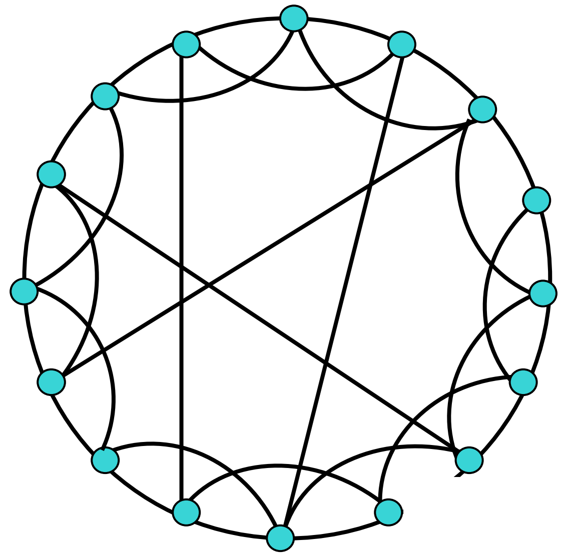
Small World Networks

Franco Zambonelli

February 2010



Outline

- Part 1: Motivations
  - The Small World Phenomena
  - Facts & Examples
- Part 2: Modeling Small Worlds
  - Random Networks
  - Lattice Networks
  - Small World Networks
- Part 3: Properties of Small World Networks
  - Percolation and Epidemics
  - Implications for Distributed Systems
- Conclusions and Open Issues

Part 1

- Motivations

Let’s Start with Social Networks

- We live in a connected social world
  - We have friends and acquaintances
  - We continuously meet new people
  - But how are we connected to the rest of the world
- We have some relationships with other persons – thus we are the nodes of a “social network”
  - Which structure such “social network” has
  - And what properties does it have?

“Hey, it’s a Small World!”

- How often has is happened to meet a new friend
  - Coming form a different neighborhood
  - Coming from a different town
- And after some talking discovering with surprise you have a common acquaintance?
  - “Ah! You know Peter too!”
  - “It’s a small world after all!”
- Is this just a chance of there is something more scientific behind that?

The Milgram Experiment (1967)

- From Harvard University, he sent out to randomly chosen person in the US a letter
- Each letter had the goal of eventually reaching a target person (typically a Milgram's friend), and it prescribe the receiver to:
  - If you know the target on a personal basis, send the letter directly to him/her
  - If you do not know the target on a personal basis, re-mail the letter to a personal acquaintance who is more likely than you to know the target person
  - Sign your name on the letter, so that I (Milgram) can keep track of the progresses to destination
- Has any of the letters eventually reached the target? How long could that have taken?
- Would you like to try it by yourself?
  - http://smallworld.columbia.edu/

Results of Milgram's Experiment

- Surprisingly
  - 42 out 160 letter made it
  - With an average of intermediate persons having received the letter of 5!
- So, USA social network is indeed a "small world"!
  - Six degrees of separation on the average between any two persons in the USA (more recent studies say 5)
  - E.g., I know who knows the Rhode Island governor who very likely knows Condoleeza Rice, who knows president Bush
  - Since very likely anyone in the world knows at least one USA person, the worldwide degree of separation is 6
- John Guare: "Six Degree of Separation", 1991
  - "Six degrees of separation. Between us and everybody else in this planet. The president of the United States. A gondolier in Venice…It's not just the big names. It's anyone. A native in the rain forest. A Tierra del Fuegan., An Eskimo. I am bound to everyone in this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds"
  - 1993 movie with Will Smith

Kevin Bacon

Apollo 13

- A great actor whose talent is being only recently recognized
  - But he’s on the screen since a long time…
  - From “Footloose” to “The Woodsman”
- Also known for being the personification of the “small world” phenomena in the actors’ network
- The “Oracle of Bacon”
  - http://www.cs.virginia.edu/oracle/

The "Bacon Distance"

- Think of an actor X
  - If this actor has made a movie with Kevin Bacon, then its Bacon Distance is 1
  - If this actor has made a movie with actor Y, which has in turn made a movie with Kevin Bacon, then its Bacon Distance is 2
  - Etc. etc.
- Examples:
  - Marcello Mastroianni: Bacon Distance 2
    - Marcello Mastroianni was in Poppies Are Also Flowers (1966) with Eli Wallach
      Eli Wallach was in Mystic River (2003) with Kevin Bacon
  - Brad Pitt: Bacon Distance 1
    - Brad Pitt was in Sleepers (1996) with Kevin Bacon
  - Elvis Presley: Bacon Distance 2
    - Elvis Presley was in Live a Little, Love a Little (1968) with John (I) Wheeler
      John (I) Wheeler was in Apollo 13 (1995) with Kevin Bacon

The Hollywood Small World

- Have a general look at Kevin Bacon numbers…
  - Global number of actors reachable within at specific Bacon Distances
  - Over a database of half a million actors of all ages and nations…
- It's a small world!!!
  - Average degree of separation around 3

Bacon Number	# of People
0	1
1	1802
2	148661
3	421696
4	102759
5	7777
6	940
7	95
8	13

The Web Small World

- Hey, weren't we talking about "social" network?
  - And the Web indeed is
    - Link are added to pages based on "social" relationships between pages holders!
    - The structure of Web links reflects indeed a social structure
- Small World phenomena in the Web:
  - The average "Web distance" (number of clicks to reach any page from anywhere) is less than 19
  - Over a number of more than a billion (1.000.000.000) documents!!!

Other Examples of Small Worlds

- The Internet Topology (routers)
  - Average degree of separation 5
  - For systems of 200.000 nodes
- The network of airlines
  - Average degree of separation between any two airports in the works around 3.5
- And more…
  - The network of industrial collaborations
  - The network of scientific collaborations
  - Etc.

- How can this phenomena emerge?

Part 2

- Modeling Small World Networks

How Can We Model Social Networks?

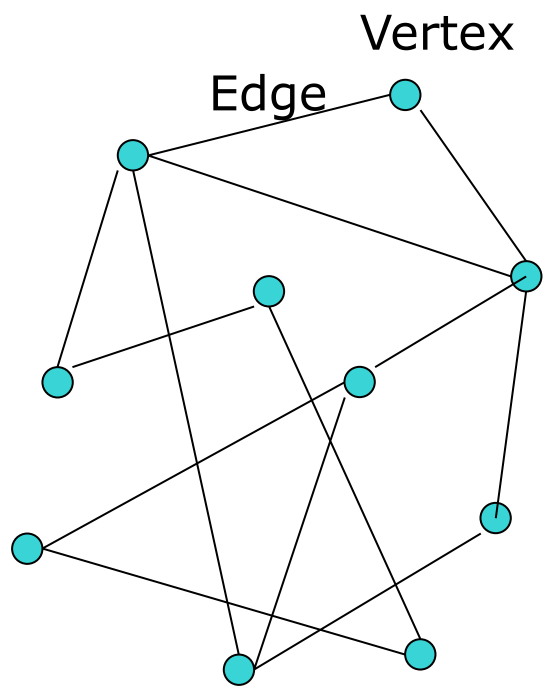
- It gets complicated…
- Relations are "fuzzy"
  - How can you really say you know a person?
- Relations are "asymmetric"
  - I may know you, you may not remember me
- Relations are not "metric"
  - They do not obey the basic trangulation rule $d(X,Z) <= d(X,Y)+d(Y,Z)$
  - If I am Y, I may know well X and Z, where X and Z may not know each other…
- So, we have to do some bold assumptions

Modeling Social Networks as Graphs

- Assume the components/nodes of the social network are vertices of a graph
  - Simple geometrical "points"
- Assume that any acquaintance relation between two vertices is simply an undirected unweighted edge between the vertices
  - Symmetric relations
  - A single edge between two vertices
  - No fuzziness, a relation either exists or does not exist
- Transitivity of relations
  - So, distance rules are respected
- The graph must be necessarily "sparse"
  - Much less edges than possible…
  - We do not know "everybody", but only a small fraction of the world..
- As you will see, we will able in any case to understand a lot about social network…
  - From now on, I will use both "graph" and "networks" as synonyms

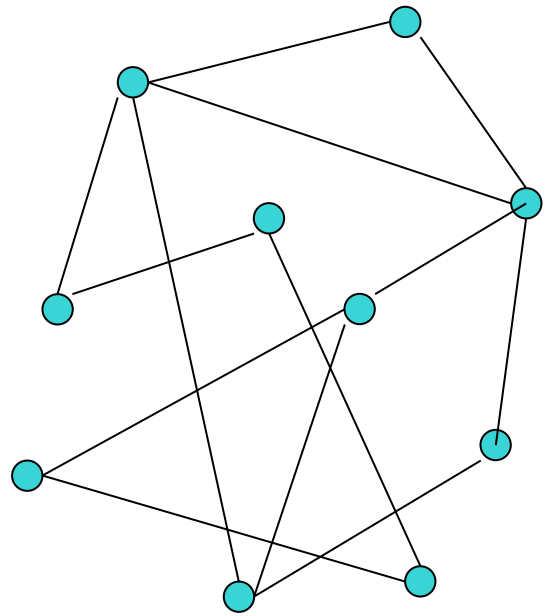
Basic of Graph Modeling (1)

