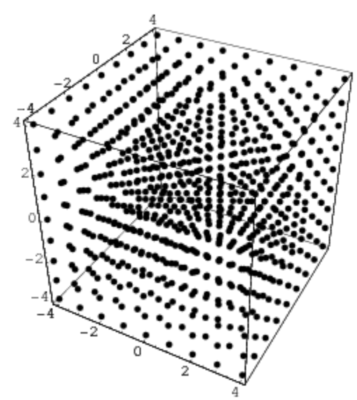


(a) cellular network



(b) triangle-free cellular network





Properties of Lattice Networks

- Deterministic degree distribution
  - Finite set of values

- All-units vector *j* is an eigenvector
- Other eigenvectors orthogonal to *j*
- CC can vary from 0 to 1 depending on degree & size
- Flat centrality

- Lattices are locally coupled
  - Locality of events – controlled propagation
- Distributed networks with low CC
- Difficult to implement dynamic processes that require global coordination (synchronization)

Applications of Lattices

- Cellular coverage schemes – mobile communications

- Sensor networks
- Smart grids
- Power grids

$\left.\begin{matrix}\\ \\ \\ \end{matrix}\right\}$ smart energy communications & management

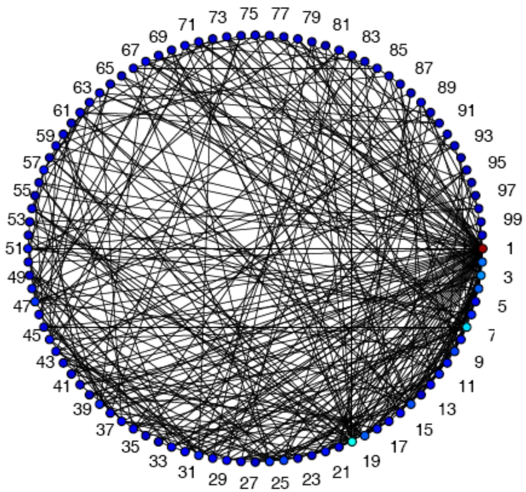
- Flexibility & strength – material science
  - Carbon fiber
  - Graphene

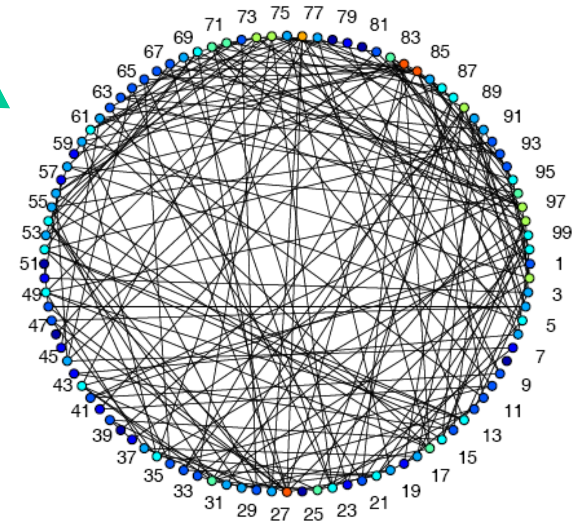
- Surveillance – security and cyber-security

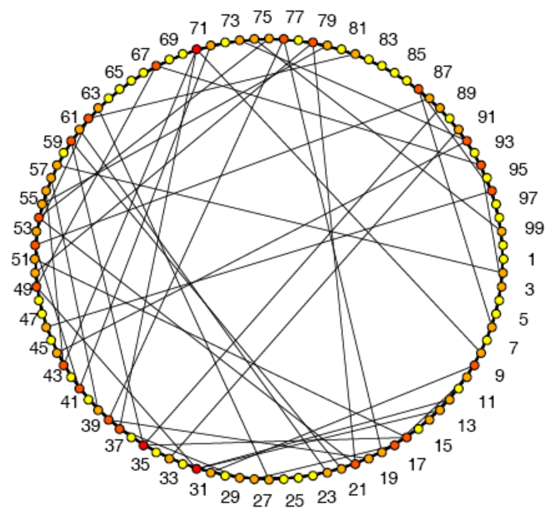
Random Graphs

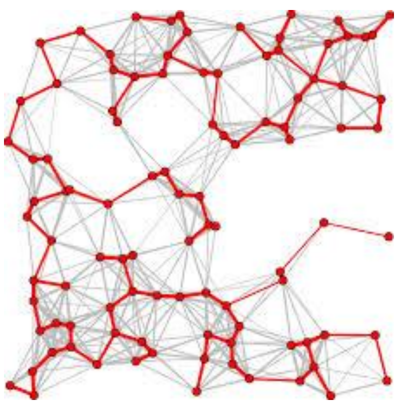
- **Random Graph (RG)**: nodes connected uniformly at random
- Probability distributions over graphs
  - Graph space becomes a probability space
- Two popular models:
  - Gilbert ***G(n,p)***
  - Erdos-Renyi ***G(N,M)***
- Relational types of graphs
  - All nodes can be potentially connected with each other
  - Idealized model, but good reference basis
  - Compare with spatial (RGG-multihop) graphs

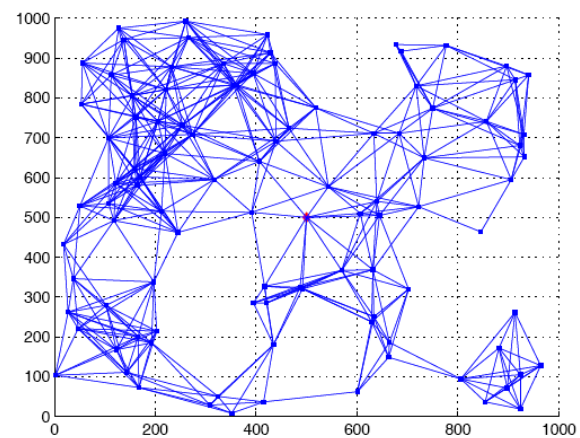
Find the Random Graphs











Gilbert Model *G(n,p)*

- Starting with a set of *n* isolated vertices, add successive edges between them at random
- Every possible edge occurs independently with probability $0 < p < 1$
- Probability of any particular RG with *m* edges:

  $p^m(1-p)^{N-m}$, where $N = \binom{n}{2}$
- The degree distribution is binomial

$$P(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k},$$

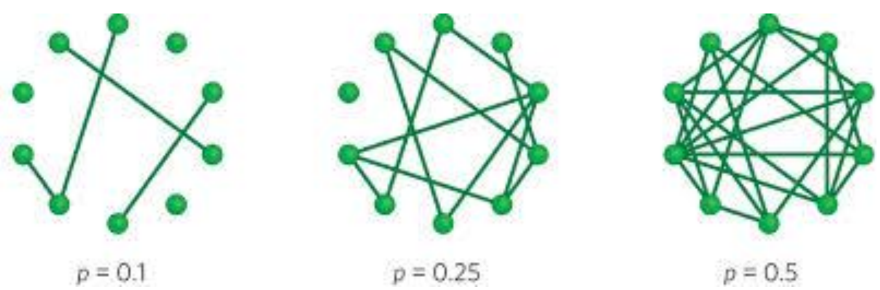
- Becomes Poisson for large *n* and *np* = constant

$$P(\deg(v) = k) \to \frac{(np)^k e^{-np}}{k!} \quad \text{as } n \to \infty \text{ and } np = \text{const},$$

