707.000
Web Science and Web Technology
„Network Evolution and Processes“

How do networks evolve? Are there „natural laws“ governing the evolution of certain networks?

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: http://www.kmi.tugraz.at/staff/markus

Overview

Agenda

- Network Creation and Evolution
  - Random Networks, Configuration Model, Barabasi and Albert
- Network Processes
  - The SIR Model

Motivation

How do networks evolve? Are there „natural laws“ governing the evolution of certain networks?

With demos from http://www-personal.umich.edu/~ladamic/NetLogo/

Examples of network evolution:

- „Invites“ to join GMail
- „Invites“ to buy Chumby
- „Invites“ to join Joost
- Vaccination strategies for epidemics
- …

Background
[Newman 2003]

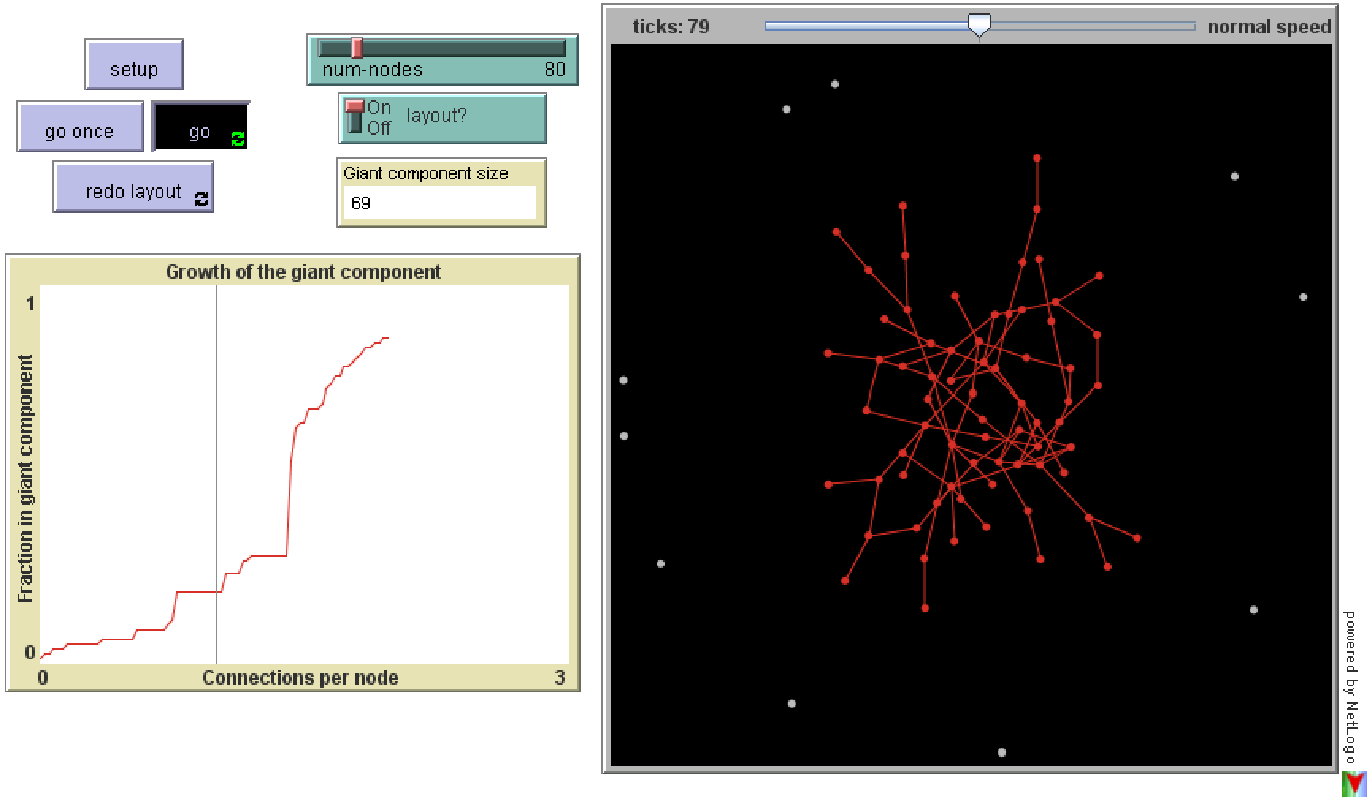
- First example of a scale-free network (Price):
  - Network of citations between scientific papers
  - Both in- and out-degrees had power-law distributions
- Answered the question: How do power law distributions emerge?
  - “the rich get richer”
  - In other words: the amount you get goes up with the amount you already have
- The “Matthew affect”
  - “For to every one that hath shall be given” (Matthew 25:29)
  - (in german ~ “wer hat dem wird gegeben”)
- Other labels
  - Cumulative advantage
  - Preferential attachment
- Evident in scientific paper citations
  - The rate at which a paper gets new citations is proportional to the number that it already has

Why do you think is that?

Giant Components - Demo

- When do Giant Components emerge?

http://ccl.northwestern.edu/netlogo/models/GiantComponent



Two Assumptions [Leskovec 2006]

"Conventional Wisdom" that networks that evolve are characterized by

- Constant average degree
  - Edges grow linearly with edges
- Slowly growing diameter
  - Growing diameter with the addition of new nodes

Empirical observations show that

- Networks are becoming denser over time (densification power laws)
- Effective diameter is in many cases decreasing as networks grow (shrinking diameter)

Empirical Observation: Densification [Leskovec 2006]

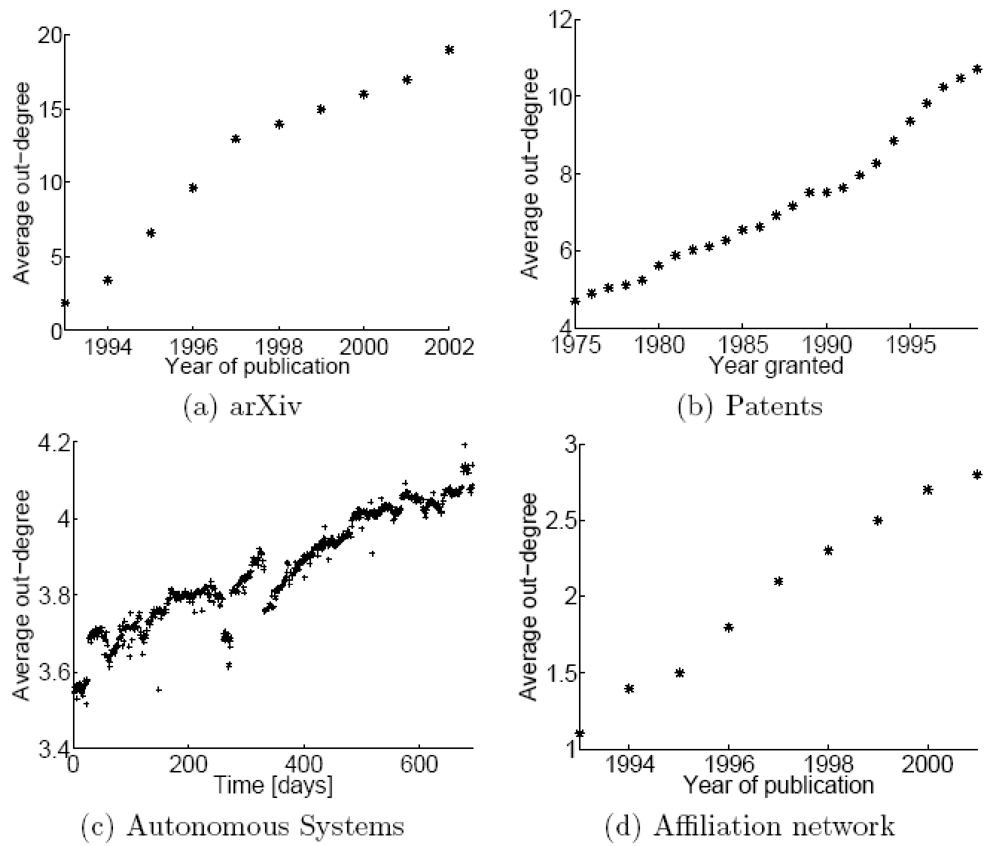


Figure 1: The average node out-degree over time. Notice that it increases, in all 4 datasets. That is, all graphs are *densifying*.