- Graph *G* as
  - A vertex set *V(G)*
  - The nodes of the network
  - An edge list *E(G)*
  - The relations between vertices
- Vertices *v* and *w* are said "connected" if
  - there is an edge in the edge list joining *v* and *w*
- For now we always assume that a graph is fully connected
  - There are not isolated nodes or clusters
  - Any vertex can be reached by any other vertex



Basic of Graph Modeling (2)

- The **order** *n* of a graph is the number of its vertices/nodes
- The **size** *M* of a size is the number of edges
  - Sorry I always get confused and call “size” the order
  - $M=n(n-1)/2$ for a fully connected graph
  - $M<<n(n-1)/2$ for a sparse graph
- The **average degree** *k* of a graph is the average number of edges on a vertices
  - $M=nk/2$
  - **k-regular** graph if all nodes have the same *k*



Graph Length Measures

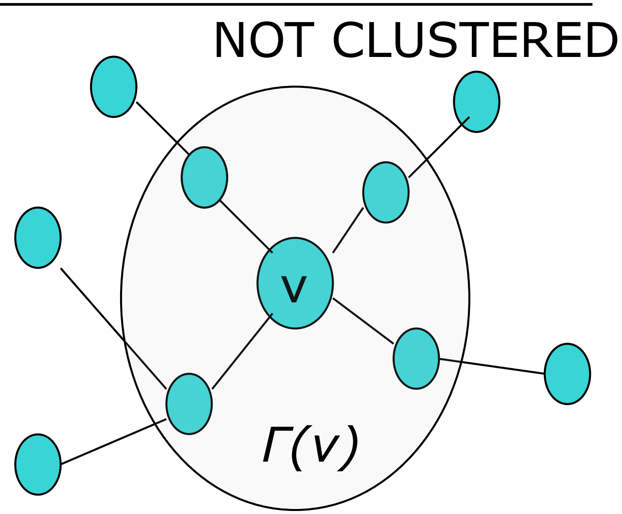
- **Distance** on a graph
  - *D(i,j)*
  - The number of edges to cross to reach node j from node i.
  - Via the shortest path!
- **Characteristic Path Length** *L(G) or simply l"*
  - The median of the means of the shortest path lengths connecting each vertex $v \in V(G)$ to all other vertices
  - Calculate d(i,j) $\forall j \in V(G)$ and find average of D. Do this for $\forall i$. Then define L(G) as the median of D
  - Since this is impossible to calculate exactly for large graphs, it is often calculated via statistical sampling
  - This is clearly the average "degree of separation"
    - A small world has a small L

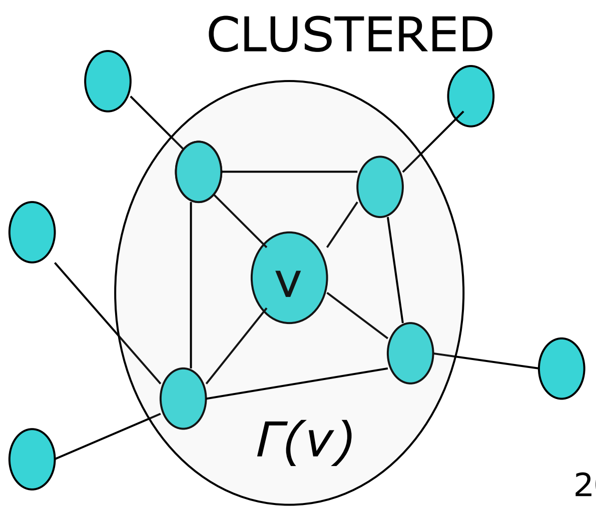
Neighborhood

- The neighborhood $\Gamma(v)$ of a vertex v
  - Is the subgraph S consisting of all the vertices adjacent to v, v excluded
  - Let us indicate a $|\Gamma(v)|$ the number of vertices of $\Gamma(v)$
- The neighborhood $\Gamma(S)$ of a subgraph S
- Is the subgraph that consists of all the vertices adjacent to any of the vertices of S, S *excluded*
  - $S=\Gamma(v)$, $\Gamma(S)= \Gamma(\Gamma(v))= \Gamma^2(v)$
  - $\Gamma^i(v)$ is the ith neighborhood of v
- Distribution sequence Λ
  - $\Lambda_i(v)=\sum_{0 \rightarrow i}|\Gamma(v)|$
  - This counts all the nodes that can be reached from v at a specific distance
  - In a small world, the distribution sequence grows very fast

Clustering (1)

- The clustering γ_v of a vertex v
  - Measures to extent to which the vertices adjacent to j are also adjacent to each other
  - i.e., measure the amount of edges in *Γ (v)*





Clustering

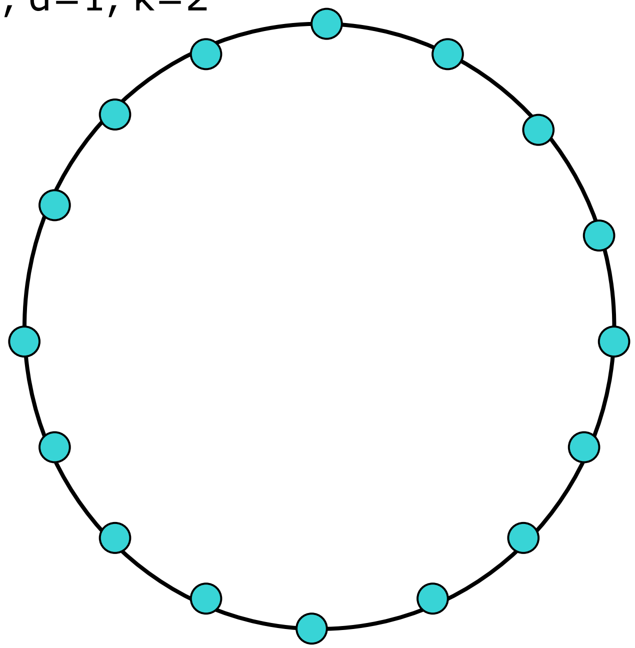
- The **clustering of a vertex** *v* γ_v (or simply C_v) is calculated as
  - $C_v = \gamma_v = |E(\Gamma(v))| / (k_v\ 2)$
  - That is: the number of edges in the neighborhood of v divided by the number of possible edges that one can draw in that neghborhood
- The clustering of a graph γ (or simply *C*) is calculated as the average of γ_v for all *v*
- Most real world social networks, are typically highly clustered
  - E.g., I know my best friends and my best friends know each other

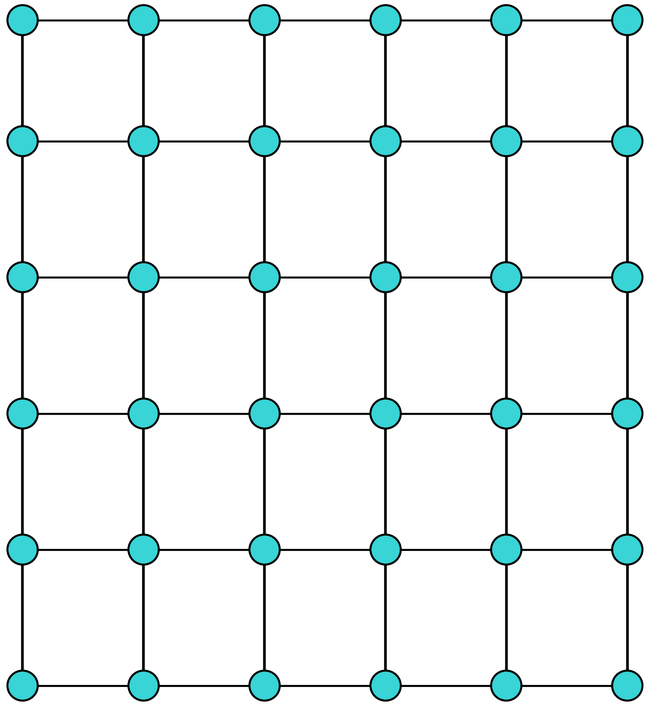
Classes of Networks

- Let's start analyzing different classes of networks
  - And see how and to what extent they exhibit "small world" characteristics
  - And to what extent they are of use in modeling social and technological networks
- Let's start with two classes at the opposite extremes
  - Lattice networks (as in cellular automata)
  - Random networks