Erdos-Renyi Model *G(n,M)*

- A graph is chosen uniformly at random from the collection of all graphs which have *n* nodes and *M* edges
  - The expected number of edges in *G*(*n*, *p*) is $\binom{n}{2}p$
- **Heuristic**: if $pn^2 \to \infty$ as *n* increases, *G*(*n*,*p*) behaves roughly as *G*(*N*,*M*), *N*=*n* with $M = \binom{n}{2}p$
- Probability that *G(N,M)* is precisely a fixed graph *H* on [*N*] with *M* edges:

$$\mathbb{P}_M(G_M = H) = \binom{N}{M}^{-1}$$



Random Graph Properties

- Theory of RG: studies typical properties that hold with high probability for graphs drawn from a particular distribution → asymptotic behavior
- Given a property *Q*, it is denoted that almost every (a.e.) graph in the probability space Ω_n consisting of graphs of order *n* has property Q if

$$\mathbb{P}(G \in \Omega_n : G \text{ has } Q) \to 1 \text{ as } n \to \infty$$

- A property *Q* of graphs is **monotone increasing** if *Q* is invariant under the addition of edges
- It is **monotone decreasing** if it is invariant under the deletion of edges

Threshold Behavior of Random Graphs

- Many RG properties exhibit phase transition → sharp change of behavior for increasing *n*
- Connectivity is the most characteristic
- For the *G(n,p)* model *p≥(logn+ω)/n* graph is asymptotically a.s connected
- Every non-trivial monotone property *A* has a threshold
- Sharp thresholds

$p_0 = p_0(n)$ is a *threshold* for a monotone property $\mathcal{A}$ if $\forall p(n)$

$$\Pr[\mathcal{G}_{n,p} \in \mathcal{A}] \to \begin{cases} 0, & \text{if } p/p_0 \to 0, \\ 1, & \text{if } p/p_0 \to \infty. \end{cases}$$

More Properties of Random Graphs

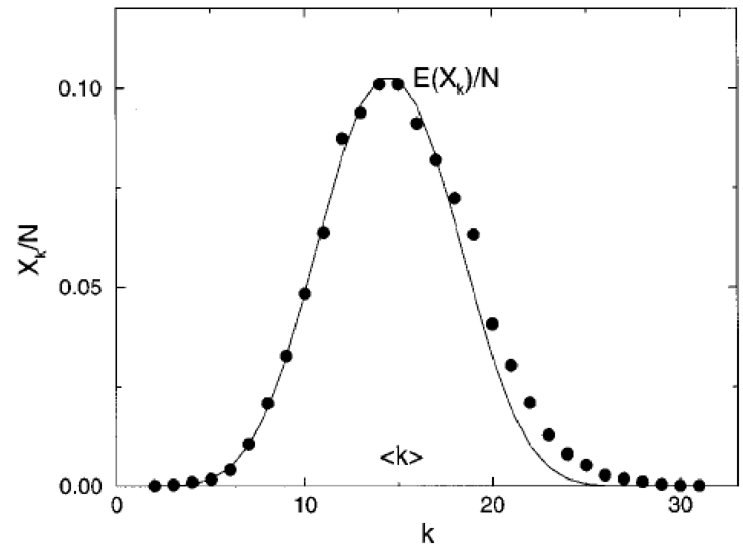
- Mean degree is *z=p(n-1)*
- Prob. vertex having degree k (Poisson distrib.)

- Mean size of the giant component $\langle s \rangle = \frac{1}{1 - z + zS}$
  - Giant component: largest connected component of graph

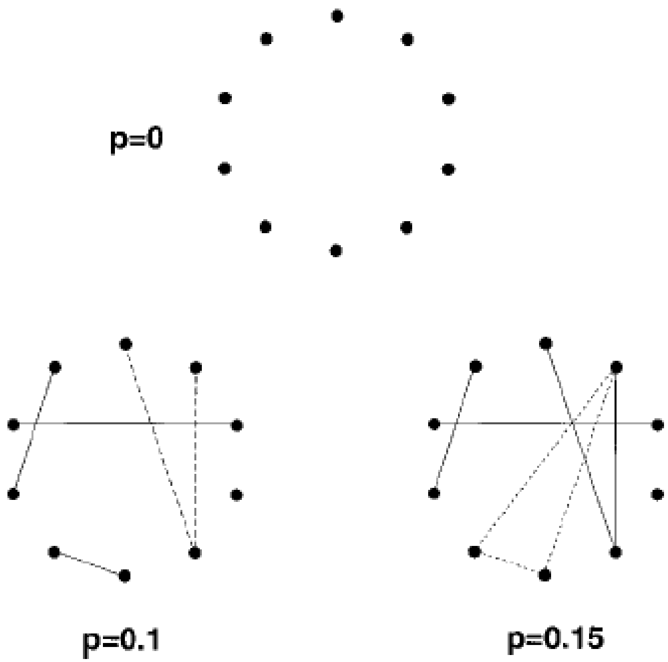
- Connectivity phase transition
  - Threshold phenomena

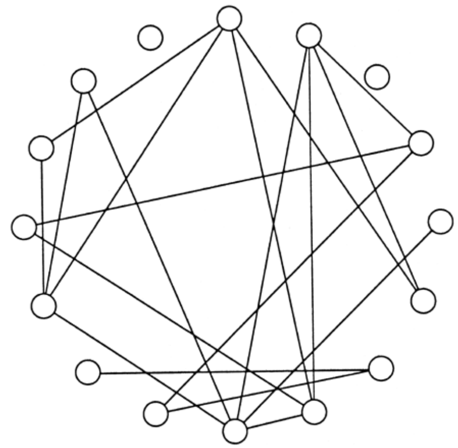
- Changing the scale does not affect results

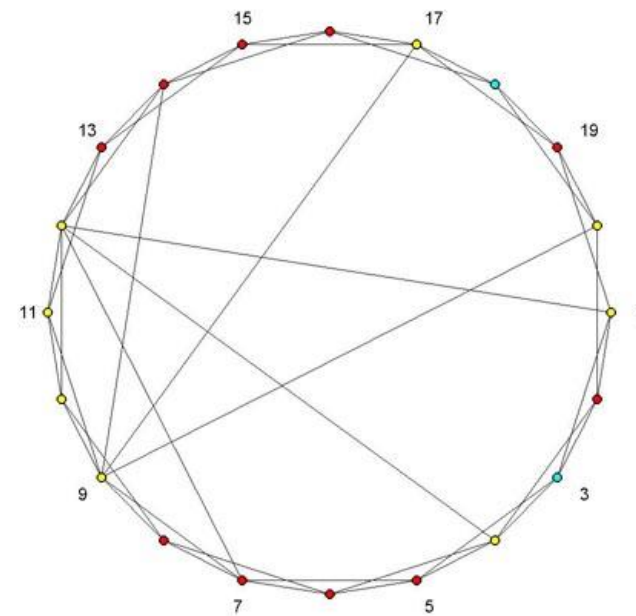
(N=10000, p=0.0015)