Empirical Observation: Densification [Leskovec 2006]

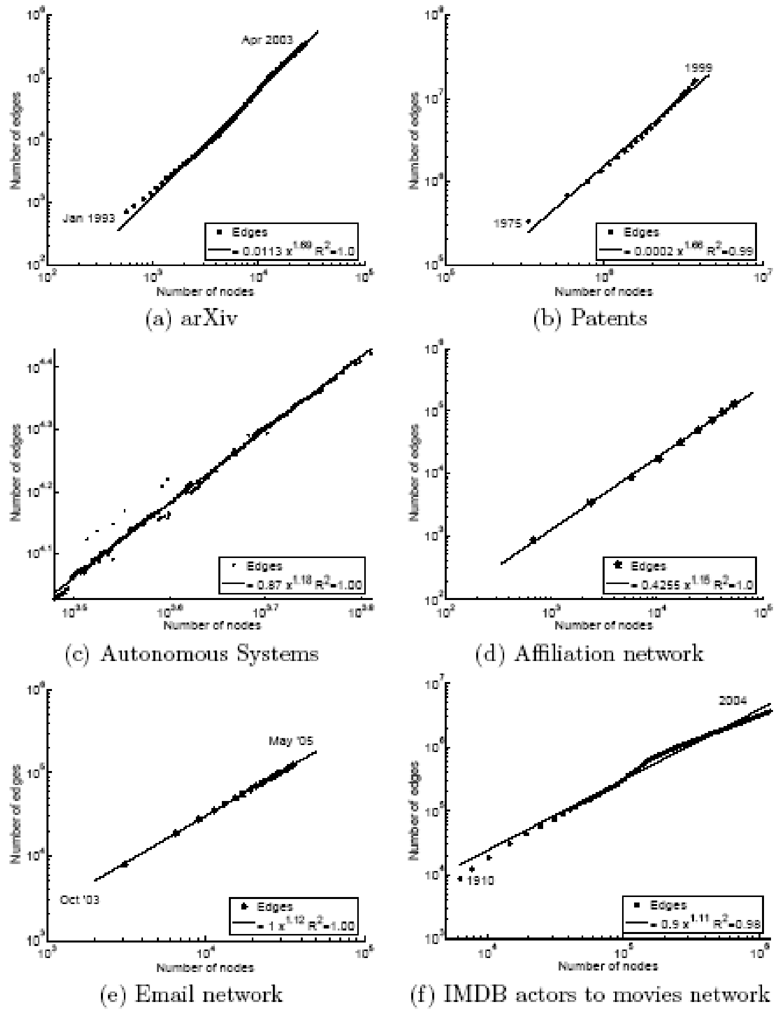


Figure 2: Number of edges $e(t)$ versus number of nodes $n(t)$, in log-log scales, for several graphs. All 4 graphs obey the Densification Power Law, with a consistently good fit. Slopes: $a = 1.68, 1.66, 1.18, 1.15, 1.12$, and 1.11 respectively.

Empirical Observation: Effective Diameter [Leskovec 2006]

Effective diameter:

The minimum distance d such that at least 90% of the connected node pairs are at distance at most d

Decreasing diameter over time

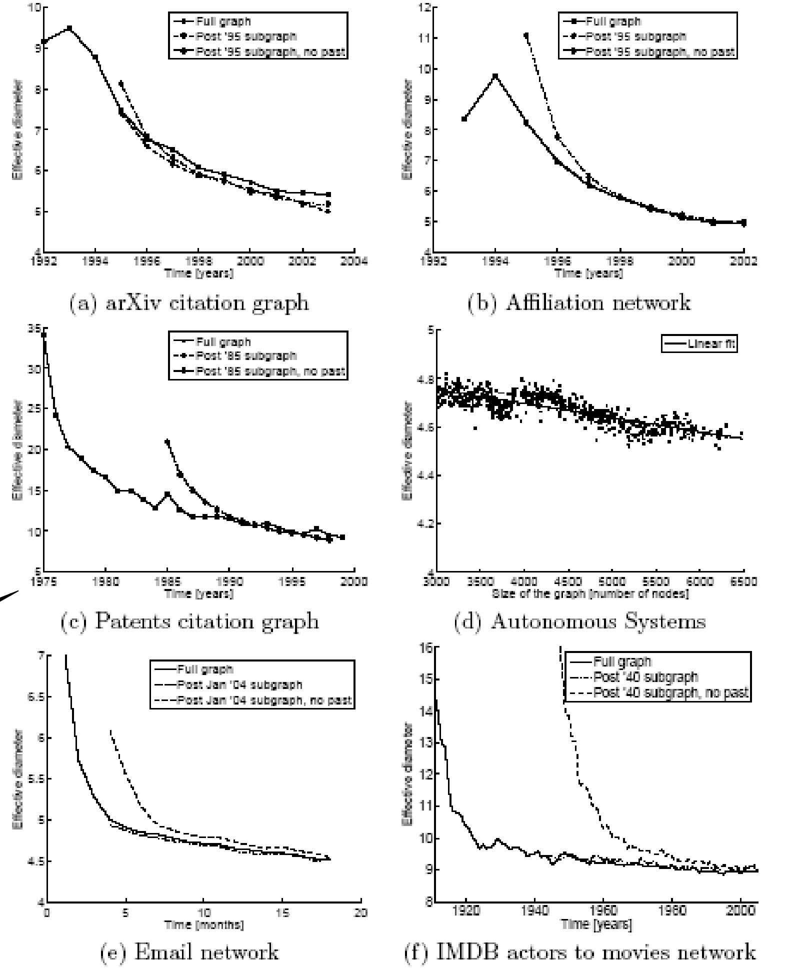


Figure 3: The effective diameter over time for 6 different datasets. Notice consistent decrease of diameter over time.

Motivation
[Leskovec 2006]

What underlying processes cause a graph to

1. systematically densify?
2. experience a decrease in effective diameter even as its size increases?

But first, let’s take a step back

Graph Generators [Leskovec 2006]

Why are we interested in simulating graph evolution?

"What if we could develop algorithms that are capable of constructing networks that exhibit similar characteristics as observed in "real-world" networks?"

We could do interesting things, such as:

- **Extrapolations**
  - predicting future network development
- **Sampling**
  - Drawing a sample and generalizing to the entire population
- **Abnormality detection**
  - Identifying deviations from "normal" network behaviour
- **Simulation**
  - Exploring "what if" scenarios, e.g. deletion of hubs, network resilience

Simple Graph Generators [Newman 2003]

Can we develop an algorithm that constructs random graphs?

Algorithm:
Take some number n of vertices and connect each pair (or not) with probability p (or 1-p).

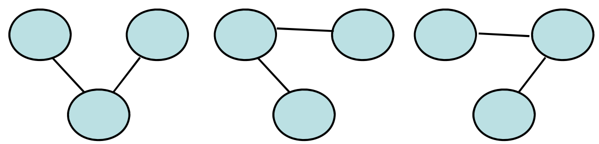
Done!

