Introduction to Social Network Analysis and Models

Lecture 2

范耀中
Yao-Chung Fan
yfan@nchu.edu.tw,
Dept. of Computer Science,
National Chung Hsing University

Lecture Outline

- Part I:
  - Social Network Analysis
- Part II:
  - Social Network Models
  - Social Network Tools

Social Network Models

Models for Real World Network

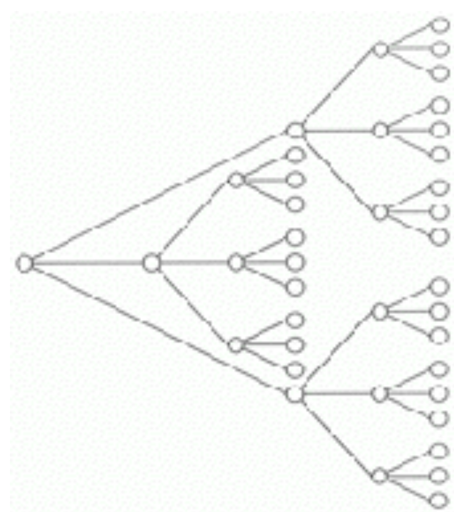
	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57		0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01		0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1 439	1 723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1 377	2 213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1 296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2 359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n; total number of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

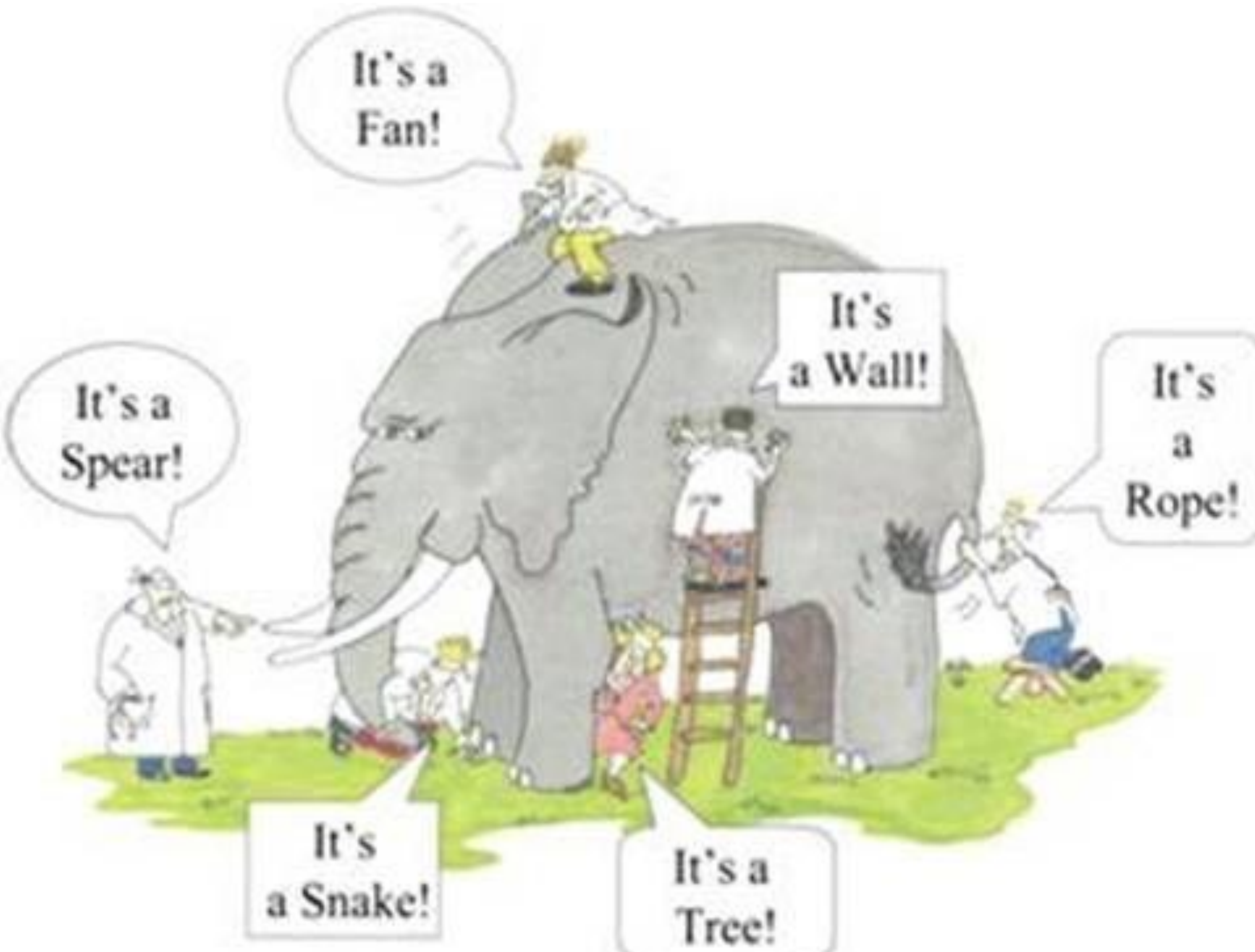
Network Properties

- Average Distance between Pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)

a graph has *k* average degree then
the first neighbours will be *k*
the second neighbours ~ k^2
……………..
the d-th neighbours ~ k^d



What the structure of a social network look like?



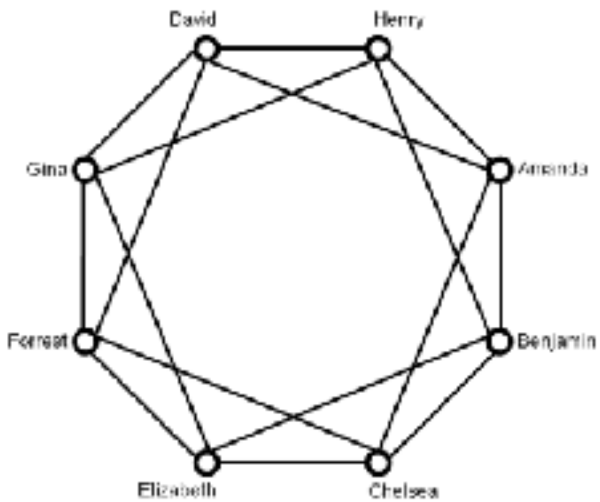
Small World Phenomenon -->
(現象)

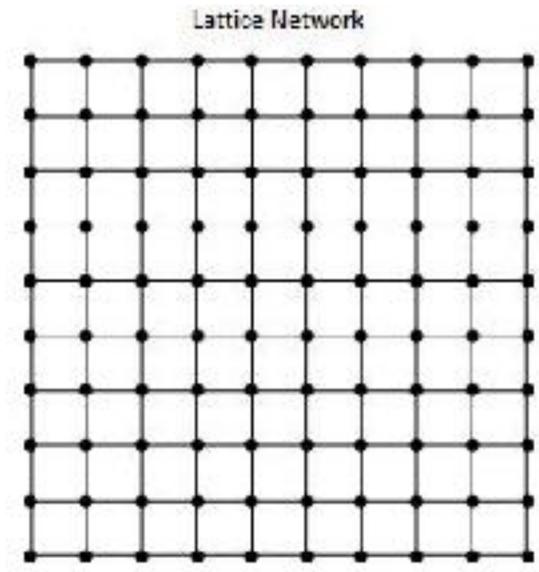
Small World Models <--
(模型)