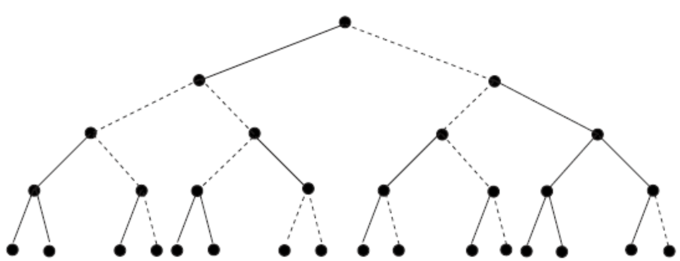


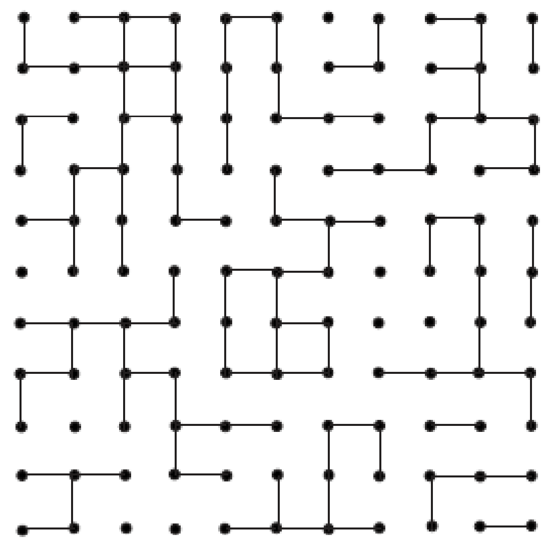
More Examples of Random Networks

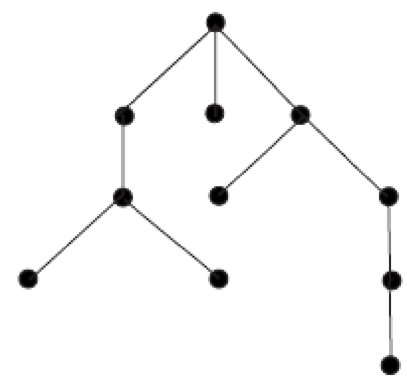












Applications of Random Graphs

- Model for social relations of people (social networks)
  - All actors potentially related – irrespective of distance
  - If random mixing → relations are random
- Random connections of users in peer-to-peer networks
  - Users are physically distant from each other
  - Directly connected at application layer – when connection
  - Examples: emule, skype, viber, email network
- Neural networks
  - Connections of neurons (brain, cortex, etc.)
  - Artificial neural networks

Random Regular Graphs – RRGs

- Regular networks and Random Networks are two extremes
- Uniform model $\rightarrow$ tough to analyze
- Algorithmic models
  - Degree-restricted process
  - Superposition models
- Random graphs demonstrating regularity in node distributions
- Probability space over all regular graphs

Properties of Random Regular Graphs

- Random Regular Graph (RRG) $G_{n,r}$ prob. space over r-regular graphs
- Essentially RRGs are RGs
    - Most of the properties hold almost surely (a.s.) , in the limit as n grows to infinity
- a.s. a $G_{n,r}$ RRG is r-connected
- For d satisfying

$$(r-1)^{d-1} \geq (2+\epsilon) rn \ln n$$

  the diameter is a.s. at most d
- The distribution of short cycles in RRGs is a.s. Poisson distributed with mean

$$\lambda_i = \frac{(r-1)^i}{2i}$$

Generalized Random Networks

- Generalized Random Graphs (GRGs) are essentially RGs
- Properties hold almost surely (a.s.) in the limit for growing network populations
- In addition: degree distribution is not Poisson (as in RGs)
- Exhibits correlations $\rightarrow$ closer to observed realistic network distributions
    - Empirical distributions fitted into probabilistic ones
- Connections remain random
- More effective for affiliation and citation networks
    - capture correlation due to relations, but retain fundamental random character

Random Geometric Graphs (RGGs)

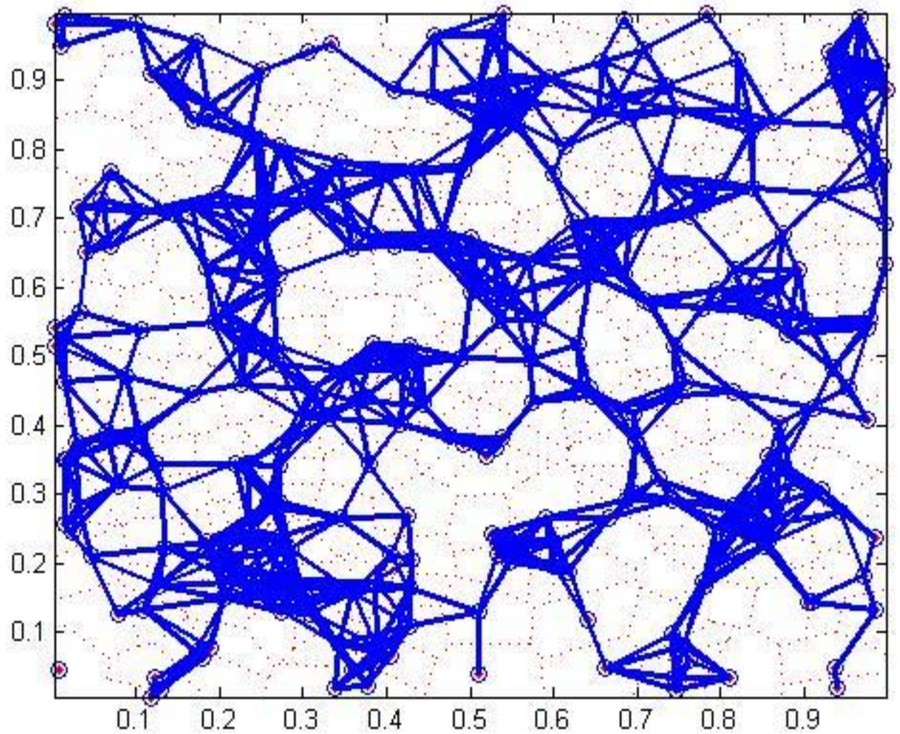
- Probabilistic spatial models
- Nodes randomly dispersed over network region
- Connections $\rightarrow$ dictated by distances
- Distances potentially defined in multiple ways:
  - Coordinates
  - Similarities
  - Context
  - Any valid measure metric