The Erdos-Renyi / Poisson random Graph

G(n,m) the set of all graphs having n vertices and m edges, each possible graph appearing with equal probability

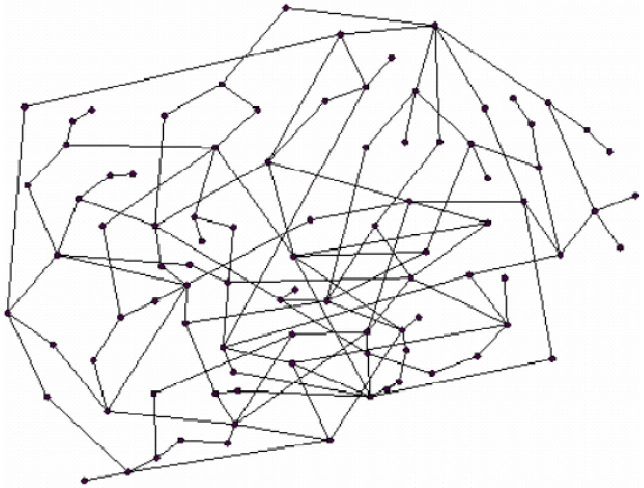
For example: G(3,2) is the set of all three graphs having 3 vertices and 2 edges, each graph has probability 1/3

->Does not mimic reality

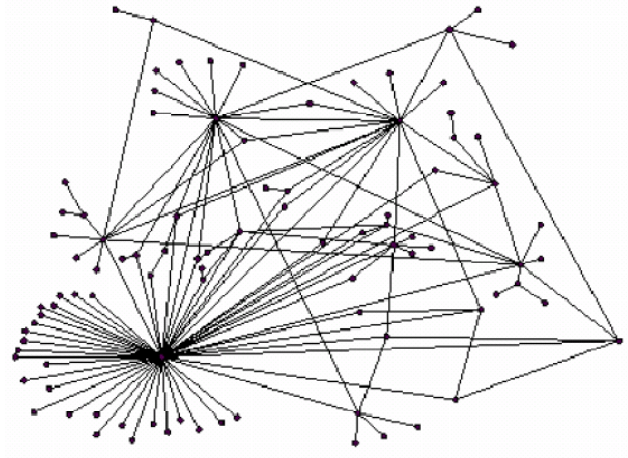


CMU SCS

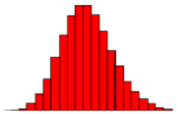
Poisson vs. Scale-free network



Poisson network
(Erdos-Renyi random graph)

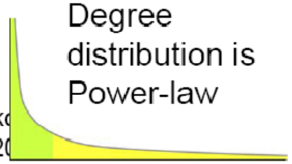


Scale-free (power-law) network



Degree distribution is Poisson

Faloutsos&Lesk
ECML/PKDD 2



Degree distribution is Power-law

Function is scale free if:
$f(ax) = c\,f(x)$

Faloutsos / Leskovec
ECML/PKDD 2007

Random Graphs [Faloutsos / Leskovec ECML/PKDD 2007]

- Pros:
  - Simple model
  - Phase transitions (giant component with avg. degree >1)
  - Giant component
- Cons:
  - Degree distribution
  - No community structure
  - No degree correlations
- Extensions:

Configuration model

- Random graphs with arbitrary degree sequence

The Configuration Model

Consider the model defined in the following way.

We specify a degree distribution p_k, such that p_k is the fraction of vertices in the network having degree k.

We choose a degree sequence, which is a set of n values of the degrees k_i of vertices i = 1 . . . n, from this distribution. We can think of this as giving each vertex i in our graph k_i “stubs” or “spokes” sticking out of it, which are the ends of edges-to-be.

[Newman 2003]

The Configuration Model

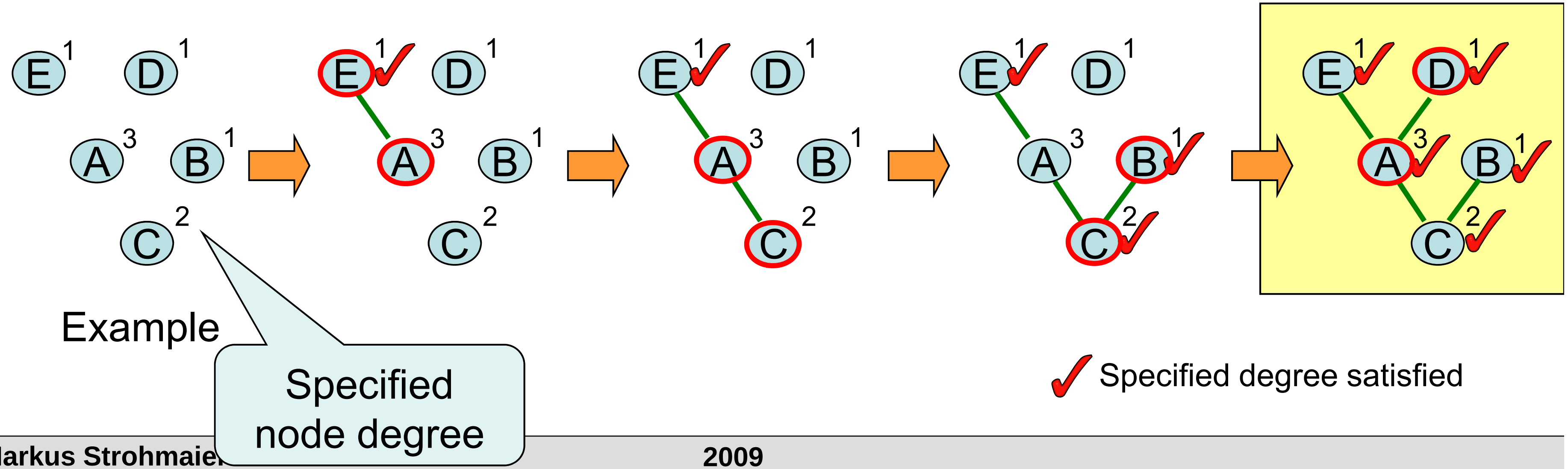
Then we choose pairs of stubs at random from the network and connect them together. It is straightforward to demonstrate that this process generates every possible topology of a graph with the given degree sequence with equal probability.

The configuration model is defined as the ensemble of graphs so produced, with each having equal weight.

[Newman 2003]

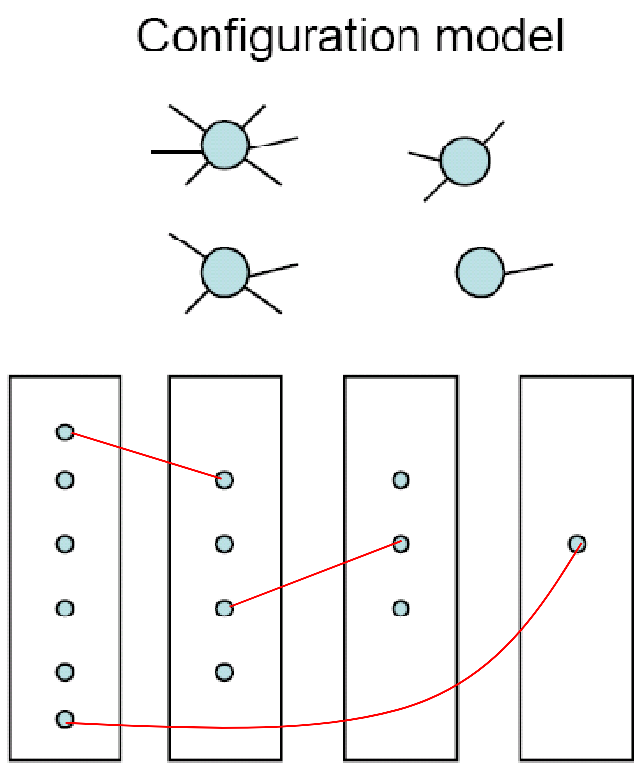
The Configuration Model: Example

1. Define a degree distribution (e.g. 3,2,1,1,1)
2. Specify degrees for each node, based on the degree distribution (e.g. A->3, B->2, C->1, D->1, E->1)
3. Insert an edge between two arbitrary nodes in your node set that have not satisfied their specified degree yet.
4. Repeat step 3 until all node degrees are satisfied.



The Configuration Model: Example II

Another perspective:



Example

Faloutsos / Leskovec
ECML/PKDD 2007

The Configuration Model

- Can reproduce networks with power-law distributions
  - Accepts arbitrary degree distributions as input

- Does not explain the natural emergence of power law networks
- Does not explain network growth / evolution

Generating Scale Free Networks

[Barabasi and Albert 1999]

To incorporate the **growing character of the network**, starting with a small number (m_0) of vertices, **at every time step we add a new vertex** with $m(\leq m_0)$ edges that link the new vertex to m different vertices already present in the system.

To incorporate preferential attachment, we assume that the probability Π that a new vertex will be connected to vertex i depends on the connectivity k_i of that vertex, so that

Probability of a new vertex attaching to a vertex i with degree k

$$\Pi(k_i) = k_i / \sum_j k_j$$

Degree of vertex i

The sum of all vertices' degrees

In other words: the probability is the degree of vertex i divided by the sum of all nodes' degrees

After t time steps, the model leads to a random network with $t+m_0$ vertices and mt edges.