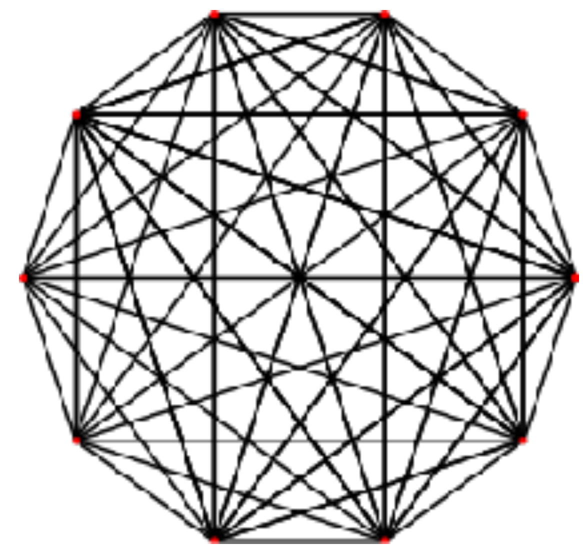
Five Models

- Regular Model
- Random Graph Model
- Small World Model/Watts-Strogatz Model
- Scale Free Model/BA Model
- Geographical Small World Model

Regular Models







$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}.$$

Why the regular models are not good ones ?

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1022	5103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
technological	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4941	6594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366
	software packages	directed	1439	1723	1.20	2.42	1.6/1.4	0.070	0.082	−0.016	318
	software classes	directed	1377	2213	1.61	1.51	–	0.033	0.012	−0.119	395
	electronic circuits	undirected	24 097	53 248	4.34	11.05	3.0	0.010	0.030	−0.154	155
	peer-to-peer network	undirected	880	1296	1.47	4.28	2.1	0.012	0.011	−0.366	6, 354
biological	metabolic network	undirected	765	3686	9.64	2.56	2.2	0.090	0.67	−0.240	214
	protein interactions	undirected	2115	2240	2.12	6.80	2.4	0.072	0.071	−0.156	212
	marine food web	directed	135	598	4.43	2.05	–	0.16	0.23	−0.263	204
	freshwater food web	directed	92	997	10.84	1.90	–	0.20	0.087	−0.326	272
	neural network	directed	307	2359	7.68	3.97	–	0.18	0.28	−0.226	416, 421

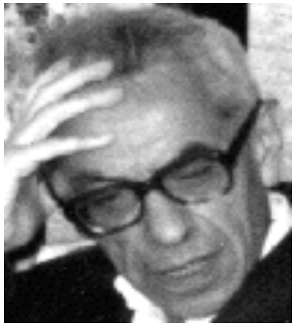
TABLE II Basic statistics for a number of published networks. The properties measured are: type of graph, directed or undirected; total number of vertices n; total number of edges m; mean degree z; mean vertex–vertex distance ℓ; exponent α of degree distribution if the distribution follows a power law (or "–" if not; in/out-degree exponents are given for directed graphs); clustering coefficient $C^{(1)}$ from Eq. (3); clustering coefficient $C^{(2)}$ from Eq. (6); and degree correlation coefficient r, Sec. III.F. The last column gives the citation(s) for the network in the bibliography. Blank entries indicate unavailable data.

Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)

Random Graph Model

Standard Theory of Random Graph
(Erdös and Rényi 1960)



Random Graphs are composed by starting with n vertices. With probability *p* two vertices are connected by an edge

$p = 0.1$ $p = 0.25$ $p = 0.5$

http://ccl.northwestern.edu/netlogo/models/run.cgi?GiantComponent.884.534

Random Graph Model Property

P

The number m of edges in a Random Graph is a random variable whose expectation value is

$$E(m) = p\frac{N(N-1)}{2}$$

The probability to form a particular Graph *G(N,m)* is given by

$$E(G(N,m)) = \underline{p^m}\,\underline{(1-p)^{\frac{N(N-1)}{2}-m}}$$

The degree has expectation value $E(k) = 2m/N = p(N-1) \cong pN$

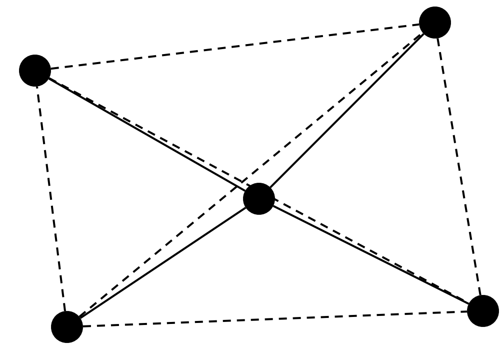
It is easy to check that the degree probability distribution is given by

* $c = p$

$$\boxed{P(k) = \binom{N-1}{k} p^k (1-p)^{(N-1)-k} \cong \frac{(pN)^k e^{-pN}}{k!}}$$

Random Graph Model Property

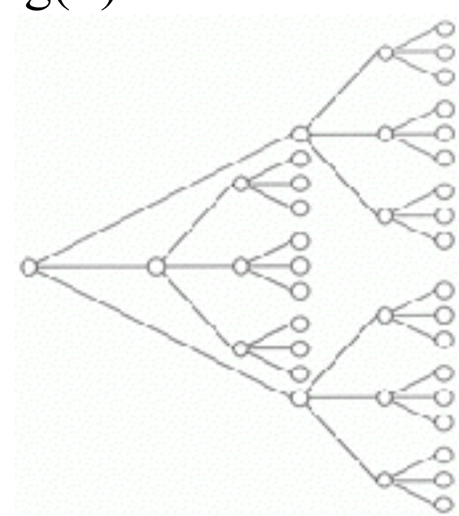
Clustering Coefficient: $E(C) \cong p$



$$C_i = \frac{\text{number of triangles connected to vertex } i}{\text{number of triples centered on vertex } i}.$$

Average Distance between pairs: $l \leq D \cong \dfrac{\log(N)}{\log(k)}$

a graph has *k* average degree then
the first neighbours will be *k*
the second neighbours ~ k^2
……………..
the d-th neighbours ~ k^d



Why the random graph model is inadequate ?