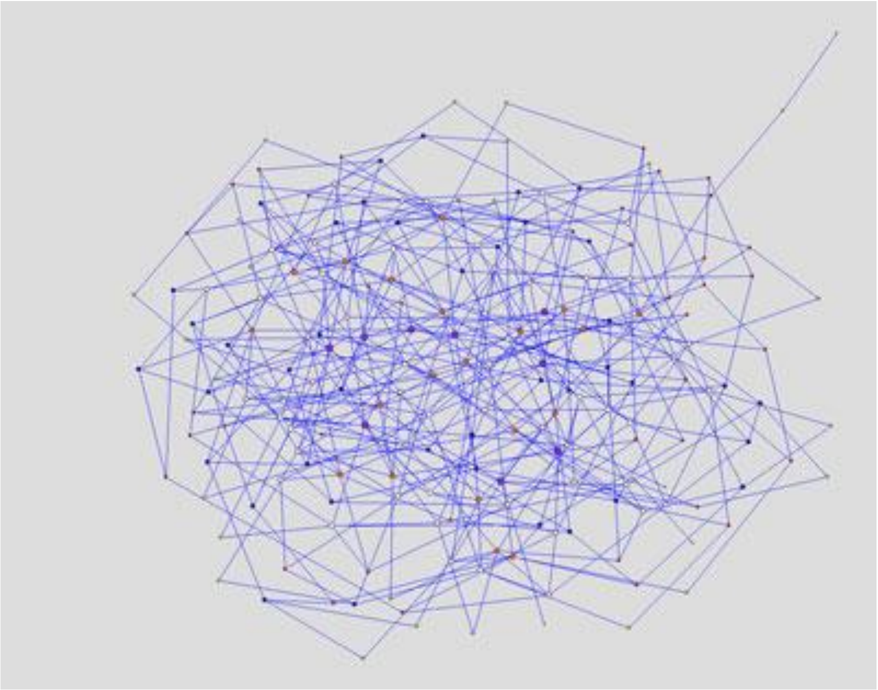
- Distance $\operatorname{dist}(X, Y) = \min \left\{ \|X + z - Y\| : z \in \mathbb{Z}^d \right\}$
- Diameter $\operatorname{diam}(A) := \sup \{ \|x - y\| : x \in A, y \in A \}$

RGG – RG Comparison

RGG

RG





RGG $G(n,r,\ell)$ with *n* nodes, radius *r*, and label ℓ for each node; square deployment region $L \times L$

Connectivity Properties of RGGs

- Typically long average path lengths → distributed network
  - significant delay e2e packet delivery (NET layer)
- Expected # of neighbors for each node: $\frac{\pi r^2}{L^2} n$
- Threshold behavior:
  - coverage

Theorem 3.2. *Consider the $G(n,r,\ell)$ model and let $r = r(\ell) = \ell^{\epsilon} f(\ell)$, for some $0 \leqslant \epsilon < 1$, and $f(\ell)$ is a function which grows strictly slower than any function of type ℓ^{γ} where $\gamma > 0$. Let $n = n(\ell) = \omega(1)$.*

- *$G(n,r,\ell)$ has full area coverage a.a.s. if $r^2 n \geqslant \ell^2((\frac{1}{2} - \frac{1}{2}\epsilon)\ln \ell + \frac{1}{2}\ln\ln \ell + h(\ell))$, for any $h(\ell) \to \infty$.*
- *$G(n,r,\ell)$ does not have full area coverage a.a.s. if $r^2 n \leqslant \ell^2((\frac{1}{2} - \frac{1}{2}\epsilon)\ln \ell + \frac{1}{2}\ln\ln \ell + g(\ell))$, for any $g(\ell) \to -\infty$.*

  - connectivity

Theorem 3.4. *Consider the $G(n,r,\ell)$ model and let $r = r(\ell) = \Theta(\ell^{\epsilon} f(\ell))$, for some $0 \leqslant \epsilon < 1$, and $f(\ell)$ is a function which grows strictly slower than any function of type ℓ^{γ} where $\gamma > 0$. Let $n = n(\ell) = \omega(1)$. Given any two constants $c_1 > 2 - 2\epsilon$ and $c_0 < \frac{1}{2} - \frac{1}{2}\epsilon$,*

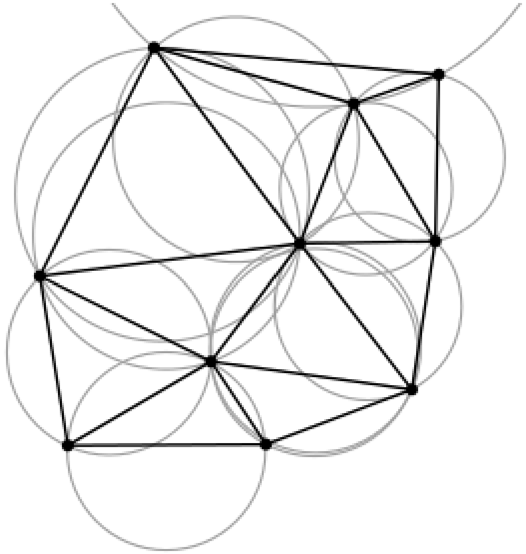
- *$G(n,r,\ell)$ is connected a.a.s. if $r^2 n \geqslant c_1 \ell^2 \ln \ell$, and*
- *$G(n,r,\ell)$ is disconnected a.a.s. if $r^2 n \leqslant c_0 \ell^2 \ln \ell$.*

  - stretch

Theorem 3.5. *In $G(n,r,\ell)$ let $r^2 n = k\ell^2 \ln \ell$, and $r = r(\ell) = \Theta(\ell^{\epsilon} f(\ell))$, for some $0 \leqslant \epsilon < 1$, as before. Let $0 < \alpha \leqslant 1$ be a fixed constant. Then for any constant $k > \frac{22(1-\epsilon)}{\alpha}$, the stretch is $1 + \alpha/2$ a.a.s. Further, if we consider only the subset F of nodes such that $D(u,v) = \omega(r)$ (i.e., strictly larger than r) for all $u, v \in F$ then the stretch restricted to this subset is 1 a.a.s.*

Generalizations of RGGs

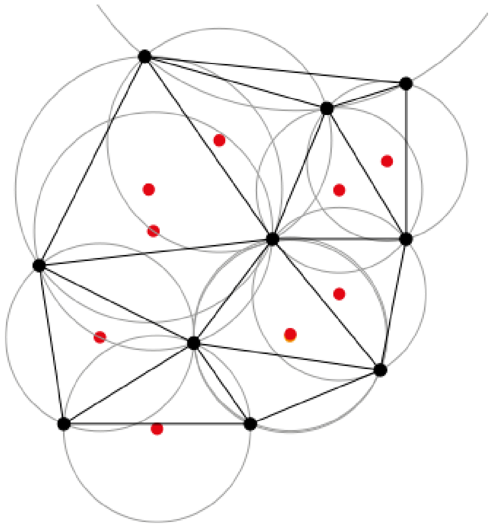
- Nodes $\rightarrow$ general point processes (PP)
  - Poisson (intensity λ) $\mathbb{P}(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$
  - Binomial, etc.
- Percolation problems