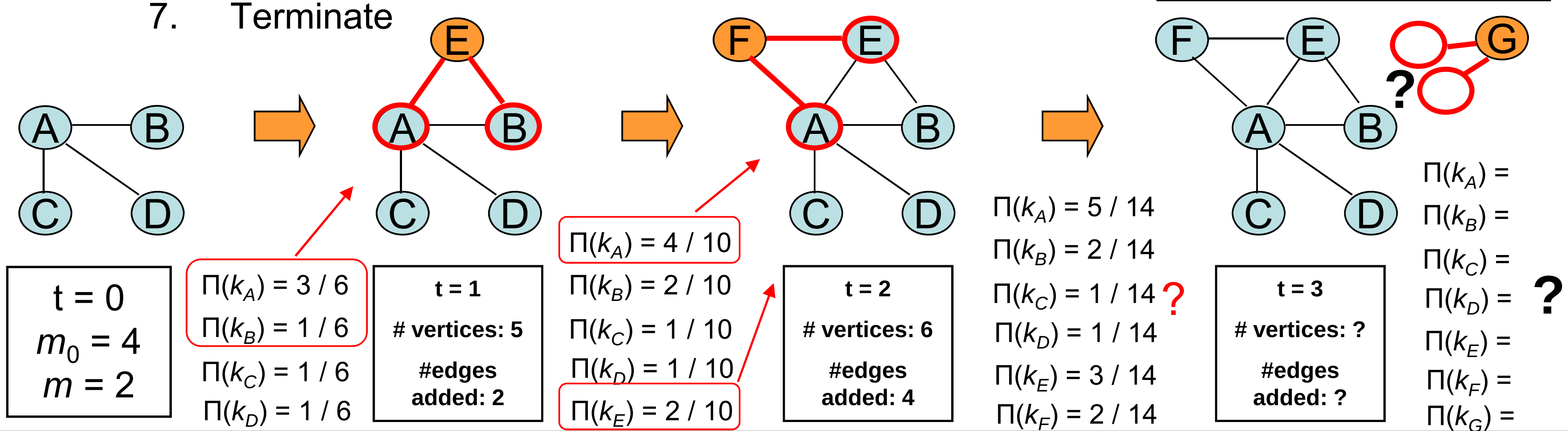
This network evolves into a scale-invariant state following a power law (satisfies the two conditions: Growth and Preferential Attachment).

Generating Scale Free Networks
[Barabasi and Albert 1999]

Example:

1. Specify a starting network with a given number of vertices m_0 and an initial set of edges (e.g.: #edges = 3); initialize t=0
2. Define the number of vertices a new node is required to link to (e.g. m=2)
3. Calculate the probabilities Π that a new vertex will be connected to vertex *i* by calculating $\Pi(k_i) = k_i / \sum_j k_j$
4. Add the new vertex. Add edges according to the calculated probabilities and m
5. Set t = t+1
6. While t≤ 3 Goto Step 3.
7. Terminate

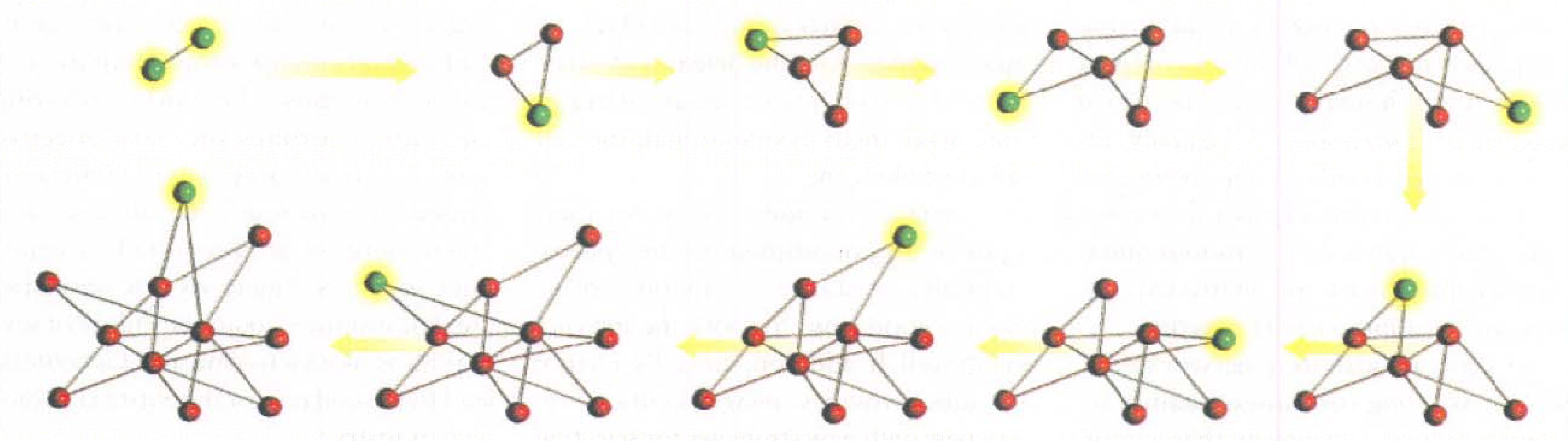
at time t: $t+m_0$ vertices
at time t: mt edges added



Generating Scale Free Networks
[Barabasi and Albert 2003]

BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



Generating Scale Free Networks

[Barabasi and Albert 1999]

Because of preferential attachment, a vertex that acquires more connections than another one will increase its connectivity at a higher rate; thus, an **initial difference** in the connectivity between two vertices **will increase further** as the network grows.

Thus **older** (with smaller t_i) **vertices increase their connectivity at the expense of the younger** (with larger t_i) ones, leading over time to some vertices that are highly connected, a "**rich-get-richer**" phenomenon that can be easily detected in real networks.

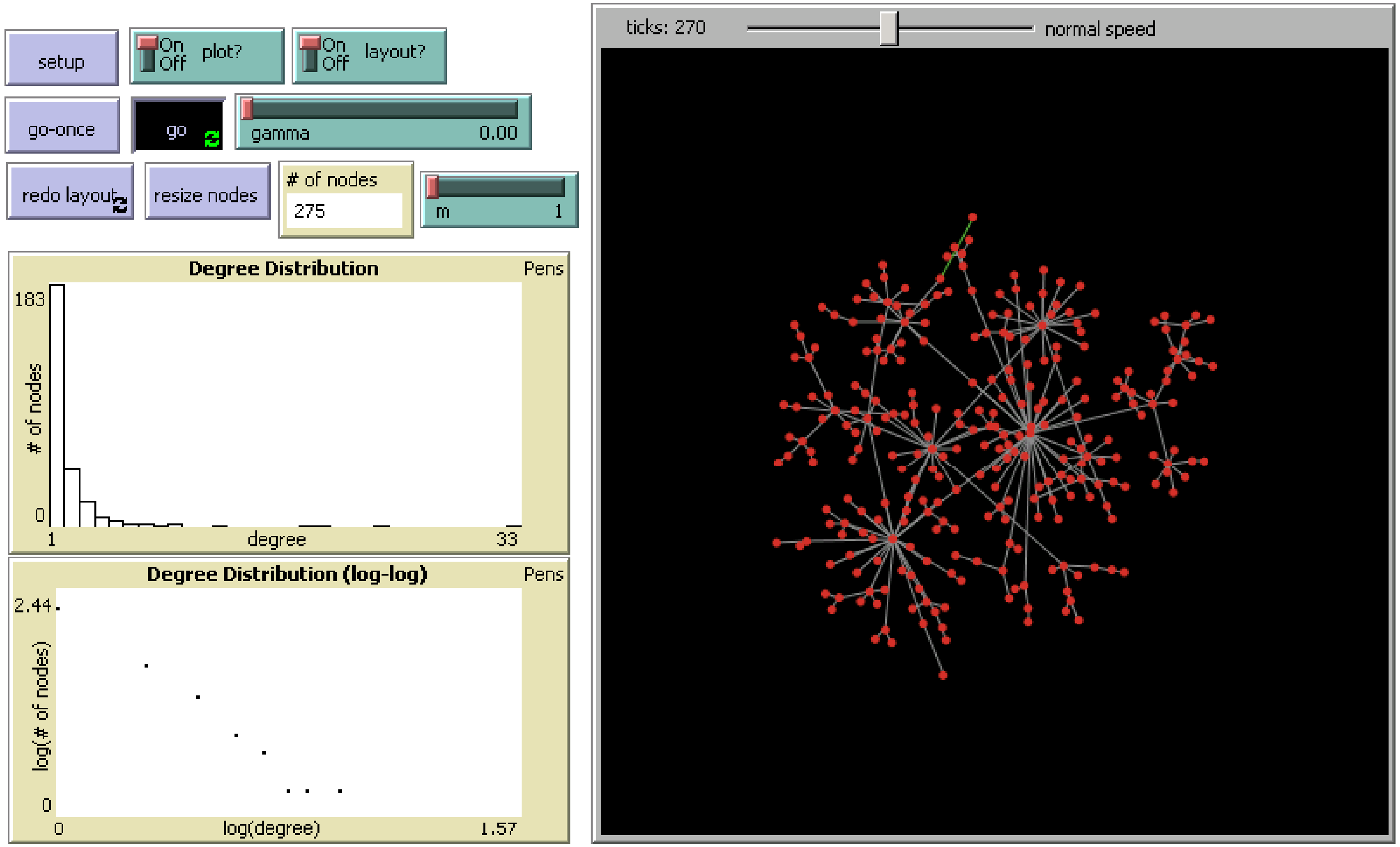
But, [Faloutsos / Leskovec ECML/PKDD 2007]