$$E(m) = p\frac{N(N-1)}{2} = C\frac{N(N-1)}{2}$$

M太大

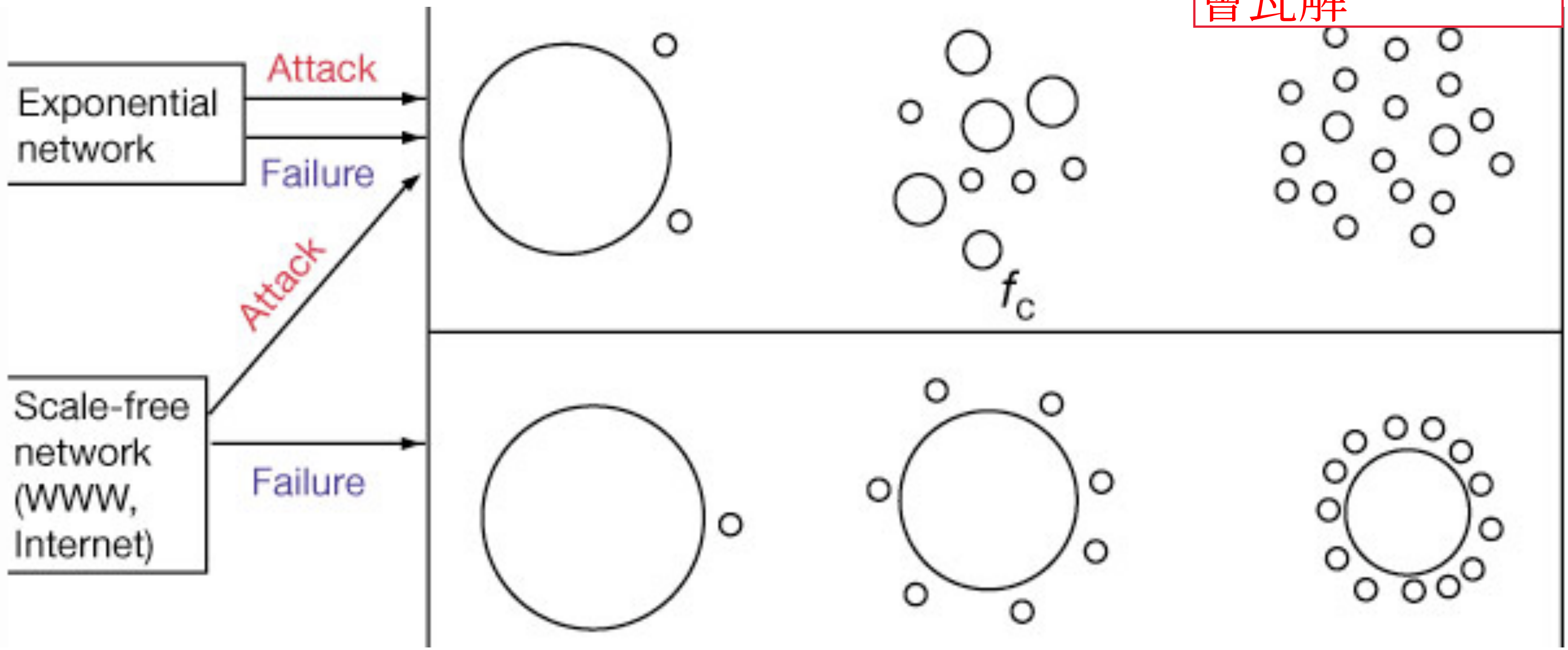
$$E(m) = 0.78 \cdot \frac{449913 \cdot 449912}{2} = 78{,}944{,}290{,}485 \approx 79 * 10^9$$

正確的
network
C大M小

	network	type	n	m	z	ℓ	α	$C^{(1)}$	$C^{(2)}$	r	Ref(s).
social	film actors	undirected	449 913	25 516 482	113.43	3.48	2.3	0.20	0.78	0.208	20, 416
	company directors	undirected	7 673	55 392	14.44	4.60	–	0.59	0.88	0.276	105, 323
	math coauthorship	undirected	253 339	496 489	3.92	7.57	–	0.15	0.34	0.120	107, 182
	physics coauthorship	undirected	52 909	245 300	9.27	6.19	–	0.45	0.56	0.363	311, 313
	biology coauthorship	undirected	1 520 251	11 803 064	15.53	4.92	–	0.088	0.60	0.127	311, 313
	telephone call graph	undirected	47 000 000	80 000 000	3.16		2.1				8, 9
	email messages	directed	59 912	86 300	1.44	4.95	1.5/2.0		0.16		136
	email address books	directed	16 881	57 029	3.38	5.22	–	0.17	0.13	0.092	321
	student relationships	undirected	573	477	1.66	16.01	–	0.005	0.001	−0.029	45
	sexual contacts	undirected	2 810				3.2				265, 266
information	WWW nd.edu	directed	269 504	1 497 135	5.55	11.27	2.1/2.4	0.11	0.29	−0.067	14, 34
	WWW Altavista	directed	203 549 046	2 130 000 000	10.46	16.18	2.1/2.7				74
	citation network	directed	783 339	6 716 198	8.57		3.0/–				351
	Roget's Thesaurus	directed	1 022	5 103	4.99	4.87	–	0.13	0.15	0.157	244
	word co-occurrence	undirected	460 902	17 000 000	70.13		2.7		0.44		119, 157
gical	Internet	undirected	10 697	31 992	5.98	3.31	2.5	0.035	0.39	−0.189	86, 148
	power grid	undirected	4 941	6 594	2.67	18.99	–	0.10	0.080	−0.003	416
	train routes	undirected	587	19 603	66.79	2.16	–		0.69	−0.033	366

Network Resilience

randrom 的不管
attack 或 failure
到一定程度時都
會瓦解



Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási.

http://ccl.northwestern.edu/netlogo/models/run.cgi?GiantComponent.884.534

Network Properties

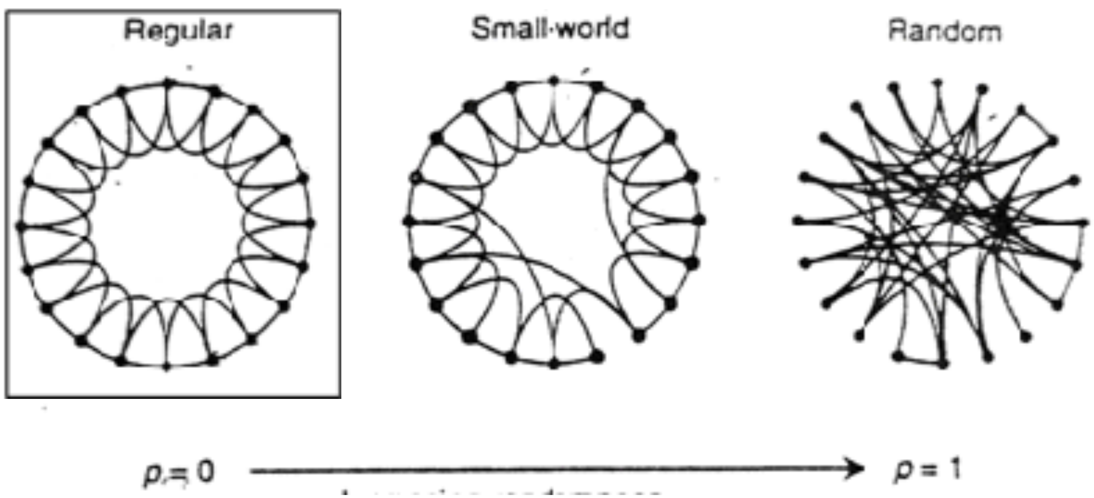
- Average Distance between Pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)

THE LAST JUDGEMENT – PART 9

Small World Model

Watts and Strogatz Model (Na

一開始正規排列，在依照p的機率隨機拆邊，並將其連到較遠的點(rewiring)，類似建捷徑的概念，使得有high clustering和small average distance



The original graph was very clustered: we keep this high clustering.