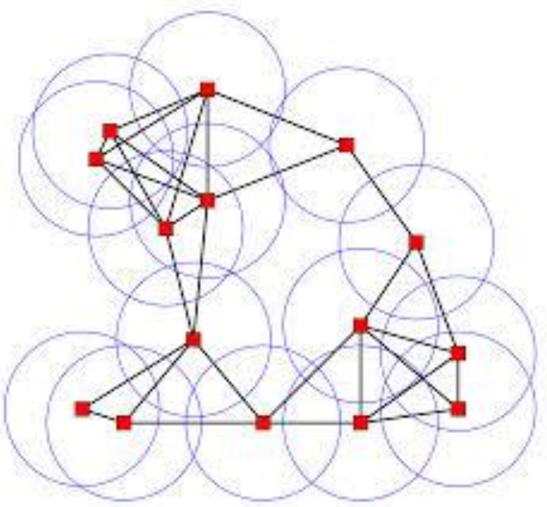


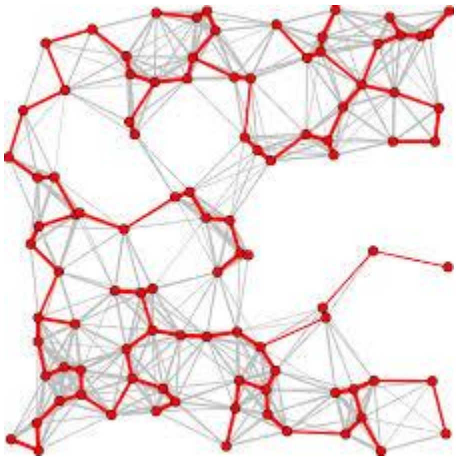
- Voronoi tesselations (Delauney triangulations)
  - Given random point sets, find equidistant clusters with these points at their centers
  - DT for a set ***P*** of points in a plane is a triangulation *DT(**P**)* s.t. no point in ***P*** is inside the circumcircle of any triangle in *DT(**P**)*
    - *Max. min. angle of all the angles of triangles in DT; avoid skinny triangles*

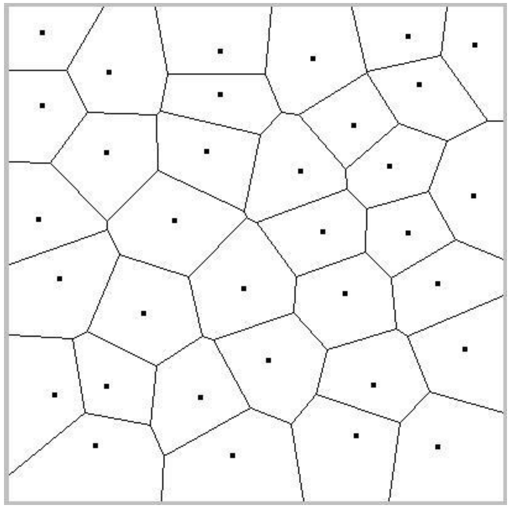
Applications of Random Geometric Graphs

- Wireless mobile (multi-hop) networks
  - Ad hoc, Sensor, Mesh, Vehicular
- Statistical data processing
- Vehicular traffic engineering
- Geographic Information Systems (GIS)



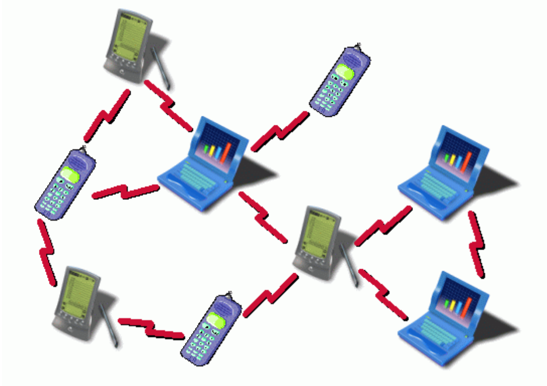






Wireless Mobile Multihop Networks

- Characteristics
  - Decentralized and dynamically organized structures
  - Nodes communicate either directly or through intermediate relay-nodes
- Examples
  - Ad hoc and mobile ad hoc networks (MANETs)
  - Wireless sensor networks (WSNs)
  - Wireless mesh networks (WMNs)



- Results
  - Capacity: logarithmic scaling
  - Mobility aids capacity (capacity-delay tradeoff)
  - Not very robust

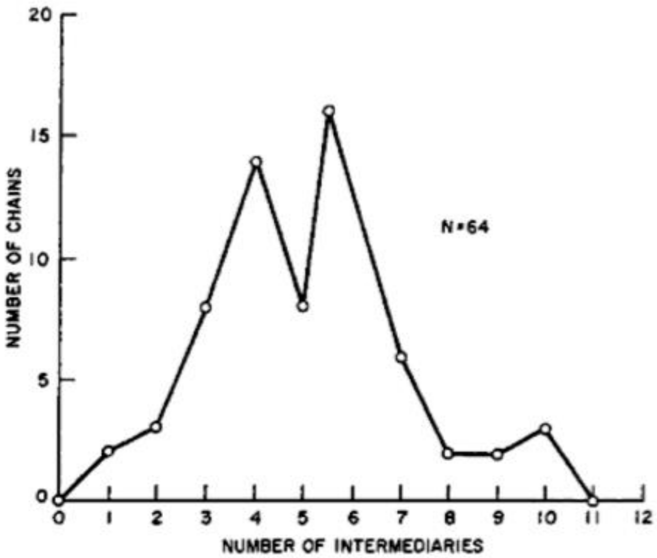
The nodes in multi-hop networks are inter-dependent, rather than independent

"How close are you to Trump?"

- **Social networks tend to have very short paths between arbitrary pais of people...**

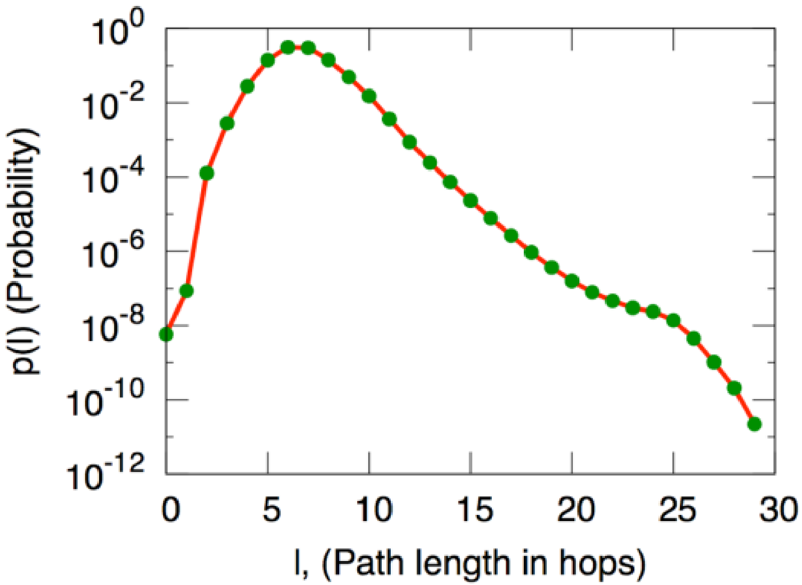
  Really!!!!