- all nodes have equal (constant) outdegree (in a directed network)
- one needs complete knowledge of the network (knowing the degrees of all nodes)

Demo – Preferential Attachment

Wilensky, U. (2005). NetLogo Preferential Attachment model. http://ccl.northwestern.edu/netlogo/models/PreferentialAttachment. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL

http://www-personal.umich.edu/~ladamic/NetLogo/index.html



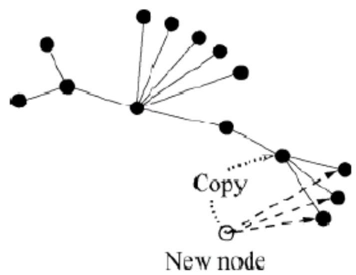
Edge copying model
[Faloutsos / Leskovec ECML/PKDD 2007]

CMU SCS

http://videolectures.net/ecml07_leskovec_mlg/

Edge copying model

- But, preferential attachment does not have communities
- Copying model [Kleinberg et al, 99]:
  - Add a node and choose k the number of edges to add
  - With prob. β select k random vertices and link to them
  - With prob. $1\text{-}\beta$ edges are copied from a randomly chosen node
- Generates power-law degree distributions with exponent $1/(1\text{-}\beta)$
- Generates communities

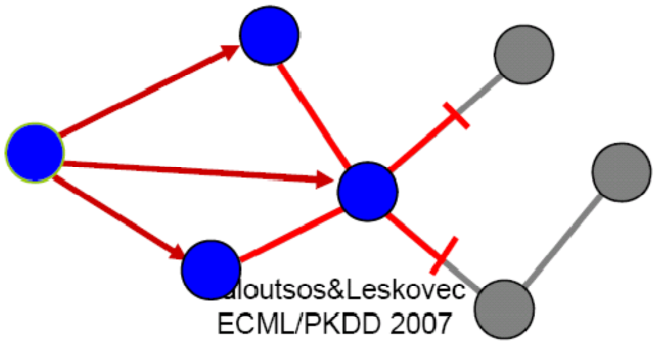


Forest Fire Model
[Faloutsos / Leskovec ECML/PKDD 2007]

CMU SCS

Forest Fire Model

- But, we do not want to have explicit communities
- Want to model graphs that density and have shrinking diameters
- Intuition:
  - How do we meet friends at a party?
  - How do we identify references when writing papers?

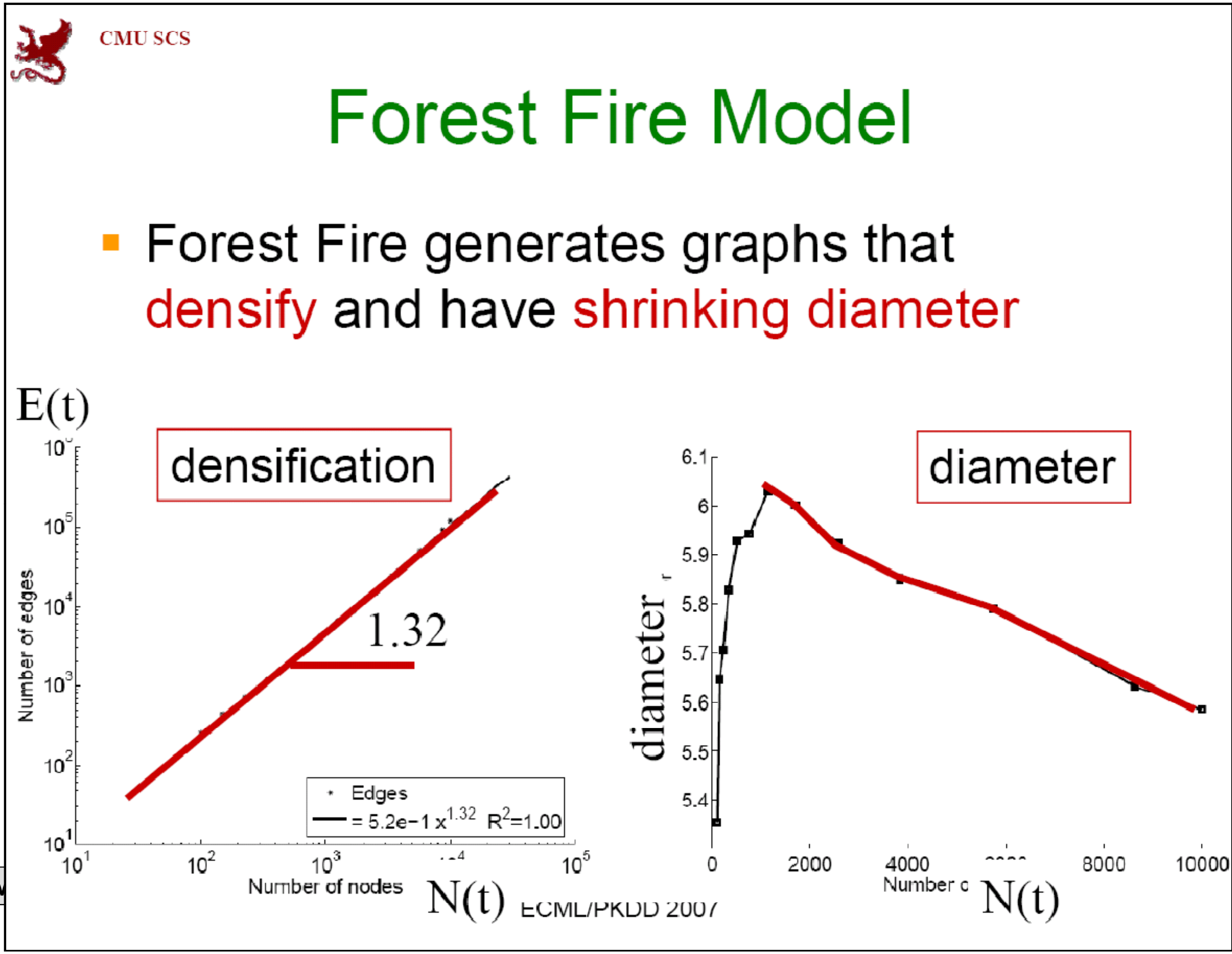


aloutsos&Leskovec ECML/PKDD 2007

Part 1-44

Forest Fire Model

[Faloutsos / Leskovec ECML/PKDD 2007]