d-Lattice Networks

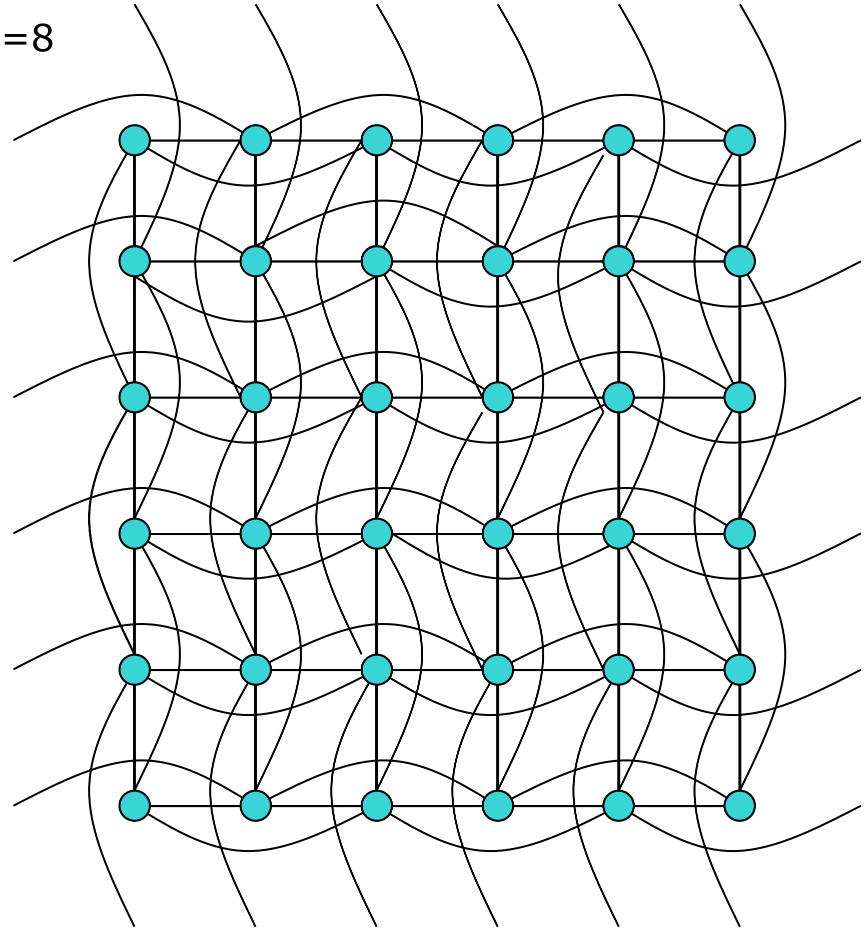
- d-Lattice networks are regular d-dimensional k-regular grids of vertices
  - 1-d, k=2 it's a ring
  - 2-d, k=4, it's a mesh



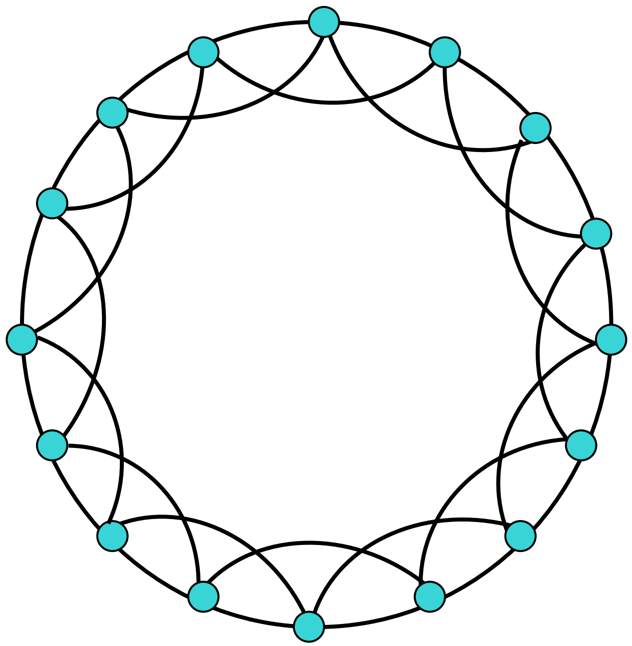


Other Examples of d-Lattices

n=36, d=2, k=8



n=16, d=1, k=4

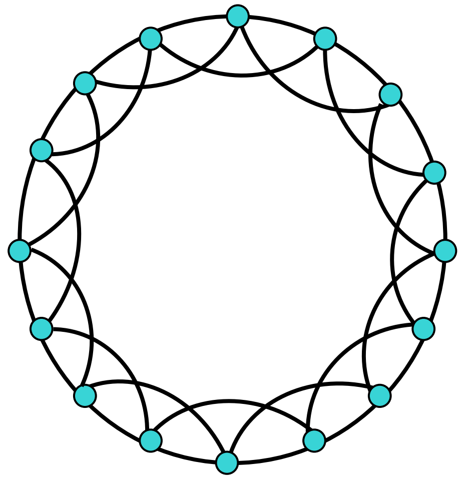


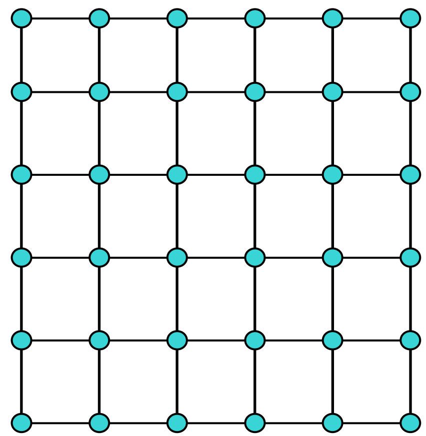
Properties of d-Lattices

- For d=1
  - $L \propto n \rightarrow L=(n(n+k-2)) / (2k(n-1))$
  - $|\Gamma^i(v)|=k$ for any v
  - $\gamma = (3k-6)/(4k-4)$ *e.g., = 0,5 for k=4*
- *For d=2*
  - $L \propto \sqrt{n}$
  - $|\Gamma^i(v)|=k$
  - $\gamma = (4k-16)/(16k-16)$ *e.g. = 0 for k=4*

Some Actual Data

- *d=1, n=1000, k=4*
  - *L =250*
  - γ *= 0,5 for k=4*
  - Good clustering, but not a small world!!!
- *d=2, n=10000, k=4*
  - *L = 50*
  - γ *= 0*
  - Not a small world and not clustered
- Lattice networks are not small world!!!
  - Not realistic representations of modern networks

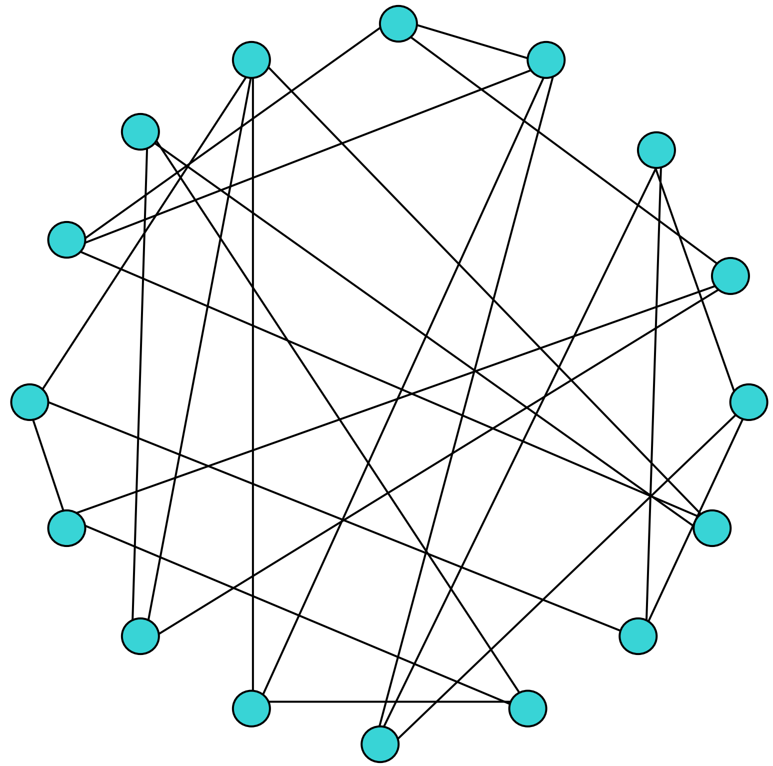




Random Networks

- Very simple to build
  - Given a set *n* of vertices
  - draw M edges each of which connect two randomly chosen vertices
- For a k-regular random networks
  - For each *i* of the *n* vertices
  - Draw k edges connecting i with k other randomly chosen vertices
  - Avoiding duplicate edges and “self” edges
- Note that the concept of dimension lose meaning

n=16, k=3



Properties of Random Networks

- For large *n*
  - Each node has *k* neighbors
  - Each connecting it to other *k* neighbors, for a total of k^2 nodes
  - And so on…
  - *In general* $\Lambda_i(v)=\sum_{i=0,n}|\Gamma(v)|\approx k^i$ for any *v*
    - Please note that for large n, the probability of cycles reducing the above estimate is very small for small *i*, while such cycles are intrinsic in lattices
    - In other words, the neighbours of a node v typically have other neighbours which, in turns, are unlikely to be neighbors with each other, in fact
- The clustering factor is low and is about
  - $\gamma = k/n$

Length of Random Networks

- Given a random network of order n
  - Since $\Lambda_i(v)=\sum_{i=0,n}|\Gamma(v)|\approx k^i$
  - There must exists a number L such that $n\approx k^L$
  - That is, a number L such that, from any v, and going at distance L, I can reach all the nodes of the network
- Then, such L will approximate the average lenght of the network
  - $n\approx k^L \rightarrow L = \log(n)/\log(k)$
- The "degree of separation" in a random network grown only logaritmically!!!
- Random networks are indeed "small world!!"