- First experiment done by Stanley Milgram in 1960s (research budget $680)
  - 296 randomly chosen starters. Asked to send a letter to a target (in Boston), by forwarding to someone they know personally and so on. Number of steps counted.
- median hop number of 6 for successful chains – six degrees of separation
  - This study has since been largely discredited



Six Degrees of Separation

- Modern experiment by Leskovec and Horvitz in 2008
- Look at the 240 million user accounts of Microsoft Instant Messenger
- Complete snapshop – no missing data
- Found a giant component with very small distances
- A random sample of 1000 users were tested and performed Breadth-first search
  - Why do they look only at a sample? Due to time and computational constraints and feasibility
- Estimated average distance of 6.6, median of 7
- **Need for another network model → SW**

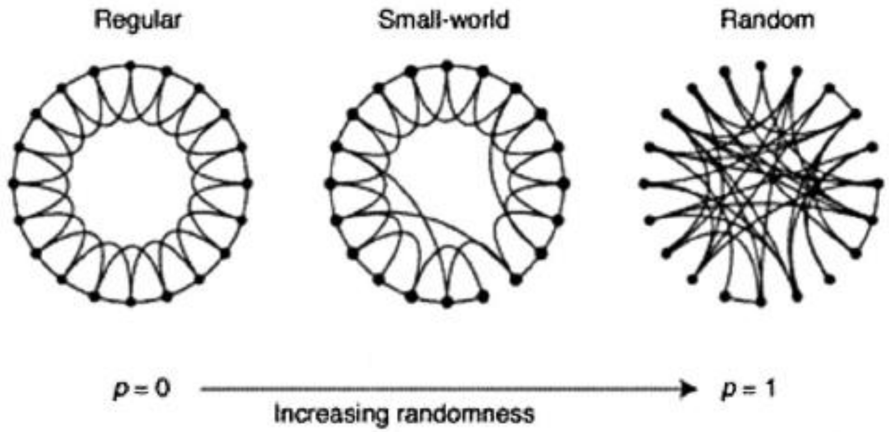


Small-world Networks

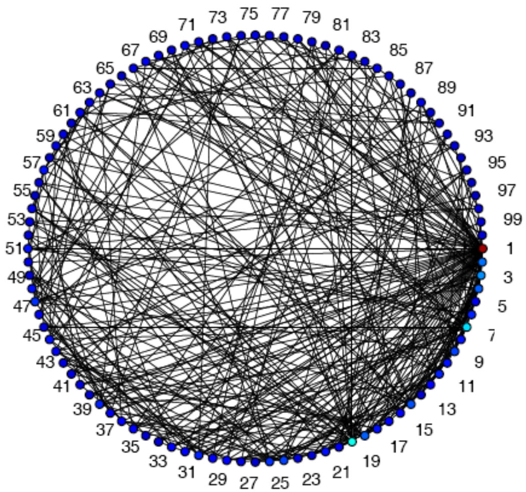
- Empirical definition

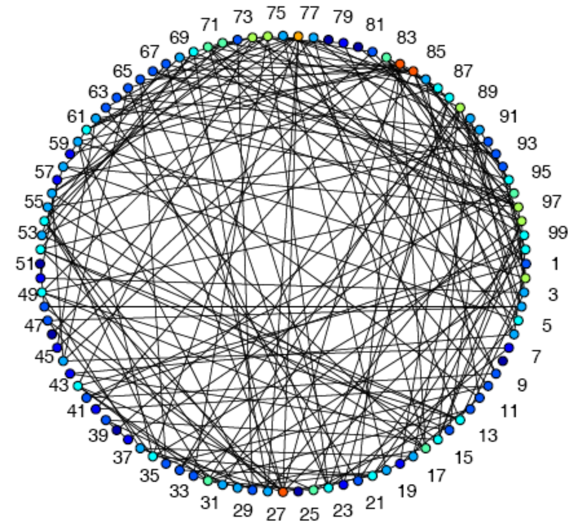
  most nodes not neighbors of one another, but most nodes can be reached from every other by a small number of hops

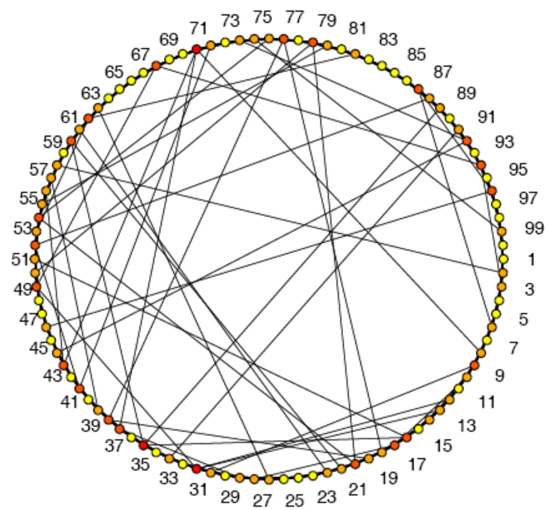
- Obtained evolutionary from ordered lattices
- **Watts-Strogatz** model
  - Start from an ordered lattice
  - Randomly rewire each edge with prob. *p* excluding self-connections and duplicate edges
  - Arbitrary long-range edges maybe added