In the model, we begin with a low-dimension regular graph

For each edge, move one of its ends to another vertex with probablity *p*.

And by creating shortcuts, we decrease the average distance, *i.e.* create a small-world effect.

Watts-Strogatz model:

Generating small world graphs

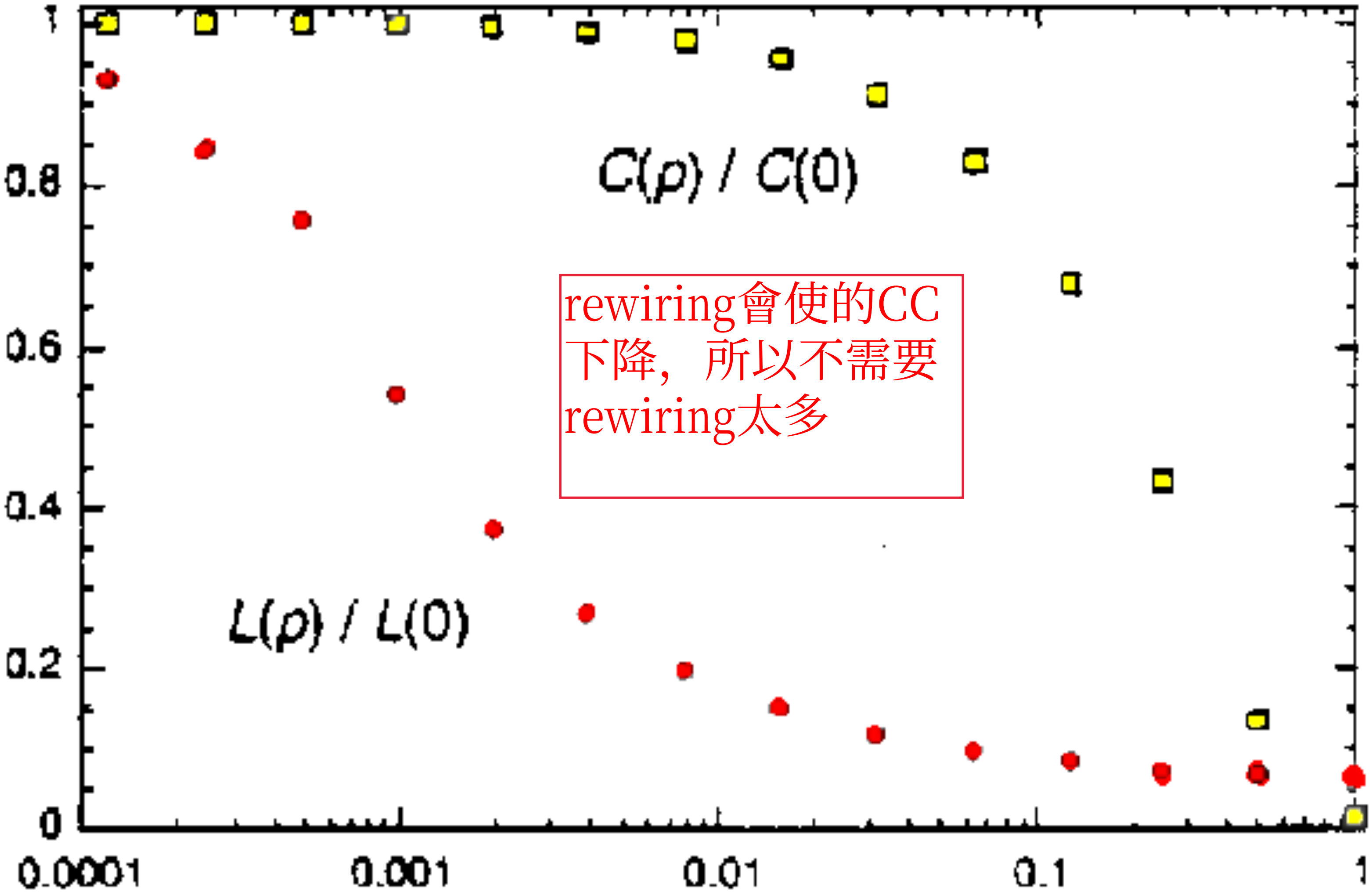
Select a fraction p of edges
Reposition on of their endpoints

Add a fraction p of additional edges leaving underlying lattice intact

- As in many network generating algorithms
  - Disallow self-edges
  - Disallow multiple edges

Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

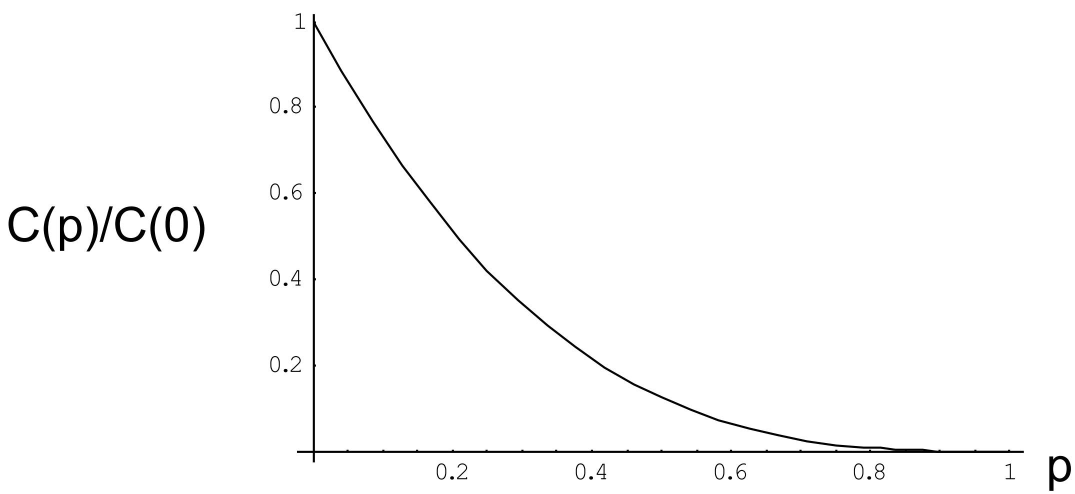
- By varying the probability of rewiring edges