Some Actual Data

- The Hollywood network
  - n=500000; k=60 (The Hollywood network)
  - L=log(500000)/log(60)=3,2
  - Matches actual data!
- The Web network
  - n=200000000; k=8
  - L=11
- Even shorter than in actual data!

However…

- Are random networks a realistic model?
  - Random networks are not clustered at all!
  - This is why they achieve very small degrees of separations!
- We know well social networks are strongly clustered
  - We know our friends and our 90% of our friends know each other
  - Web pages of correlated information strongly link to each others in clustered data
  - The network of actors is strongly clustered
    - E.g., dramatic actors meet often in movies, while they seldom meet comedians…

- We can see this from real data, comparing a random network model with the real statistical data….
  - Making it clear that social networks are different…

Random Networks vs. Real Networks (1)

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Small world

R. Albert, A. Barabási, Reviews of Modern Physics **74**, 47 (2002).

Random Networks vs. Real Networks (2)

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Clustered!

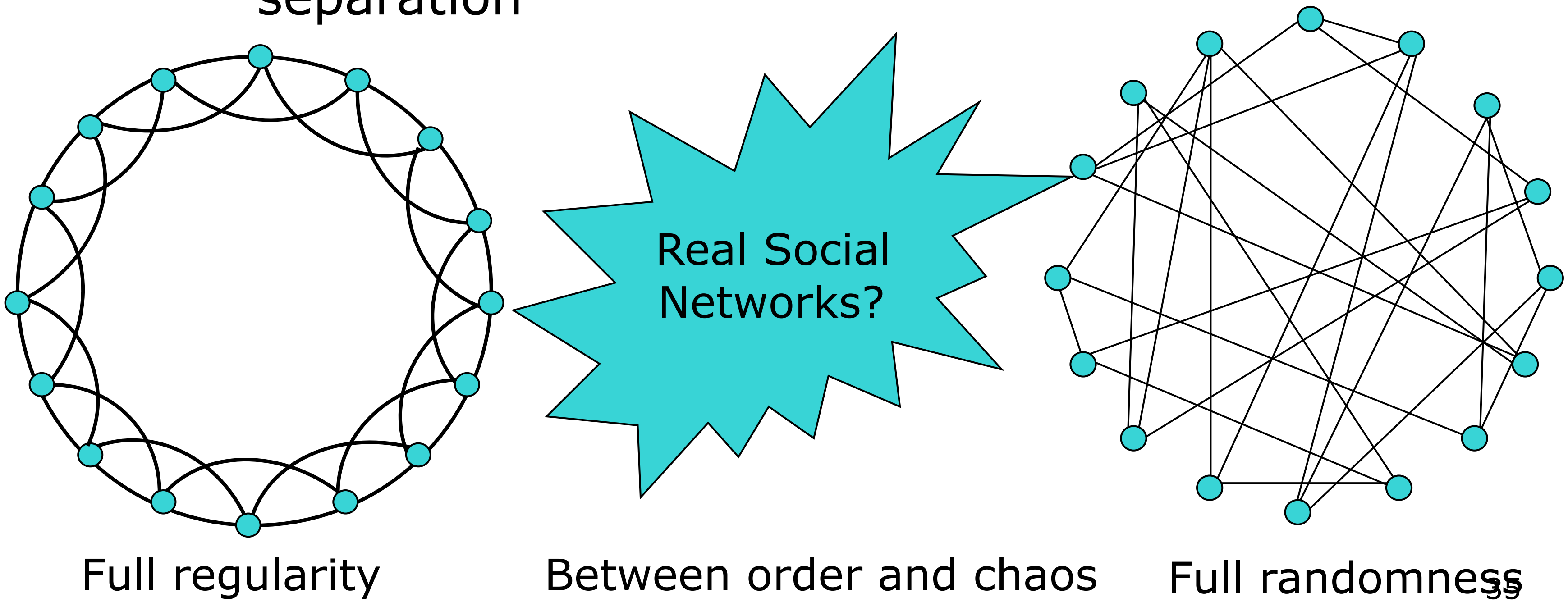
R. Albert, A. Barabási, Reviews of Modern Physics **74**, 47 (2002).

Towards Realistic Small World Models

- The key consideration
  - Real social network are somewhere in between the “order” of lattices and the “chaos” of random network, In fact
- We know well a limited number of persons that also know each other
  - Defining connected clusters, as in d-lattices with reasonably high k
  - And thus with **reasonably high clustering factors**
- At the same time, we also know some people here and there, far from our usual group of friends
  - Thus we have some way of escaping far from our usual group of friends
  - Via acquaintance that are like random edges in the network..

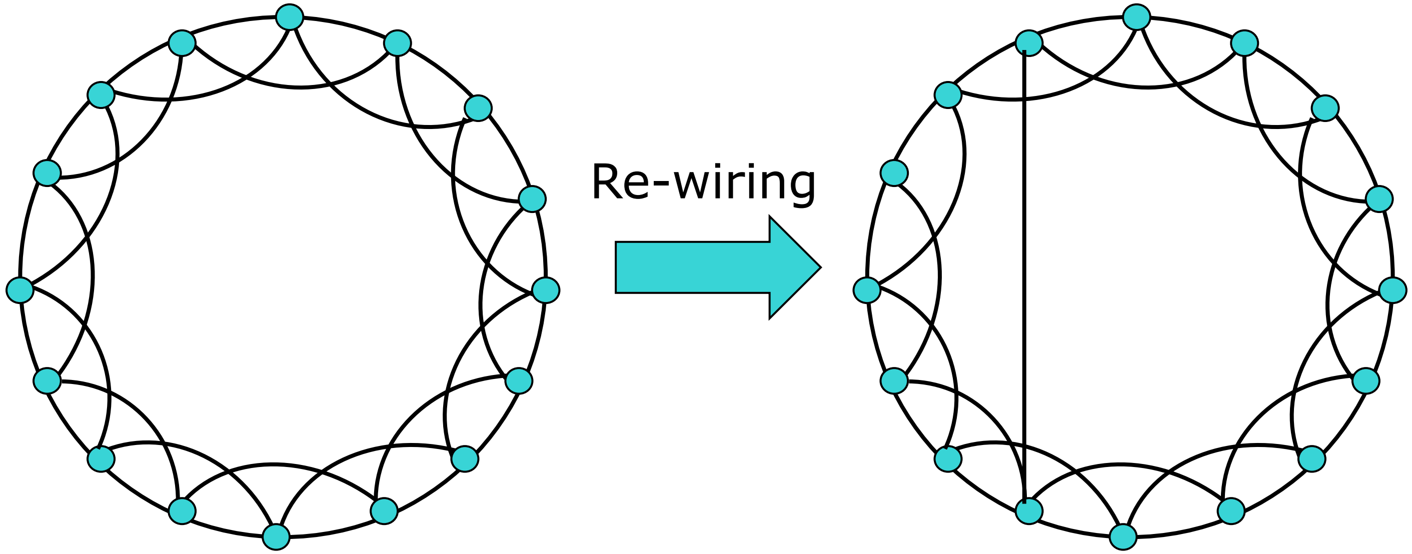
Networks at the Edges of Chaos

- The key idea (Watts and Strogatz, 1999)
- Social networks must thus be
  - Regular enough to promote clustering
  - Chaotic enough to promote small degrees of separation