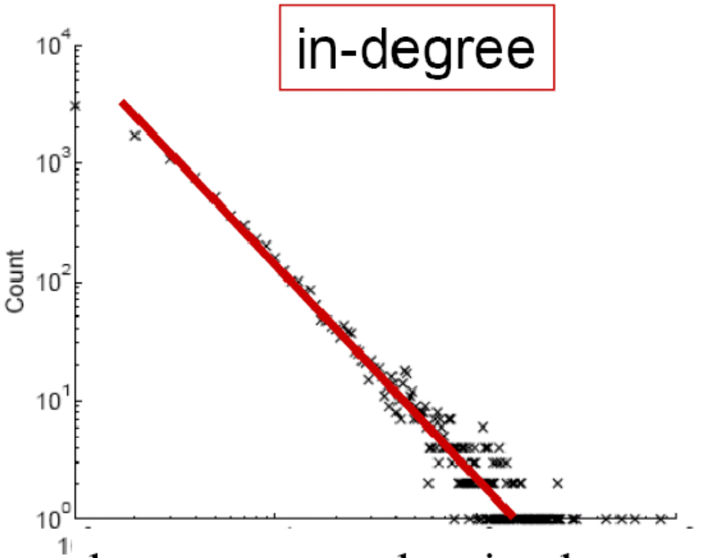


Forest Fire Model
[Faloutsos / Leskovec ECML/PKDD 2007]

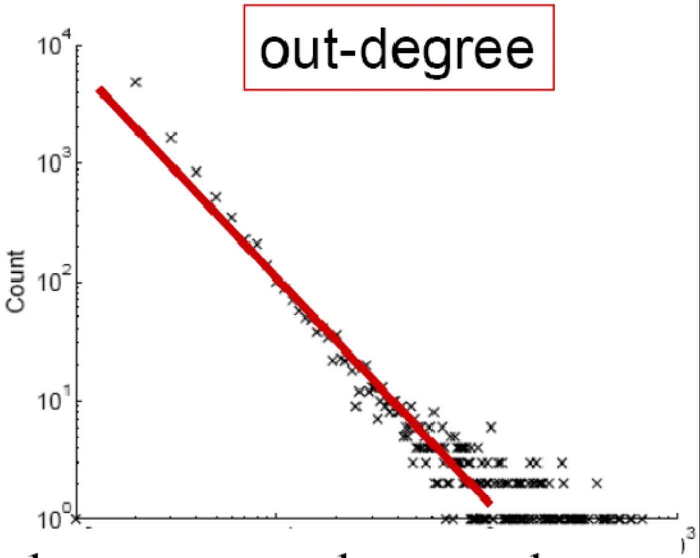
CMU SCS

Forest Fire Model

- Forest Fire also generates graphs with Power-Law degree distribution



log count vs. log in-degree



log count vs. log out-degree

Network Generators: Description and Survey

D. Chakrabarti and C. Faloutsos. Graph mining: Laws, generators, and algorithms. ACM Comput. Surv., 38(1), 2006.

Table II. Taxonomy of Graph Generators

Generator	Graph type						Degree distributions: Power law			
	Undir.	Dir.	Bip.	Self loops	Mult. edges	Geog. info	Plain	Exp. cutoff	Deviation	Exponential
Erdös–Rényi [1960]	√			√	√					√
PLRG [Aiello et al. 2000], PLOD [Palmer and Steffan 2000]	√			√	√		any γ (Eq. 15) (user-defined)			
Exponential cutoff [Newman et al. 2001]	√			√	√		any γ (Eq. 16) (user-defined)	√		
BA [Barabási and Albert 1999]	√						$\gamma = 3$			
Initial attractiveness [Dorogovtsev and Mendes 2003]		√		√	√		$\gamma \in [2, \infty)$ (Eq. 3.2.2)			
AB [Albert and Barabási 2000]	√			√	√		$\gamma \in [2, \infty)$ (Eq. 22)			√
Edge Copying [Kumar et al. 1999], [Kleinberg et al. 1999]		√		√	$\gamma \in (1, \infty)$		√ (Eqs. 23, 24)			
GLP [Bu and Towsley 2002]	√			√	√		$\gamma \in (2, \infty)$ (Eq. 26)			
Accelerated growth [Dorogovtsev and Mendes 2003], [Barabási et al. 2002]	√			√	√		Power-law mixture of $\gamma = 2$ and $\gamma = 3$			
Fitness model [Bianconi and Barabási 2001]	√						$\gamma = 2.255$[1]			
Aiello et al. [2001]		√					$\gamma \in [2, \infty)$ (Eq. 30)			
Pandurangan et al. [2002]		√		√	$\gamma = ?$		√			
Inet-3.0 [Winick and Jamin 2002]	√						$\gamma = ?$[2]	√		
Forest Fire [Leskovec et al. 2005]		√					$\gamma = ?$			
Pennock et al. [2002]	√			√	√		$\gamma \in [2, \infty)$[3]		√	
Small-world [Watts and Strogatz 1998]	√					√				√
Waxman [1988]	√					√		√		
BRITE [Medina et al. 2000]	√					√	$\gamma = ?$			
Yook et al. [2002]	√					√	$\gamma = ?$		√	
Fabrikant et al. [2002]	√					√	$\gamma = ?$			
R-MAT [Chakrabarti et al. 2004]	√	√	√	√	√		$\gamma = ?$		√ (DGX)	

This table shows the graph types and degree distributions that different graph generators can create. The graph type can be undirected, directed, bipartite, allowing self-loops or multigraph (multiple edges possible between nodes). The degree distributions can be power-law (with possible exponential cutoffs, or other deviations such as lognormal/DGX) or exponential decay. If it can generate a power law, the possible range of the exponent γ is provided. Empty cells indicate that the corresponding property does not occur in the corresponding model.

Network Attacks

Informed vs. Random Attacks:

http://www-personal.umich.edu/~ladamic/GUESS/resiliencedegree.html

Network Resilience [Newman 2003]

The resilience of networks with respect to vertex removal and network connectivity.

If vertices are removed from a network, the typical length of paths between pairs of vertices will increase – vertex pairs will be disconnected.

Examples:

1. Deletion of a hub
2. Deletion of a leaf node element

The web is highly resilient against random failure of vertices, but highly vulnerable to deliberate attack on its highest-degree vertices

Network Resilience [Newman 2003]

Delete the node with the highest degree, what happens to the network?

Deleting which nodes introduces a new component?

Example

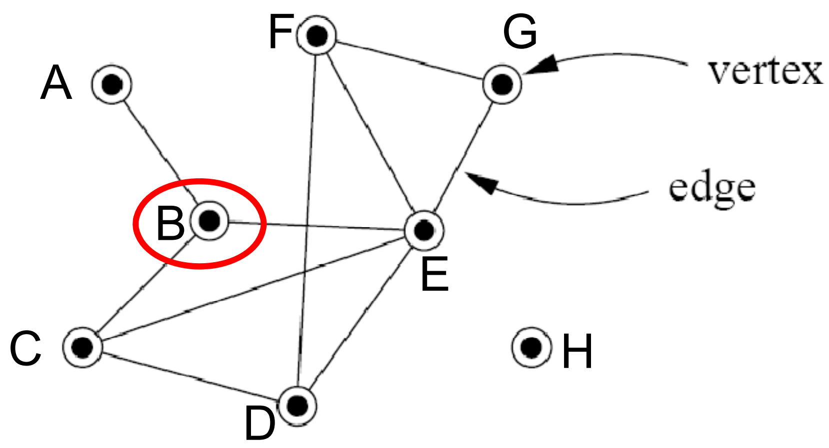


FIG. 1 A small example network with eight vertices and ten edges.

[Newman 2003]

Connectivity: a function of whether a graph remains connected when nodes and/or lines are deleted. [Wassermann 1994]

Network Resilience [Newman 2003]

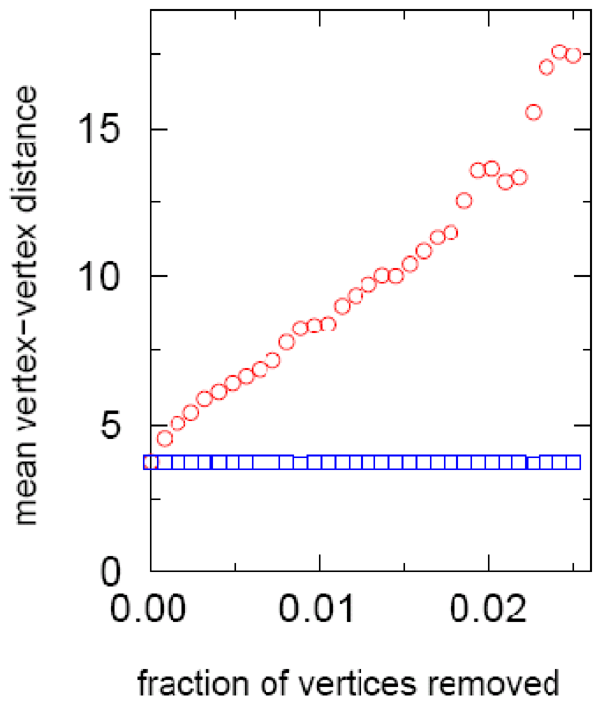


FIG. 7 Mean vertex–vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert *et al.* [15].