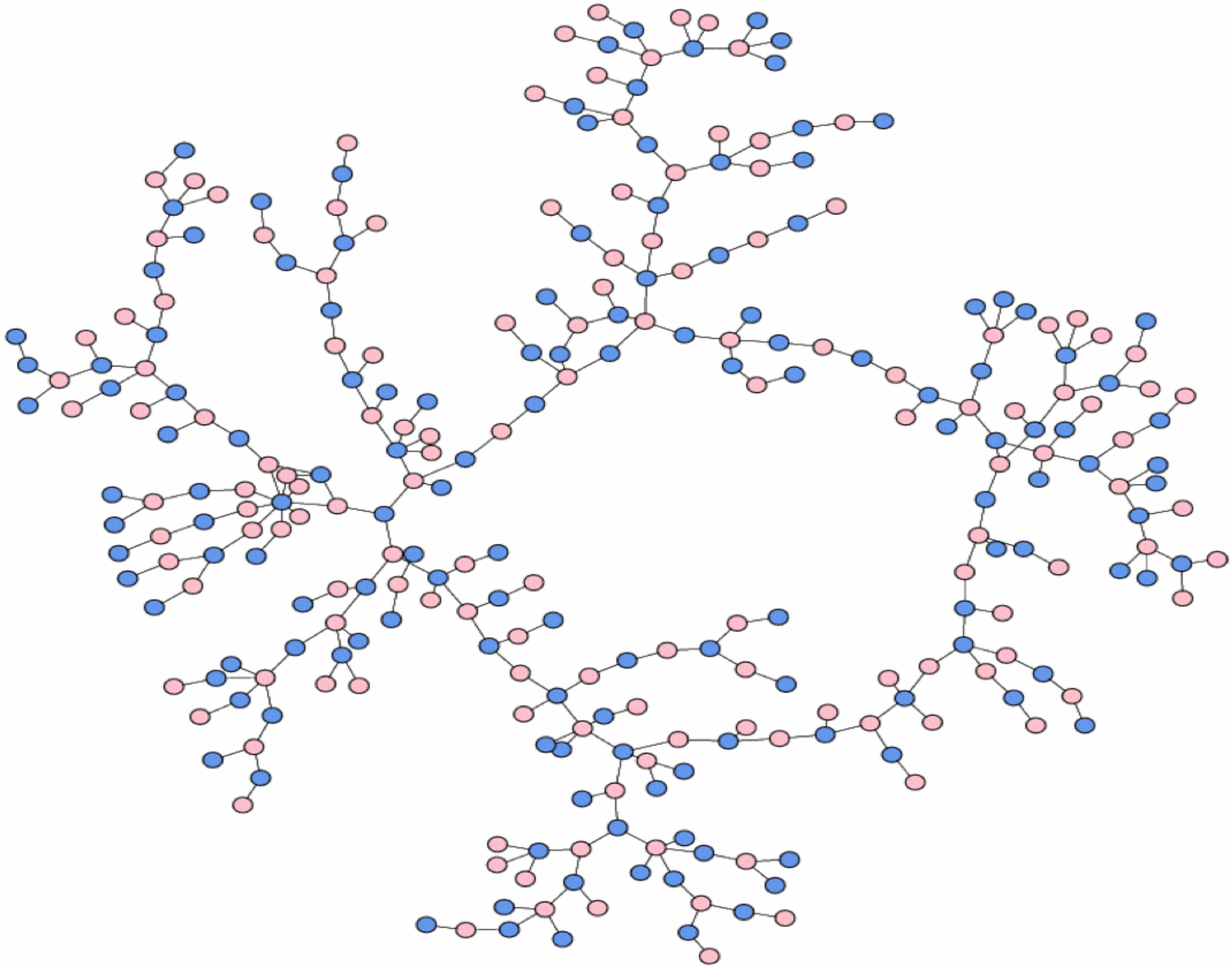
Duncan J. Watts & Steven H. Strogatz, Nature 393, 440-442 (1998)

Watts/Strogatz model: Clustering coefficient

- The probability that a connected triple stays connected after rewiring
  - probability that none of the 3 edges were rewired $(1-p)^3$
- Clustering coefficient = $C(p) = C(p=0)*(1-p)^3$

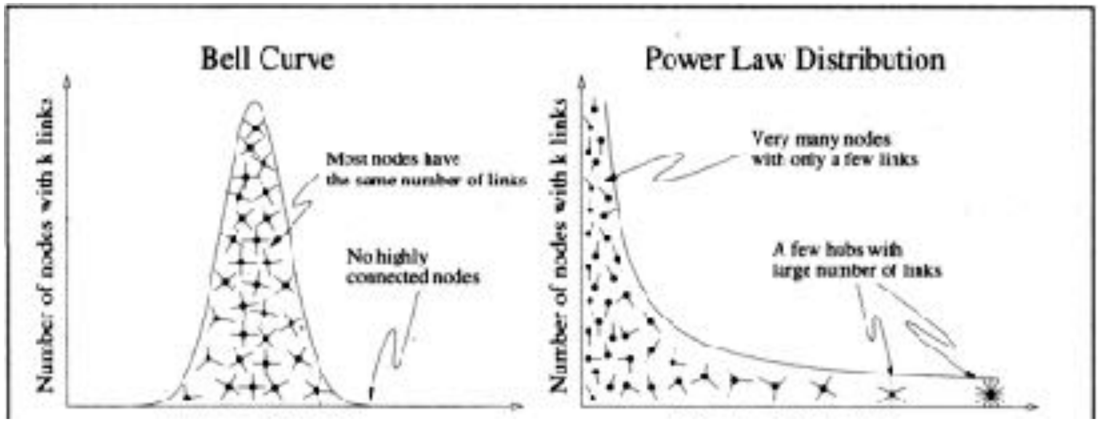


Source: Watts, D.J., Strogatz, S.H.(1998) Collective dynamics of 'small-world' networks. Nature 393:440-442.

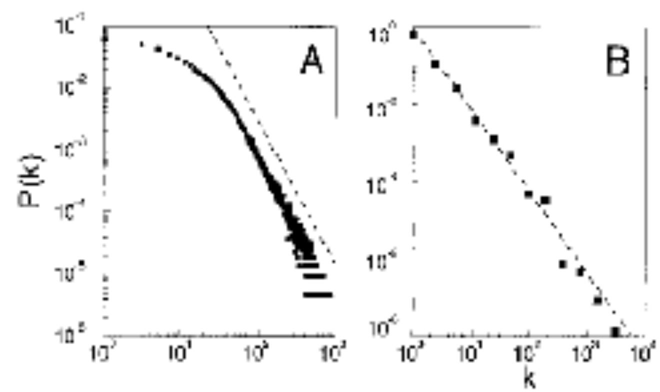


Watts-Strogatz Model

Did Watts and Strogatz Model Explain All ?