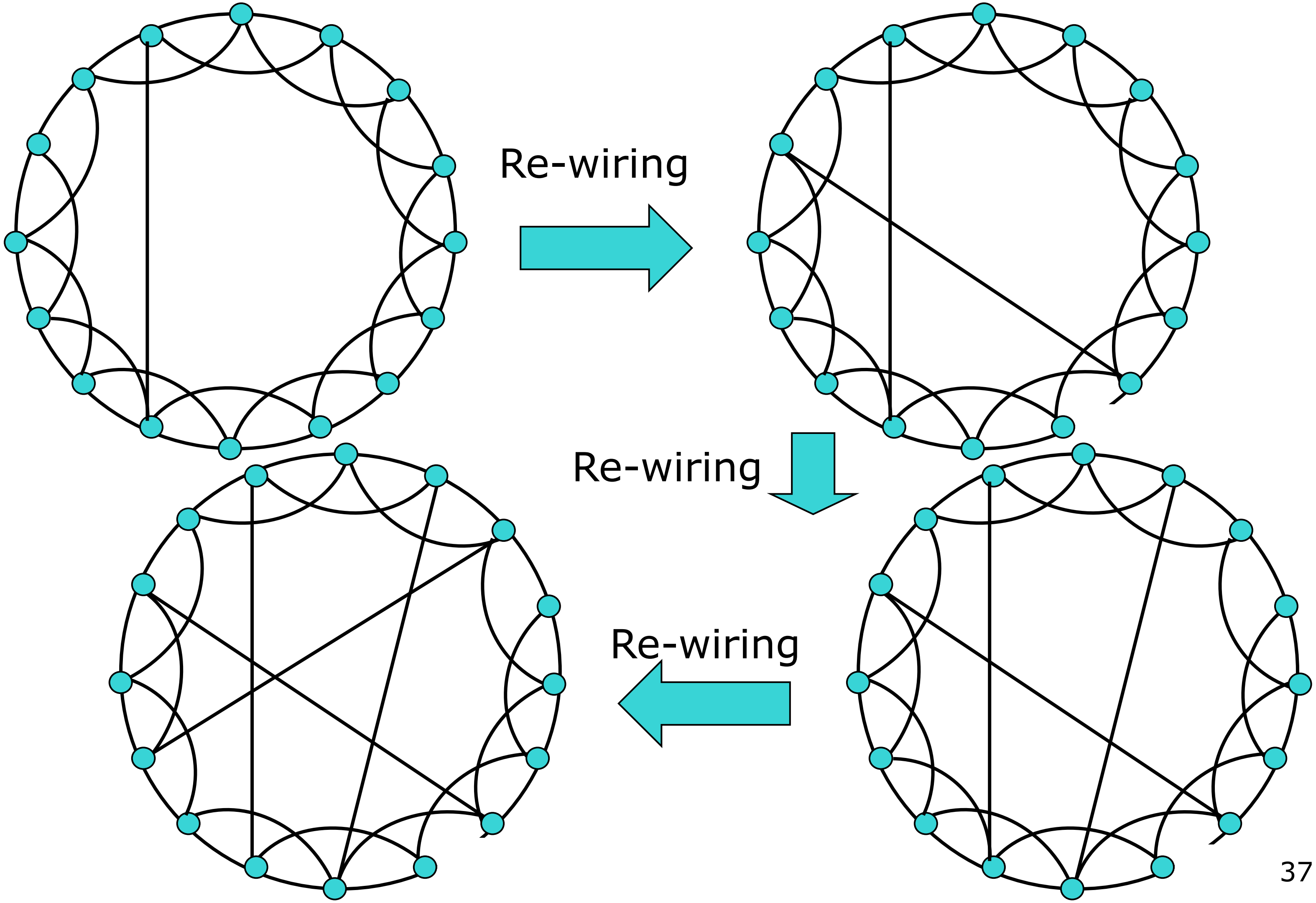


The Watts-Strogatz Model

- Let’s start with a regular lattice
  - And start re-wiring one of the edges at random
  - Continue re-wiring edges one by one
  - By continuing this process, the regular lattice gets progressively transformed into a random network



Let's Re-wire!



So What?

- Re-wired network are between order and chaos
- For limited re-wiring, they preserve a reasonable regularity
  - And thus a reasonable clustering
- Still, the exhibit “short-cuts”, i.e., edges that connect parts of the network that would have been far away from each other
  - This provides for shortening the length of the network

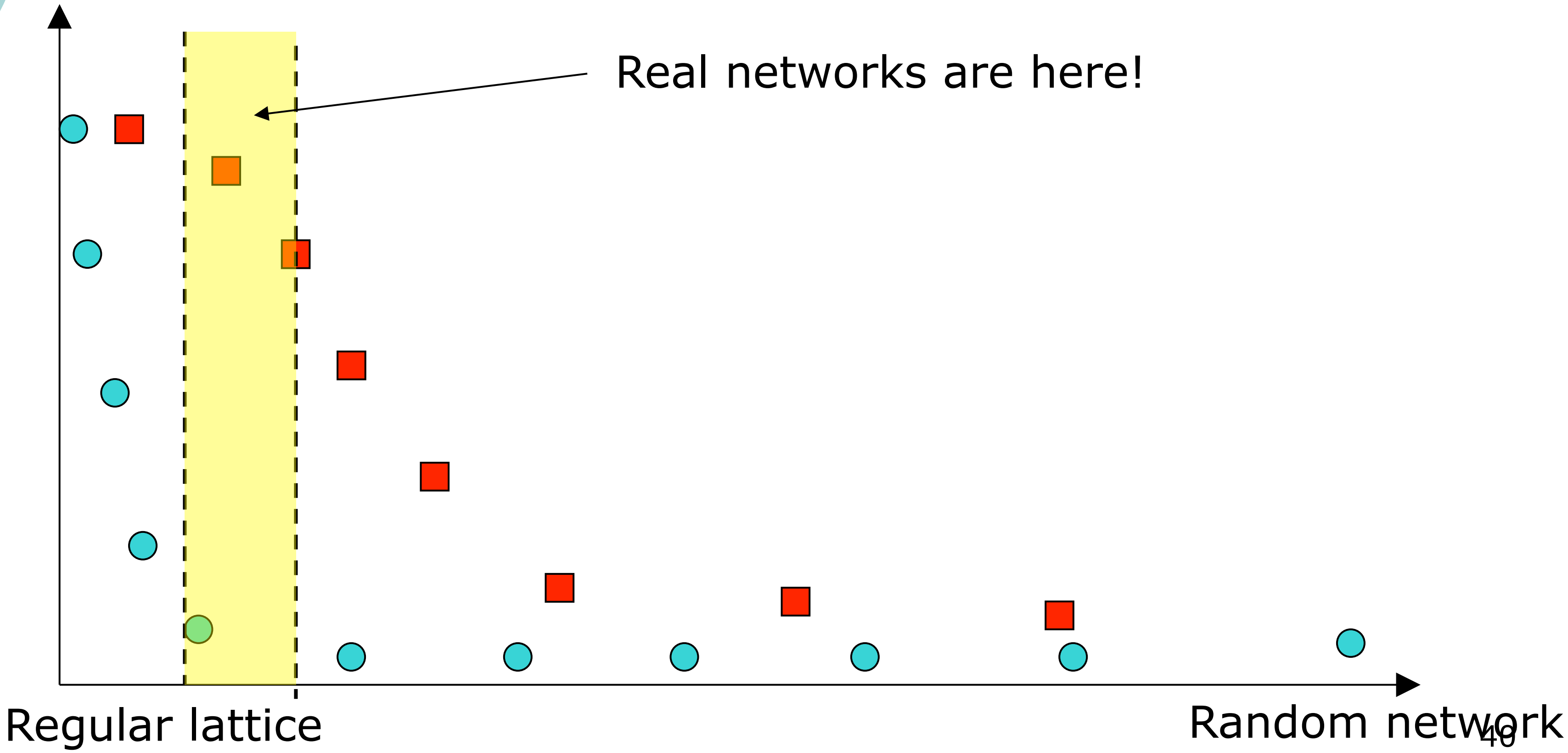
- But how much re-wiring is necessary?

Experimental Results (1)

- Doing exact calculation is impossible
- But experiments shows that:
  - A very very limited amount of re-wiring is enough to dramatically shorten the length of the network
  - To make is as short as a random network
- At the same time
  - The clustering of the network start decreasing later, for a higher degree of re-wiring
  - Thus
- There exists a moderate regime of re-wiring "between order and chaos" for which the network
  - Exhibit "small world behavior"
  - Is still clustered!
- The same as real-world social networks does!