Removal of high degree nodes first

Removal of random nodes

Percolation Theory [Newman 2003]

A percolation process is one in which vertices or edges on a graph are randomly designated either "occupied" or "unoccupied".

One of the main motivations for the percolation model when it was first proposed in the 1950s was the modeling of the spread of disease.

Connectivity of the Web [Newman 2003, Broder et al 2000]

What does it need to destroy the connectivity of the web?

According to Broder et al 2000, you need to remove all vertices with a degree greater than five.

Because of the highly skewed degree distribution of the web, the fraction of vertices with degree greater than five is only a small fraction of all vertices.

Percolation Theory [Newman 2003]

Why are we interested in percolation theory in the context of web science?

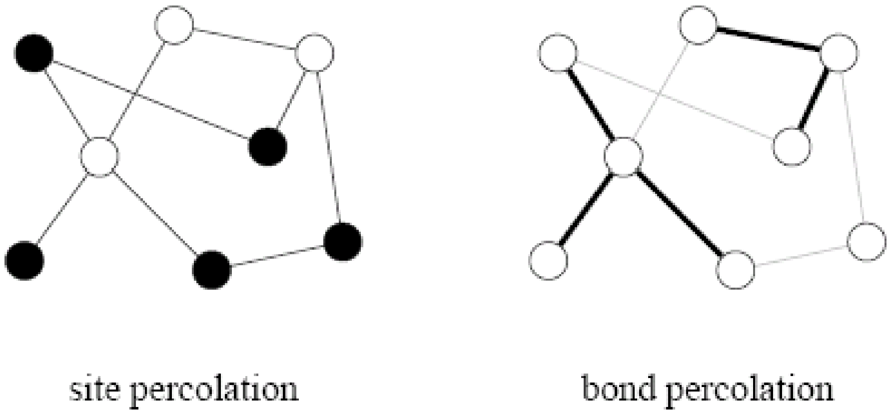


FIG. 13 Site and bond percolation on a network. In site percolation, vertices ("sites" in the physics parlance) are either occupied (solid circles) or unoccupied (open circles) and studies focus on the shape and size of the contiguous clusters of occupied sites, of which there are three in this small example. In bond percolation, it is the edges ("bonds" in physics) that are occupied or not (black or gray lines) and the vertices that are connected together by occupied edges that form the clusters of interest.

Two Fundamental Network Process Distinctions [Newman 2003]

Can you name examples of these processes on the web?

Epidemic processes

- such as influenza, which sweeps through the population rapidly and infects a significant fraction of individuals in a short outbreak (cf. the SIR model)

Endemic processes

- such as measles, which persists within the population at a level roughly constant over time. The disease can persist indefinitely, circulating around the population and never dying out (cf. the SIS model)

The SIR Model
[Watts 2004]

The SIR model of network epidemics

S **Susceptible**
Vulnerable to infection, but not yet been infected

I **Infected**
infected and infectious (can infect others)

R **Removed**
either recovered or ceased to pose a threat

Rules:

- New infections can only occur when an infected individual (an infective) comes into direct contact with a susceptible.
- The susceptible can become infected, with probability p depending on infectiousness of the disease and the characteristics of the susceptible
- Who comes into contact with whom will depend on the populations‘ network structure.

The SIR Model [Watts 2004]

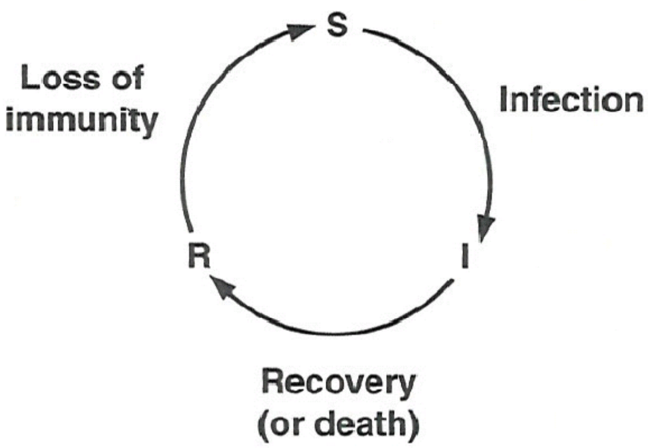


Figure 6.1. The three states of the SIR model. Each member of the population can be susceptible, infected, or removed. Susceptible individuals can become infected by interacting with infectives. Infectives can either recover or die, thus ceasing to take part in the dynamics. If they recover, they might become susceptible again through loss of immunity.

The SIR Model [Watts 2004]

In its simplest version,

- based on purely random interactions
- Rate of infection depends only on the relative population sizes

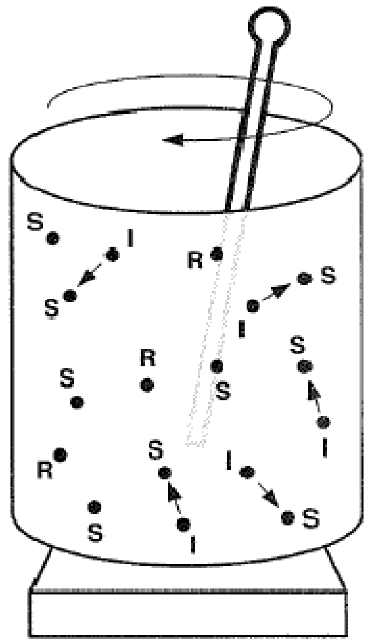


Figure 6.2. In the classical version of the SIR model, interactions are assumed to be purely random. One way to think of random interactions is as individuals being mixed together in a large vat. The main consequence of the random mixing assumption is that interaction probabilities depend only on the relative population sizes, a feature that greatly simplifies analysis.

The SIR Model [Watts 2004]

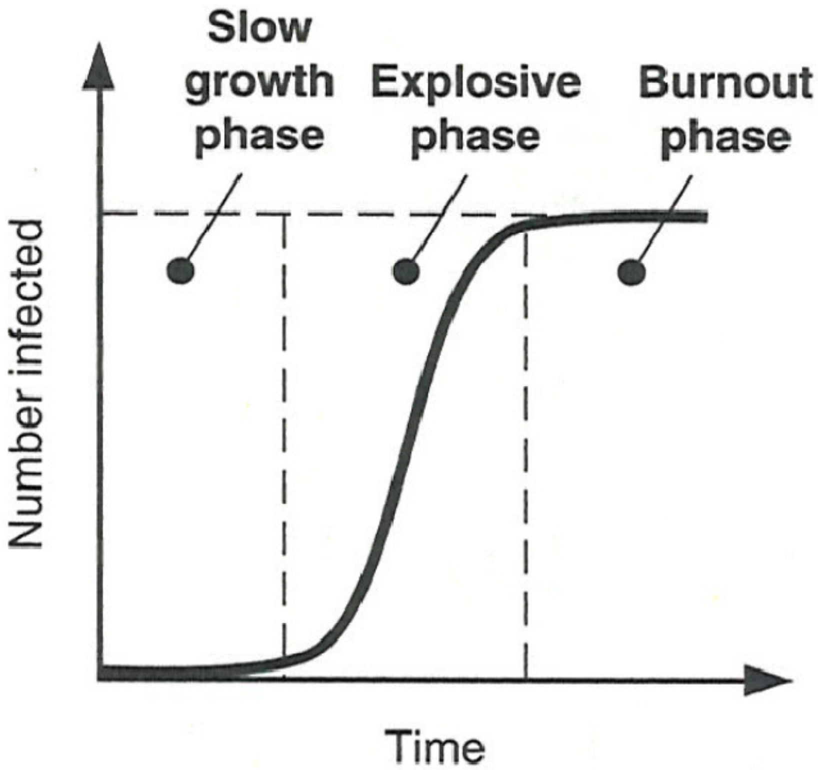


Figure 6.4. Logistic growth, displaying the slow-growth phase, explosive phase, and burnout phase.