Network	*n*	*k*	*L*	*C*
WWW pages	153127	35.21	3.1	0.1078
Internet AS	6209	4.11	3.76	.3
Math co-authors	70975	3.9	9.5	.59
Power Grid	4941	2.67	18.7	0.08
E-coli reaction	315	28.3	2.62	.59



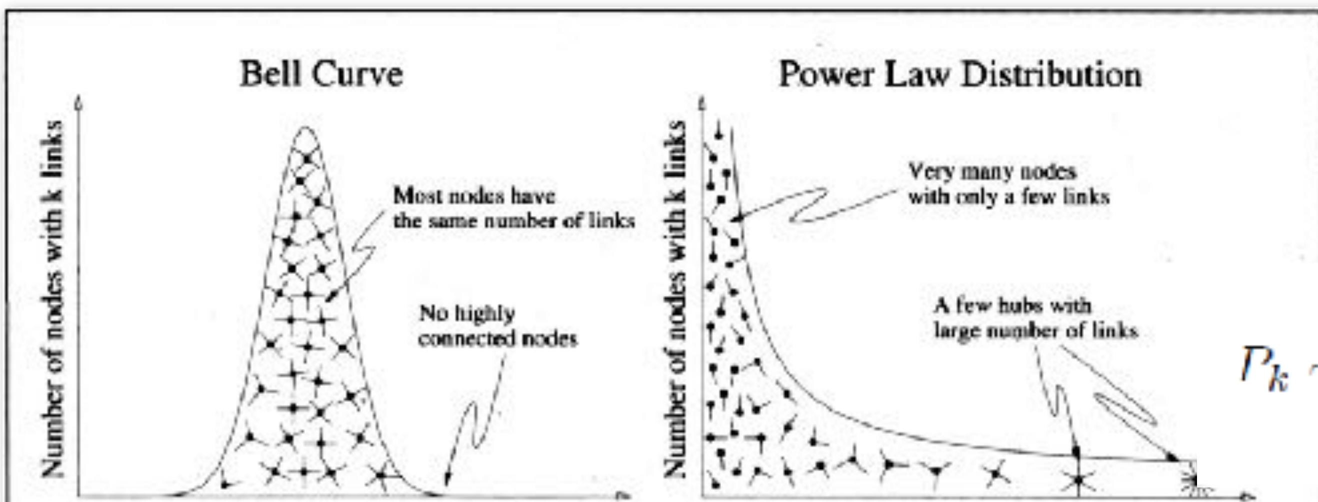
Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)

Barabasi-Albert model

Each node connects to other nodes with probability proportional to their degree the process starts with some initial subgraph each node comes with m edges

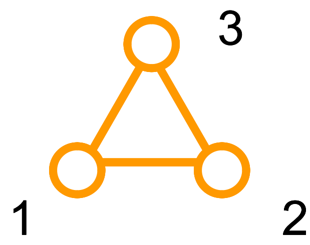
Results in power-law with exponent α = 3



$$P_k \sim \sum_{k'=k}^{\infty} k'^{-\alpha} \sim k^{-(\alpha-1)}.$$

Basic BA-model ()

- start with an initial set of m_0 fully connected nodes
  - e.g. $m_0 = 3$



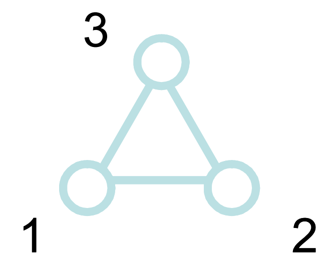
- now add new vertices one by one, each one with exactly *m* edges
- each new edge connects to an existing vertex in proportion to the number of edges that vertex already has → ***preferential attachment***
- easiest if you keep track of edge endpoints in one large array and select an element from this array at random
  - the probability of selecting any one vertex will be proportional to the number of times it appears in the array – which corresponds to its degree

http://ccl.northwestern.edu/netlogo/models/run.cgi?PreferentialAttachment.836.527

generating BA graphs – cont'd

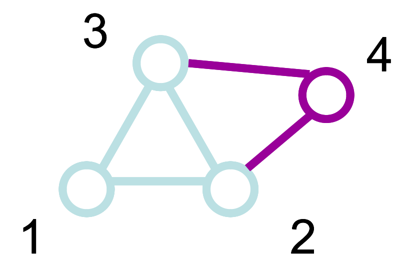
- To start, each vertex has an equal number of edges (2)
  - the probability of choosing any vertex is 1/3

1 1 2 2 3 3



- We add a new vertex, and it will have m edges, here take m=2
  - draw 2 random elements from the array – suppose they are 2 and 3