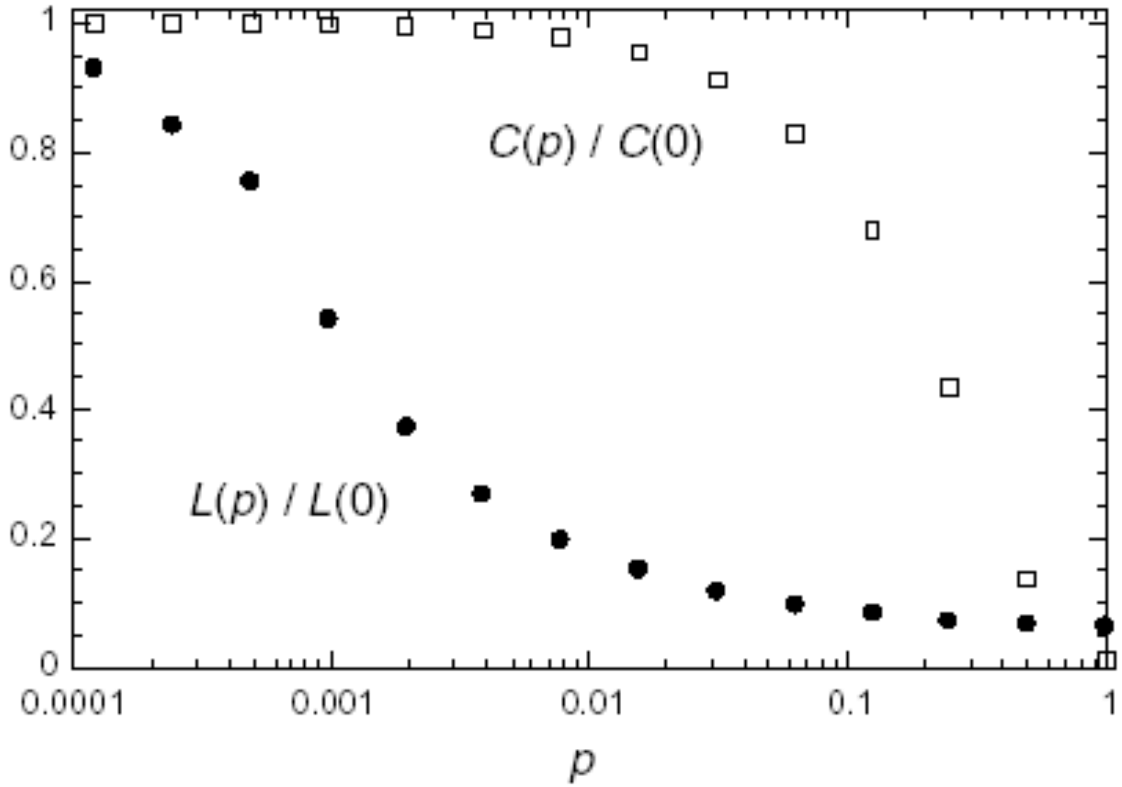
Experimental Results (2)

- Here's how performance graphics could look like…
  - The length diminishes immediately
  - The clustering a bit later…



Experimental Results (3)

- Here's the original graphics of Duncan and Watts, published on "Nature" (length *L* and clustering *C* normalized)



Getting Back to Real Networks (1)

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Small world

R. Albert, A. Barabási, Reviews of Modern Physics **74**, 47 (2002).

As Random Networks!

Getting Back to Real Networks

Network	Size	$\langle k \rangle$	ℓ	ℓ_{rand}	C	C_{rand}	Reference	Nr.
WWW, site level, undir.	153 127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015–6209	3.52–4.11	3.7–3.76	6.36–6.18	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001	2
Movie actors	225 226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52 909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE co-authorship	1 520 251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56 627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11 994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70 975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabási *et al.*, 2001	8
Neurosci. co-authorship	209 293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabási *et al.*, 2001	9
E. coli, substrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Solé, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Solé, 2000	13
Words, co-occurrence	460.902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Solé, 2001	14
Words, synonyms	22 311	13.48	4.5	3.84	0.7	0.0006	Yook *et al.*, 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C. Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Clustered!

R. Albert, A. Barabási, Reviews of Modern Physics **74**, 47 (2002).

Unlike Random Networks

Summarizing

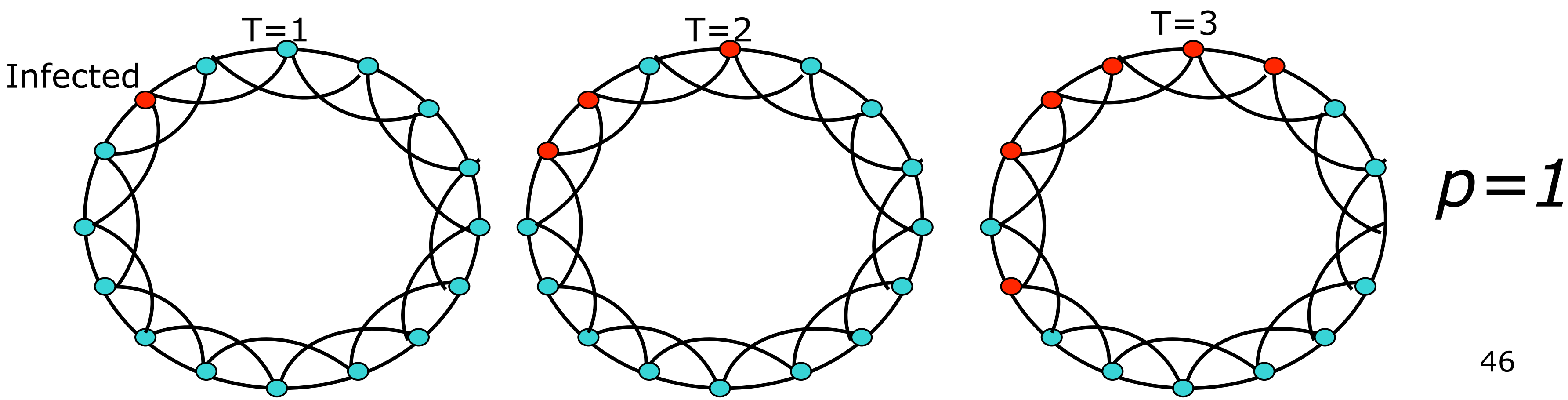
- May real-world networks exhibit the “small world” phenomena
  - Social networks
  - Technological networks
  - Biological networks
- This emerges because these networks have
  - Clustering and “Short-Cuts”
  - Getting the best from both regular and random networks!
  - “At the Edges of Chaos”
- And this may have dramatic impact on the dynamics of the processes that take place on such networks!
  - As we analyze later on…

Part 3

- Properties and Dynamics of Small World Networks

The Spread of Infectious Diseases

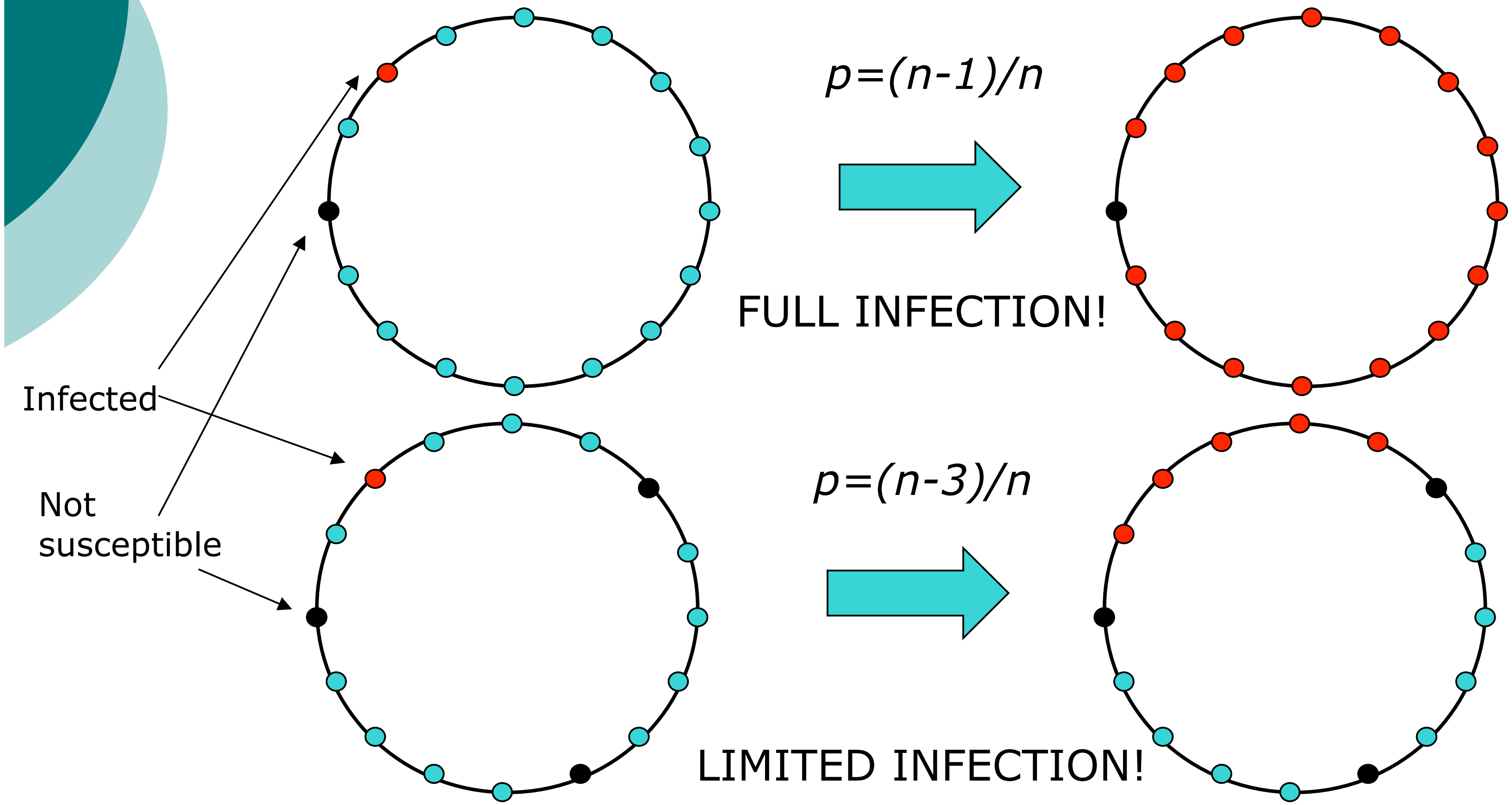
- Let us assume that a single individual, in a social network, is initially infected
- And that it has a probability ***0≤p≤1*** to infect its neighbors (due to the presence of non-susceptible individuals that **do not contract** and **do not further re-propagate** the infection)
  - For *p=0*, the infection do not spread
  - For *p=1*, the infection spread across the whole network, if the network is fully connected, in the fastest way
  - What happens when *0<p<1* ??? (the most realistic case)