Reproduction rate	Low	High	Low

In terms of the SIR model, stopping **an epidemic is roughly equivalent to preventing it from reaching the explosive growth phase**.

This implies focusing **not on the size of the initial outbreak** but on its **rate of growth**.

The SIR Model
[Watts 2004]

Each infection requires the **participation of both an infected and a susceptibel** individual.

The **rate** at which new infections ca be generated depends on the **size of both populations**.

Reproduction rate: the average number of **new infectives generated** by each currently infected.

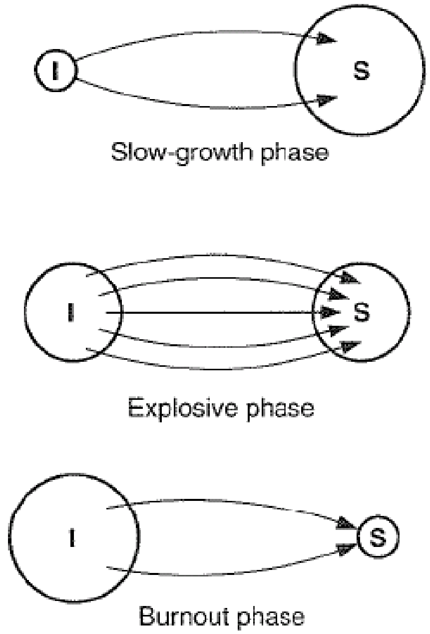


Figure 6.3. In logistic growth, the rate of new infections depends on the size of the susceptible and infected populations. When either population is small (top and bottom diagrams), new infections are rare. But when both populations are intermediate in size (middle diagram), infection rates are maximized.

The SIR Model
[Watts 2004]

Condition for epidemics: reproduction rate >1 (threshold)

Note: That's the same threshold at which a giant component occurs in networks

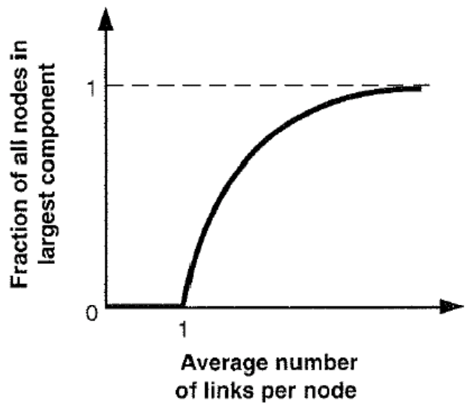


Figure 2.2. Connectivity of a random graph. The fraction of the nodes connected in a single component changes suddenly when the average number of links per node exceeds one.

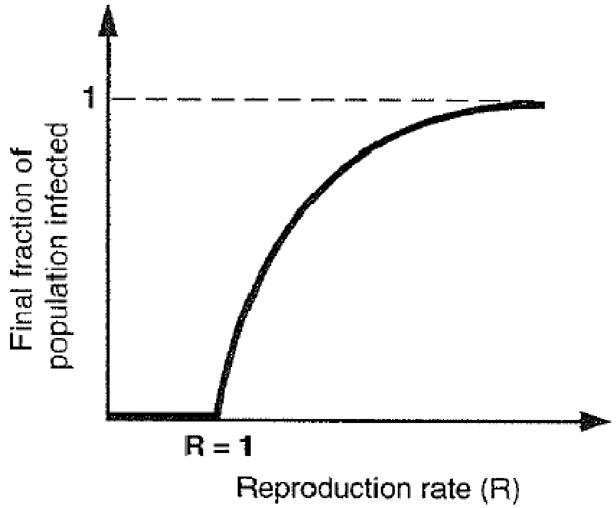


Figure 6.5. Phase transition in the SIR model. When the *reproduction rate* (R) of the disease exceeds one (the epidemic threshold), an epidemic occurs.

SIR simulation: e.g.
http://www.uni-tuebingen.de/modeling/Mod_Pub_Software_SIR_en.html
SI Diffusion in random networks: http://www-personal.umich.edu/~ladamic/NetLogo/ERdiffusion.html
SI Diffusion in scale-free networks: http://www-personal.umich.edu/~ladamic/NetLogo/BADiffusion.html

When Zombies Attack

http://www.wiskundemeisjes.nl/wp-content/uploads/2009/08/zombies.pdf

Chapter 4

WHEN ZOMBIES ATTACK!: MATHEMATICAL MODELLING OF AN OUTBREAK OF ZOMBIE INFECTION

Philip Munz[1]*, Ioan Hudea*[1]*, Joe Imad*[2]*, Robert J. Smith?*[3§]
[1]School of Mathematics and Statistics, Carleton University,
1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada
[2]Department of Mathematics, The University of Ottawa,
585 King Edward Ave, Ottawa ON K1N 6N5, Canada
[2]Department of Mathematics and Faculty of Medicine, The University of Ottawa,
585 King Edward Ave, Ottawa ON K1N 6N5, Canada

Abstract

Zombies are a popular figure in pop culture/entertainment and they are usually portrayed as being brought about through an outbreak or epidemic. Consequently, we model a zombie attack, using biological assumptions based on popular zombie movies. We introduce a basic model for zombie infection, determine equilibria and their stability, and illustrate the outcome with numerical solutions. We then refine the model to introduce a latent period of zombification, whereby humans are infected, but not infectious, before becoming undead. We then modify the model to include the effects of possible quarantine or a cure. Finally, we examine the impact of regular, impulsive reductions in the number of zombies and derive conditions under which eradication can occur. We show that only quick, aggressive attacks can stave off the doomsday scenario: the collapse of society as zombies overtake us all.

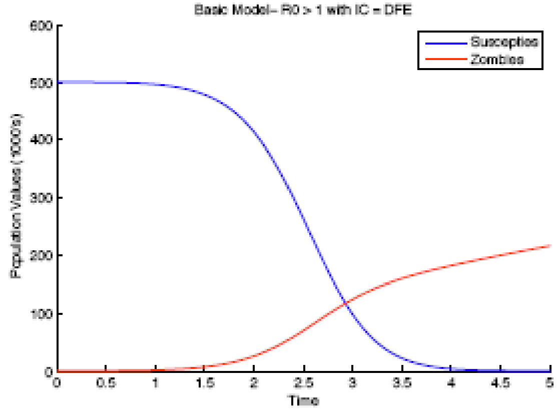


Figure 3. Basic model outbreak scenario. Susceptibles are quickly eradicated and zombies take over, infecting everyone.

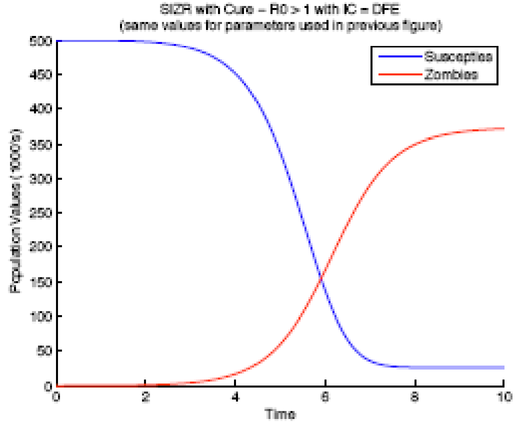


Figure 9. The model with treatment, using the same parameter values as the basic model.

Applications of Graph Generators and Growth Models [Leskovec 2006]

Recapitulation:

- „ What if“ scenarios
- Forecasting future parameters of computer and social networks
- Anomaly detection
- Graph sampling algorithms
- Realistic graph generators

Examples:

- „Invites“ to join GMail
- „Invites“ to buy Chumby
- „Invites“ to join Joost
- Vaccination strategies for epidemics

Any questions?

See you in two weeks!