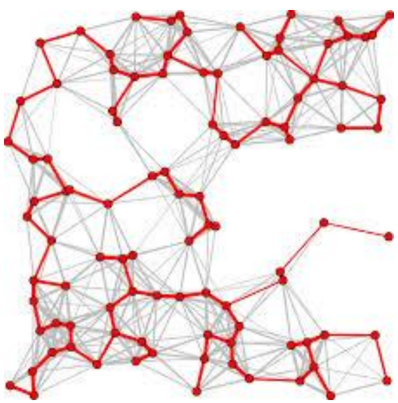


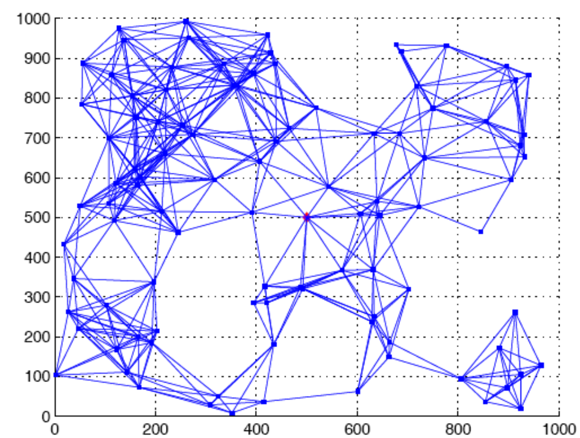
Find the Small-world Graphs











Applications of Small-world Networks

- Socio-economic system development

  Opinion formation

  Leadership

- Epidemiology

  Virus

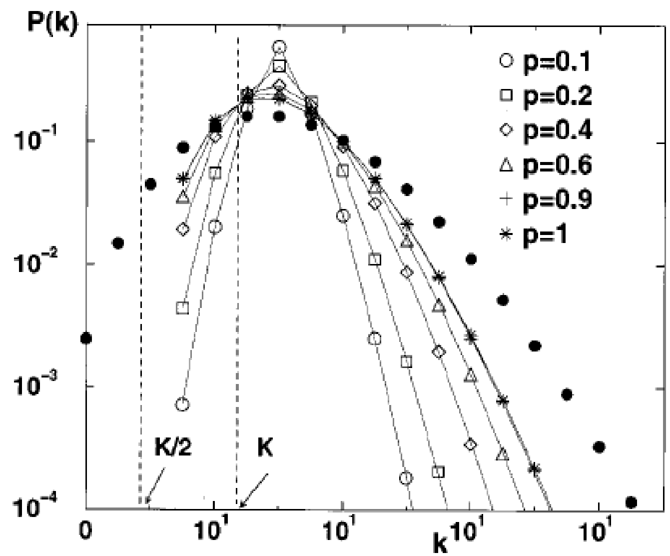
  alcohol

- Peer-to-peer networks
- Affiliation networks

  Affiliation spreading

Properties of Small-world Networks

- Small average path length
- Relative high clustering coefficient
- Relative homogeneous topology
  - Most nodes having approximately the same number of edges
- Spectrum depends on *p*
- High number of triangles in the network



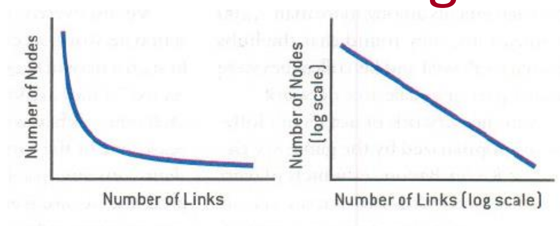
"Who's the most popular?"

- Perhaps the most popular question in social circles is "who's the most popular?"
- Which airport is the busiest? The most well-connected? Depends on country/airliner?
- Why epidemics break abruptly? What are the speed of epidemics? And why pandemia can be reality?
  - Black plague: ~ 75 to 200 mil. people dead in Europe (1346–53)
  - "I am Legend" and "World War Z" can be reality….really!
- And many more questions related to popularity & robustness of networks….one answer:

Scale-free Networks

Scale-free (Exponential) Networks

- Power-law distributed small-world network $P(k) \sim k^{-\gamma}$
  - Small percentage of nodes with large degree values
  - Majority of nodes with small degree values

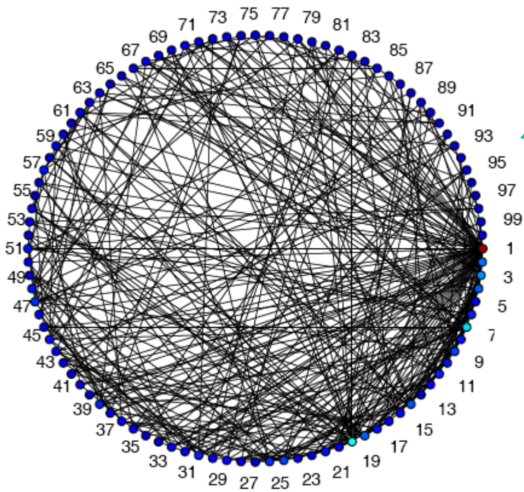


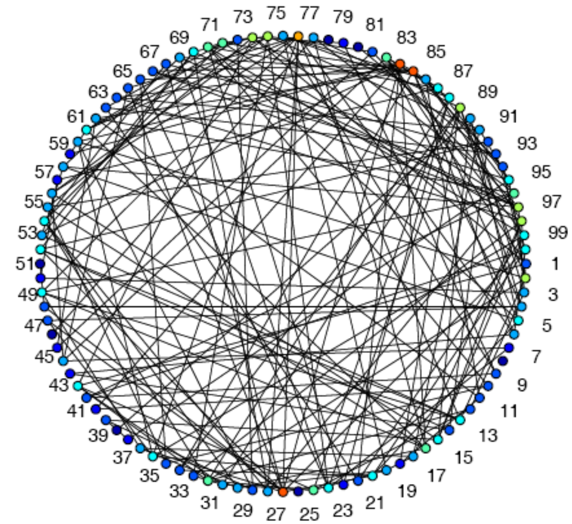
- Obtained by growth + preferential attachment

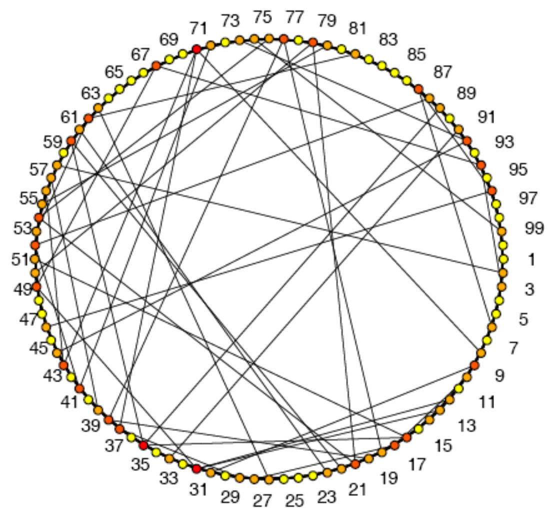
$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

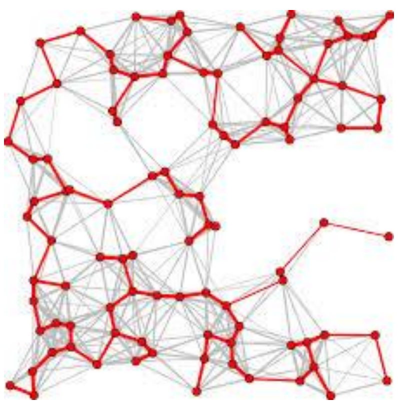
- Many empirically observed networks appear to be scale-free $\rightarrow$ seems the most natural emerging network structure