Percolation

- **Percolation**: the process by which something (a fluid, a particle, a disease) diffuse across a medium (a fluid, a labyrinth, a network)
- **Percolation threshold**: the critical value of a parameter over which the diffusion process can complete
  - i.e., can diffuse over the (nerarly) whole network
  - It is a sort of state transition
- In the case of **epidemics on social networks**
  - The percolation threshold is the value p_c of p at which the epidemy diffuse over all the network ("gian epidemy")

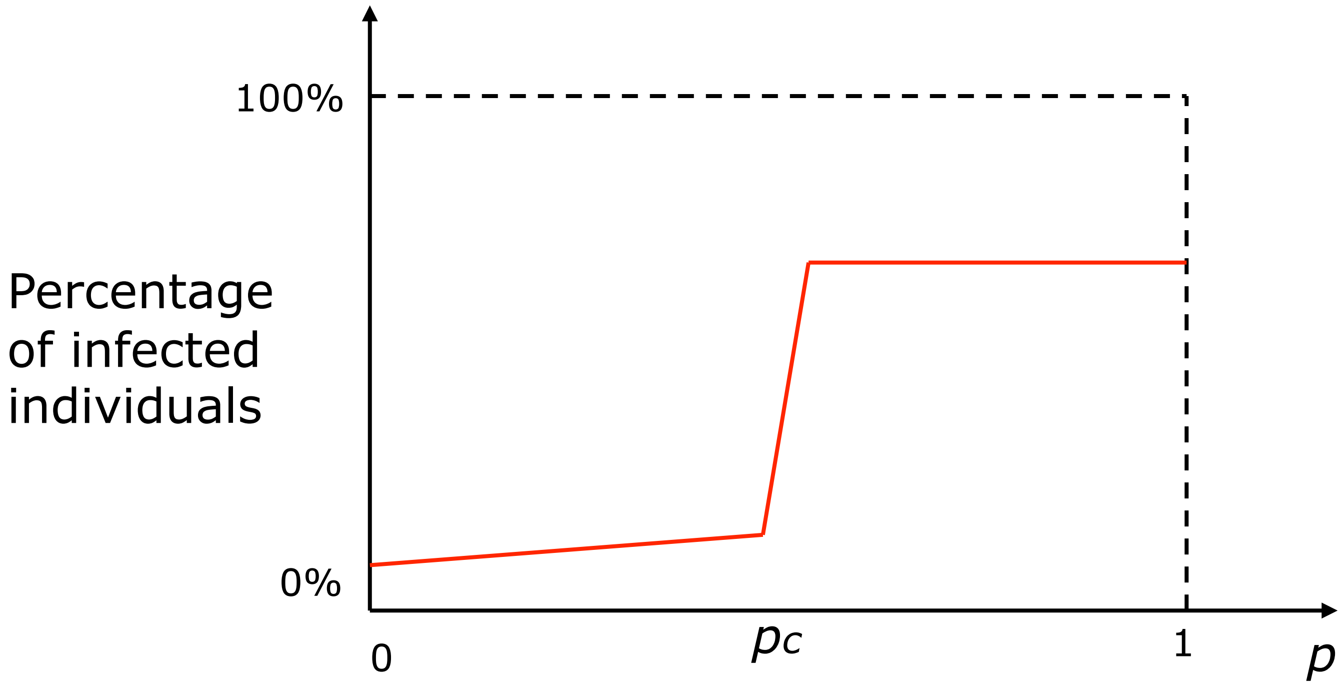
Examples of Percolation on a Ring



PERCOLATION THRESHOLD $p_c=(n-2)/n$

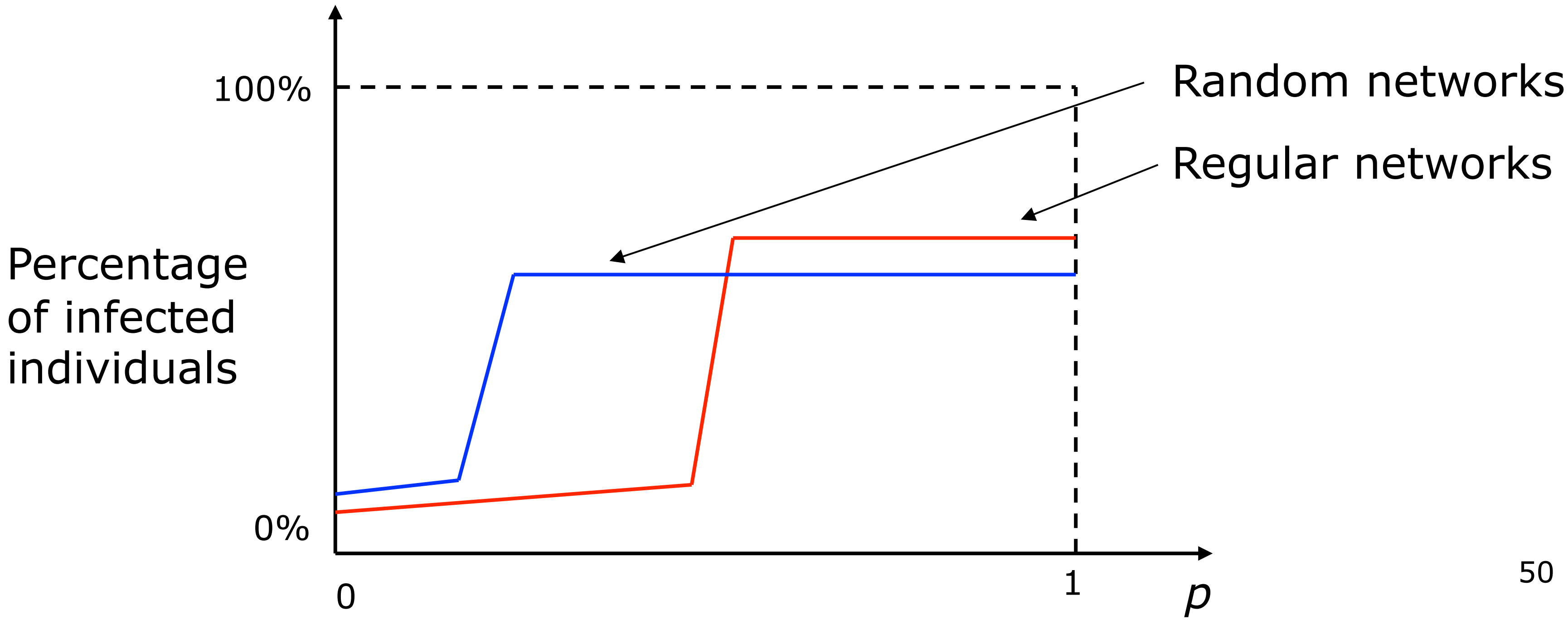
Non Linearity of Percolation

- In general, given a network, percolation exhibit a "state transition"
  - Suddenly, over the threshold, the epidemy becomes "giant"