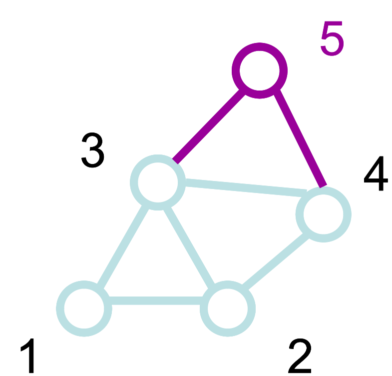
1 1 2 2 2 3 3 3 4 4

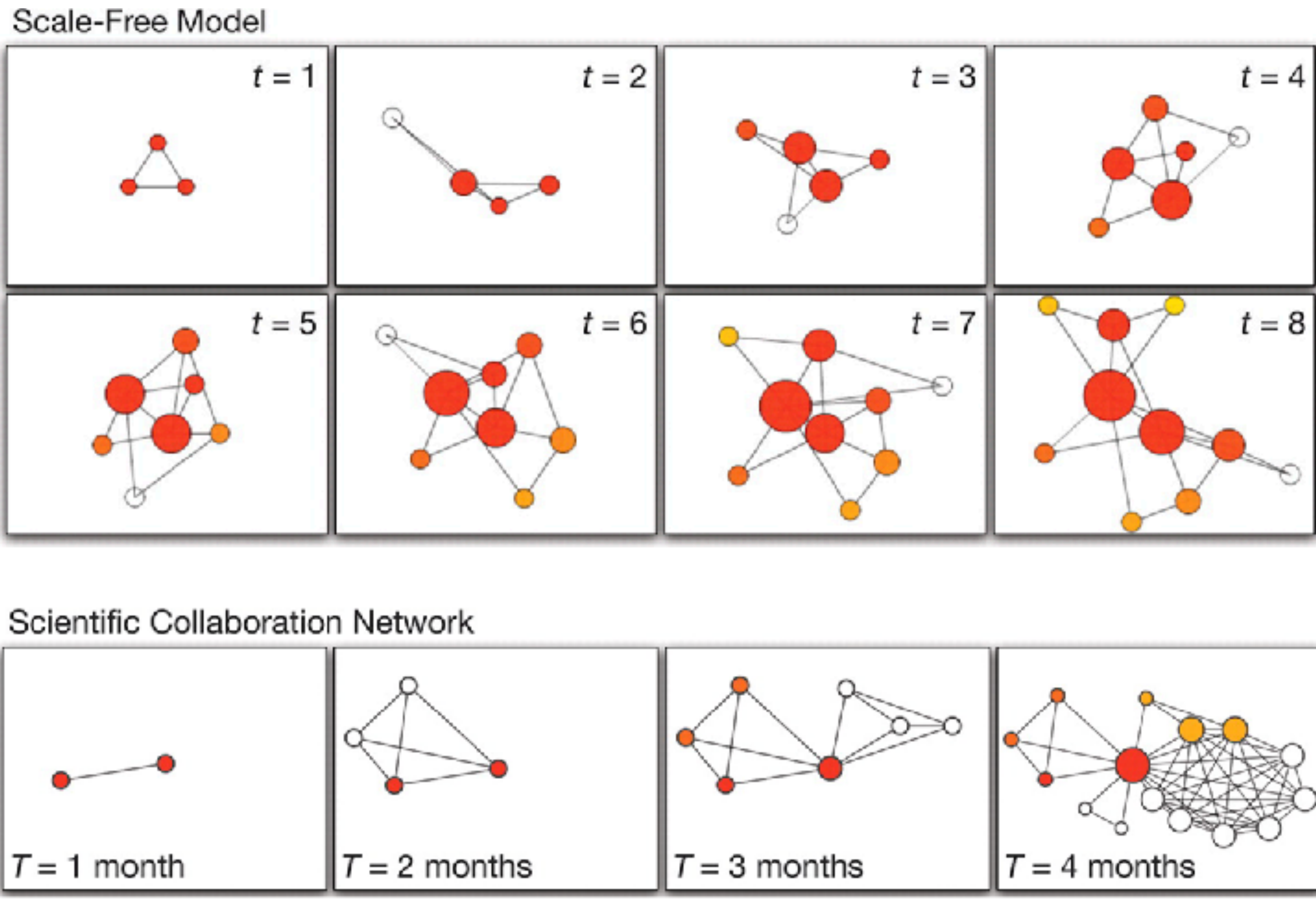


- Now the probabilities of selecting 1,2,3,or 4 are 1/5, 3/10, 3/10, 1/5

- Add a new vertex, draw a vertex for it to connect from the array
  - etc.

1 1 2 2 2 3 3 3 3 4 4 4 5 5



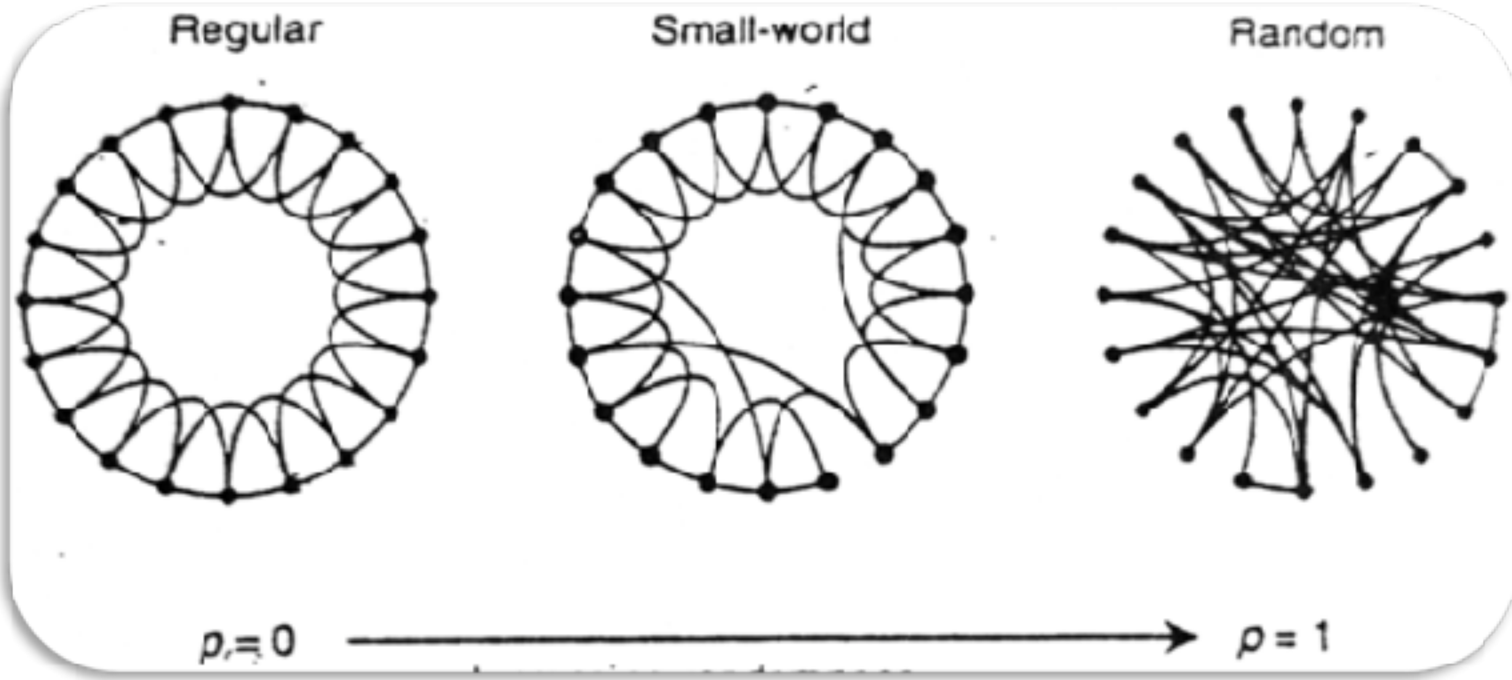


Scale-Free Model
t = 1
t = 2
t = 3
t = 4
t = 5
t = 6
t = 7
t = 8
Scientific Collaboration Network
T = 1 month
T = 2 months
T = 3 months
T = 4 months

Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)

BA Model + Watts/Strogatz Model



Network Properties

- Average Distance between pairs (small)
- Transitivity (high)
- Degree Distribution (power law)
- Network Resilience (weak under attack)


死亡之握
VS
梵帝岡
聖光

Summary So Far

- Small World Phenomena
- Four Models
  - Regular model
  - Random model
  - WS model
  - BA model
- Applications

What we learn from the small world phenomenon ?

A very big network in which any pairs of nodes are reachable via short chains of intermediates

It tells you that exchanging information in small-world networks will be very fast !

A very big computer network in which exchanging information between nodes are very fast !