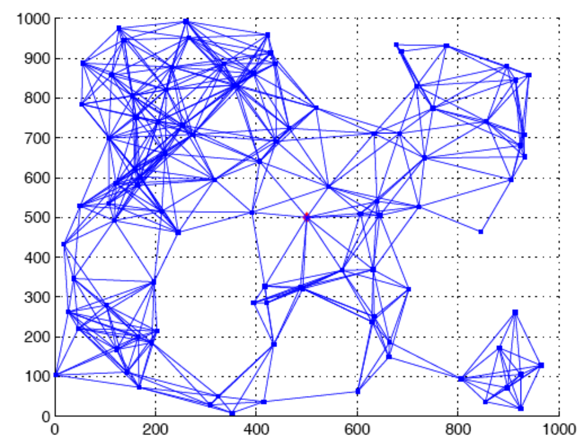
Find the Scale-free Graphs



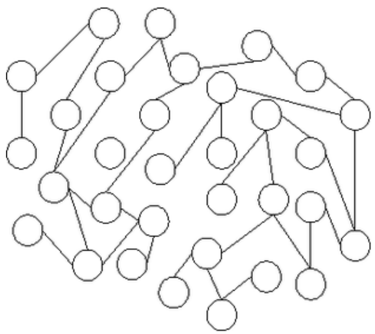




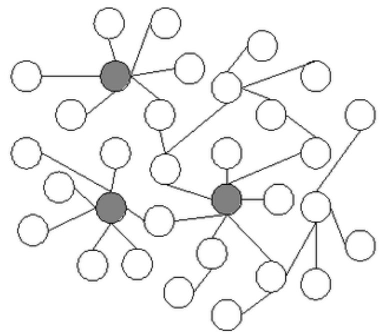




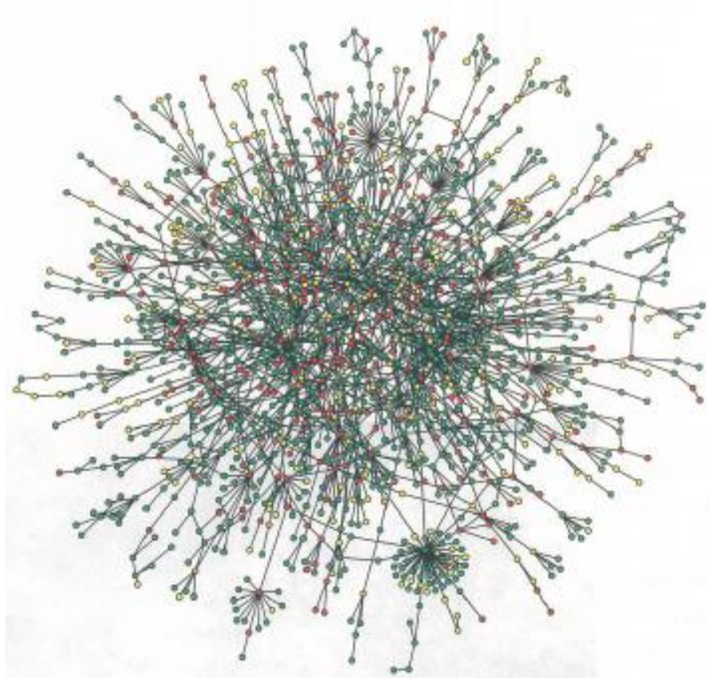
Examples of Scale-free Networks

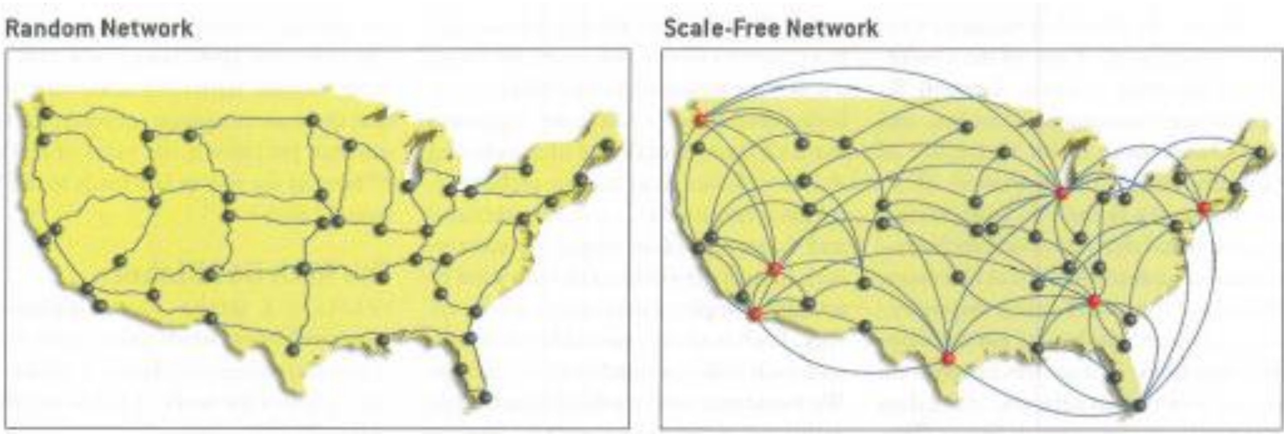


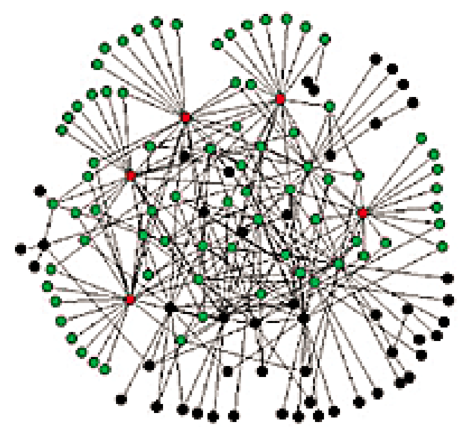
(a) Random network



(b) Scale-free network







Applications of Scale-free Networks

- Internet/WWW
- Science collaboration graphs
- Hollywood co-starring graphs
- Cellular networks
- Citation networks
- transcriptional networks in which genes correspond to nodes
- road maps
- food chains
- electric power grids
- metabolite processing networks
- neural networks
- voter networks
- telephone call graphs
- social influence networks
- Human sexual contacts

Properties of Scale-free Networks

- Power law degree distribution $P(k) \sim k^{-\gamma}$
- Exponent γ is empirically computed
- Clustering coefficient ~x5 than in RGs
- Smaller average path length than a RG
  - ‘six degrees of separation’
- Correlations occur spontaneously between connected nodes
- Percolation and phase transitions
- Robustness under random attacks
- Vulnerable to targeted attacks

Network Feature Comparison

Network type	Degree distrib.	Av. Path Len.	Centrality
Regular	dirac function	constant	constant
S-W	heavy-tailed	small	varying
S-F	power-law	small	varying
RG	Poisson	average	uniform
RGG	uniform	long	uniform

Next in SNA ….

- Complex network analysis metrics
- Community detection
- Evolutionary computing
- Epidemics
  - Infection models
  - Epidemic thresholds