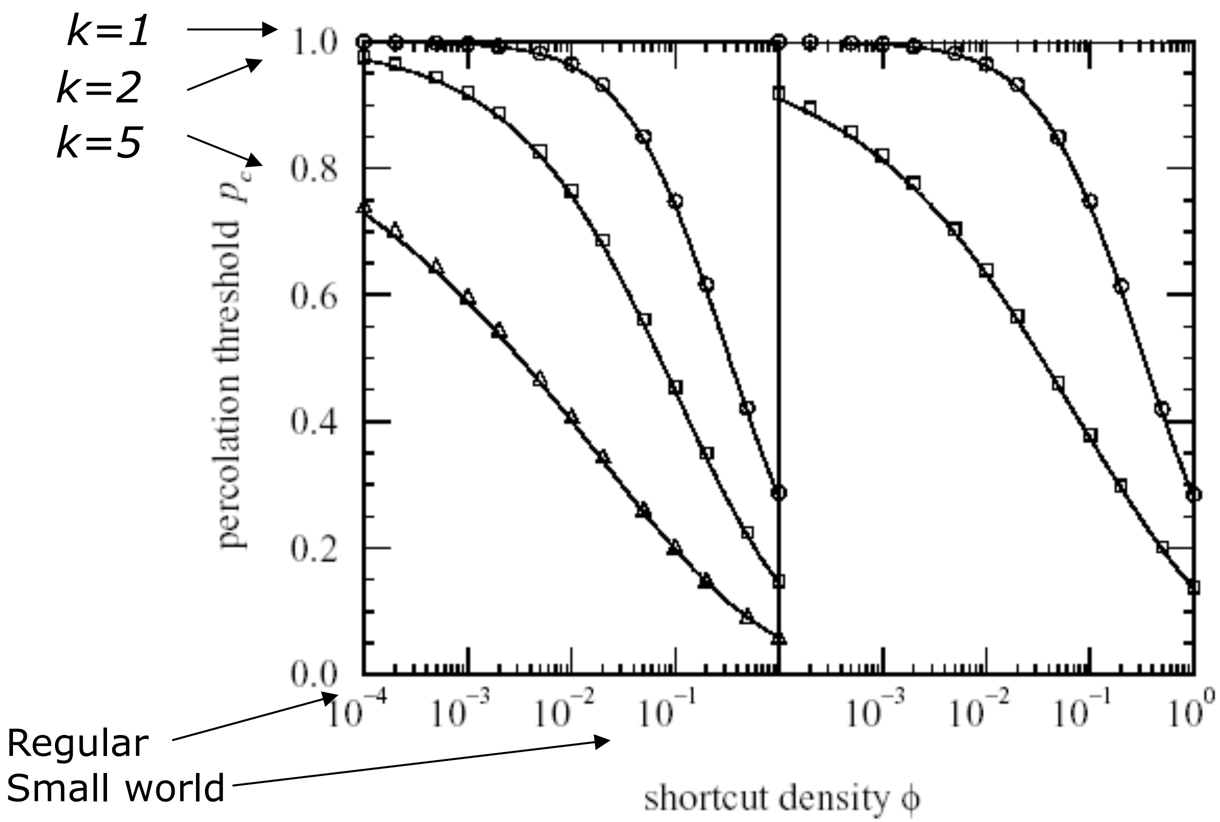
Percolation in Small World Networks (1)

- In random networks
  - The percolation threshold is rather low
- In strongly clustered networks
  - The percolation threshold is high
- In small world networks
  - The percolation thresholds approaches that of random networks, even in the presence of very limited re-wiring



Percolation in Small World Networks (2)

- The original data of Watts & Newmann for percolation on a 1-d re-wired network



Implications

- Epidemics spread much faster than predicted by normal models (assuming regular networks)
  - Even with a low percentage of susceptible individuals
- More in general
- The effects of local actions spread very fast in small world networks
  - Viruses, but also information, data, gossip, traffic jams, trends

The Spread of Internet Viruses

- The Internet and the Web are small worlds
  - So, a virus in the Internet can easily spread even if there is a low percentage of susceptible computers (e.g., without antivirus)
- The network of e-mails reflect a social network
  - So, viruses that diffuses by e-mail
    - By arriving on a site
    - And by re-sending themselves to all the e-mail addresses captured on that site
  - Actually diffuses across a social small world networks!!
- No wonder that each that a “New Virus Alert” is launched, the virus has already spread whenever possible…

The Gnutella Network (1)

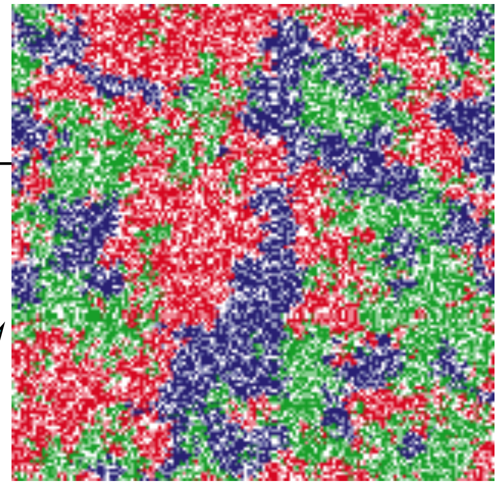
- Gnutella, in its golden months (end of 2001) counted
  - An average of 500.000 nodes for each connected clusters
  - With a maximum node degree of 20 (average 10)
- If Gnutella is a "small world" network (and this has been confirmed by tests) then
  - The average degree of separation should be, as in a random network, around
  - Log(500000)/Log(10) = 5,7
  - (why Gnutella is a small world network will be analyzed later during the course…)
- The Gnutella protocol consider 9 steps of **broadcasting** (**flooding**) requests to ALL neighbors ($p=1$)
  - Each request reach all nodes at a network distance of 9
  - Thus, a single request reaches the whole Gnutella network
  - If a file exists, we will find it!

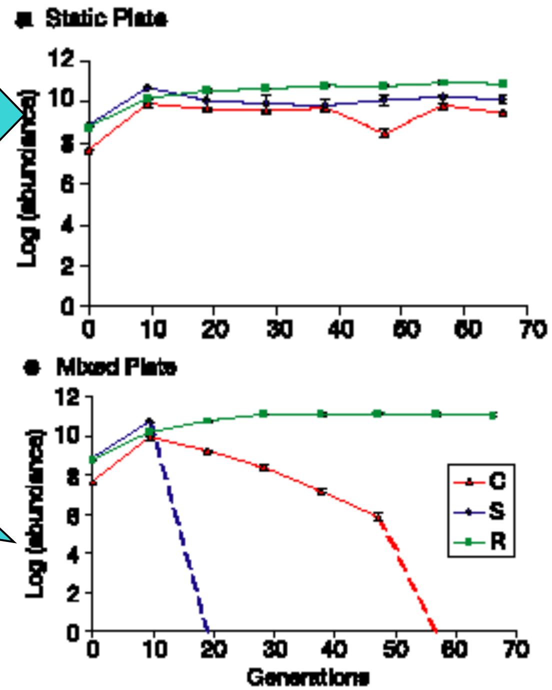
The Gnutella Network (2)

- Gnutella is based on broadcasting of request messages
  - Incurring in a dramatic traffic
  - This is why DHT architectures have been proposed
- However, given the small world nature of Gnutella
  - One can assume the presence of a percolation threshold
- Thus, one could also think at spreading requests probabilistically
- This would notably reduces the traffic on the network
  - While preserving the capability of a message of reaching the whole network
- This is a powerful technique also known as **probabilistic multicast**
  - Avoid the traffic of broadcast, and probabilistically preserve the capability of reaching the whole network

Implications for Ecology

- Experiments (Kerr, Nature, 2002)
- Three species of bacteria
  - Competing via a circular chains (paper-rock-scissors)
  - Thus, no evident winner!
- Does biodiversity get preserved?
  - **YES**, if only *local interactions* (as normal since these bacteria are sedentary)
  - **NO**, if *non-local interactions* (if wind or other phenomena creates small world shortcuts
- In general
  - Our acting in species distribution via forces migration of species
  - May strongly endanger biodiversity equilibrium





Kerr at al., Nature, June. 2002.

And more implications…

- This explains why
  - Gossip propagates so fast…
  - Trends propagates so fast
  - HIV has started propagating fast, but only recently
    - The infection in central Africa was missing short-cuts!
- And suggest that
  - We must keep into account the structure of our surrounding social networks for a variety of problems
  - E.g., marketing and advertising

Conclusions and Open Issues

- The small world model is interesting
  - It explains some properties of social networks
  - As well as some properties of different types of technological networks
  - And specific of modern distributed systems
- However
  - It is not true (by experience and measurement) that all nodes in a network have the same k
  - Networks are not static but continuously evolve
  - We have not explained how these networks actually form and evolve. This may be somewhat clear for social networks, it is not clear for technological ones
  - Studying network formation and evolution we can discover additional properties and phenomena