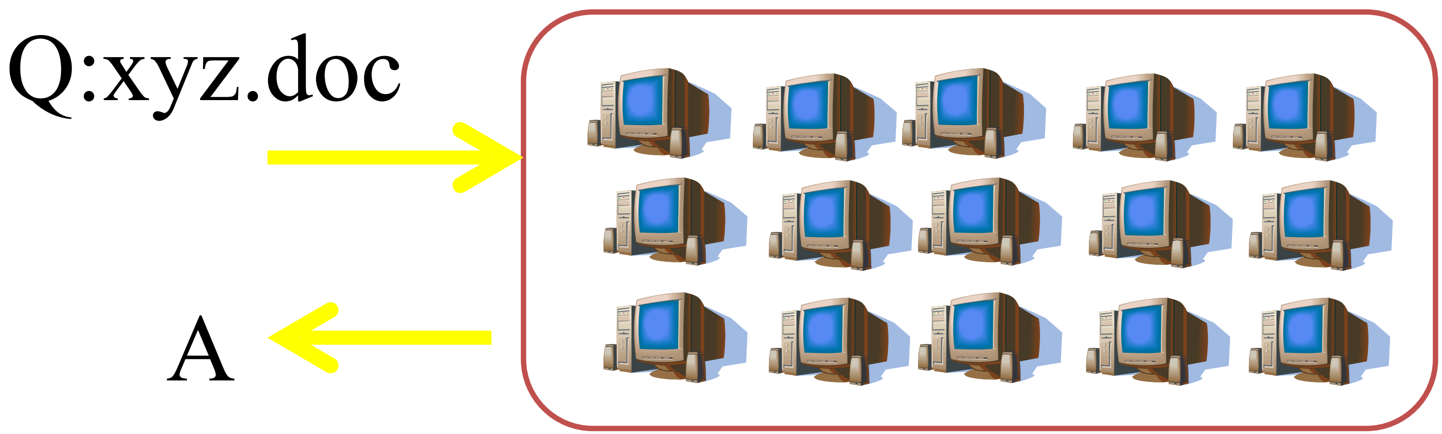
Cloud System Efficient

Efficient Cloud System

Mimic a small world system to design a data storage system

Given a huge set of computers, how to utilize them to have a data storage system

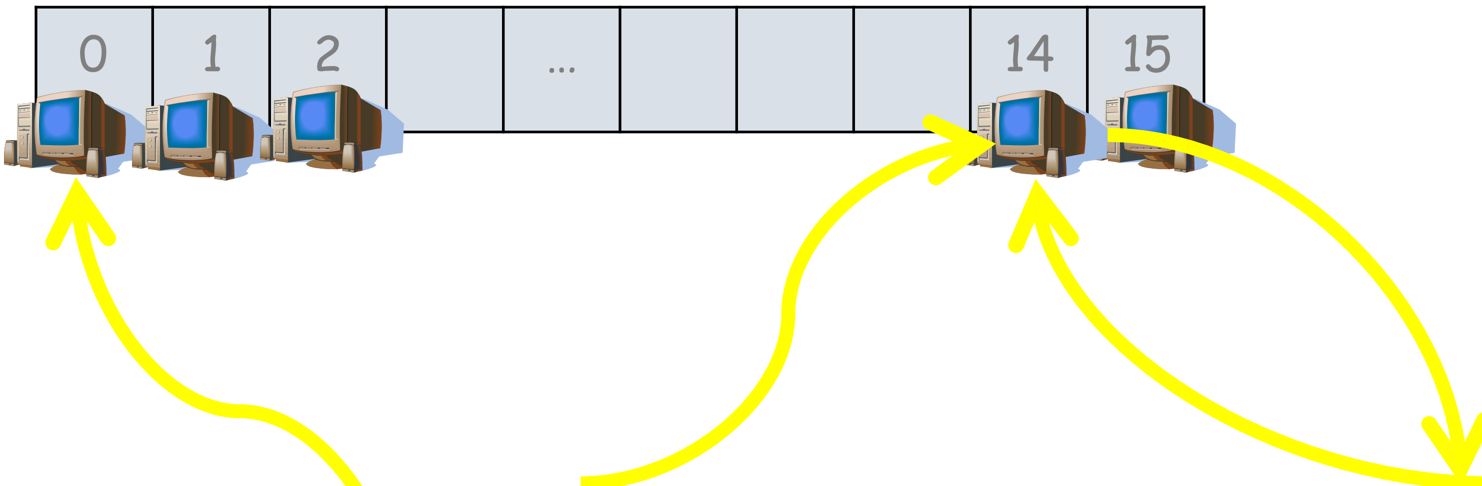
P1: how data are stored
P2: how data are efficiently retrieved



As a system

P1: How data are stored ?

Content Addressable



Hash(abc.doc) = 14

Hash(xyz.doc) = 0

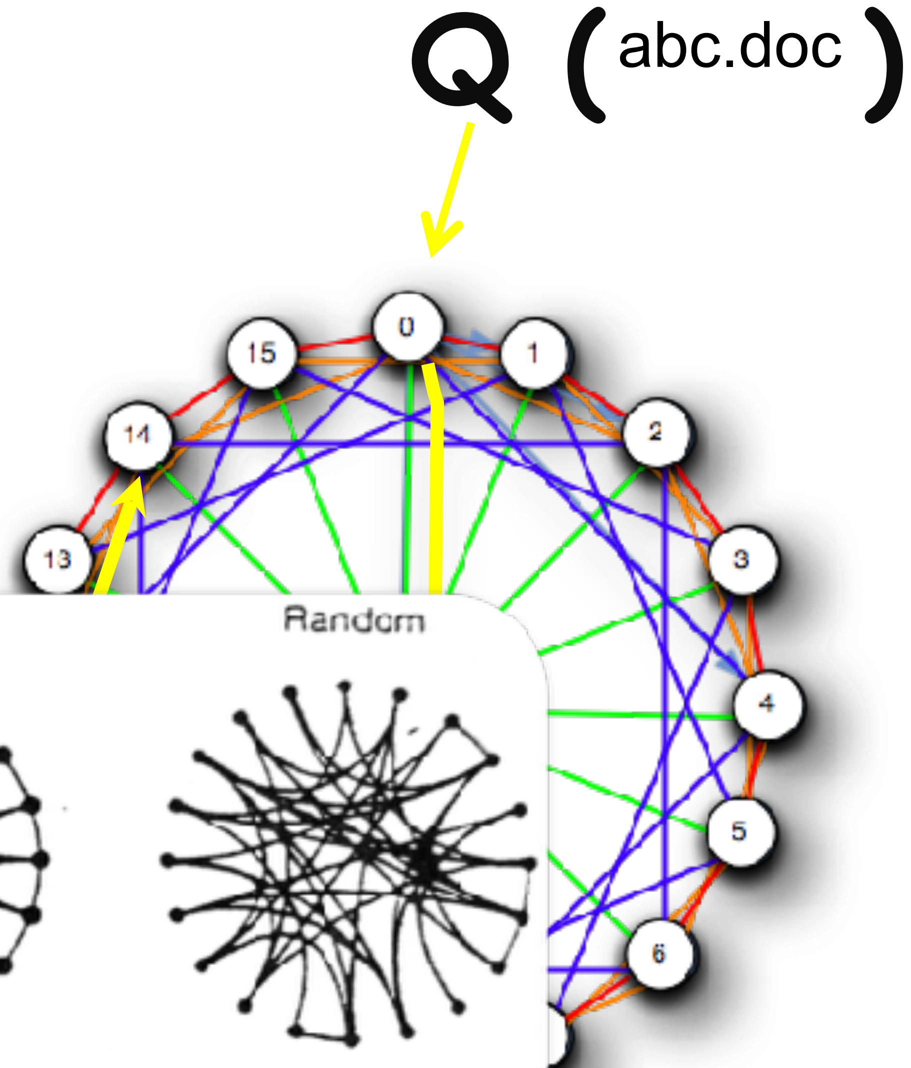
Q(abc.doc)

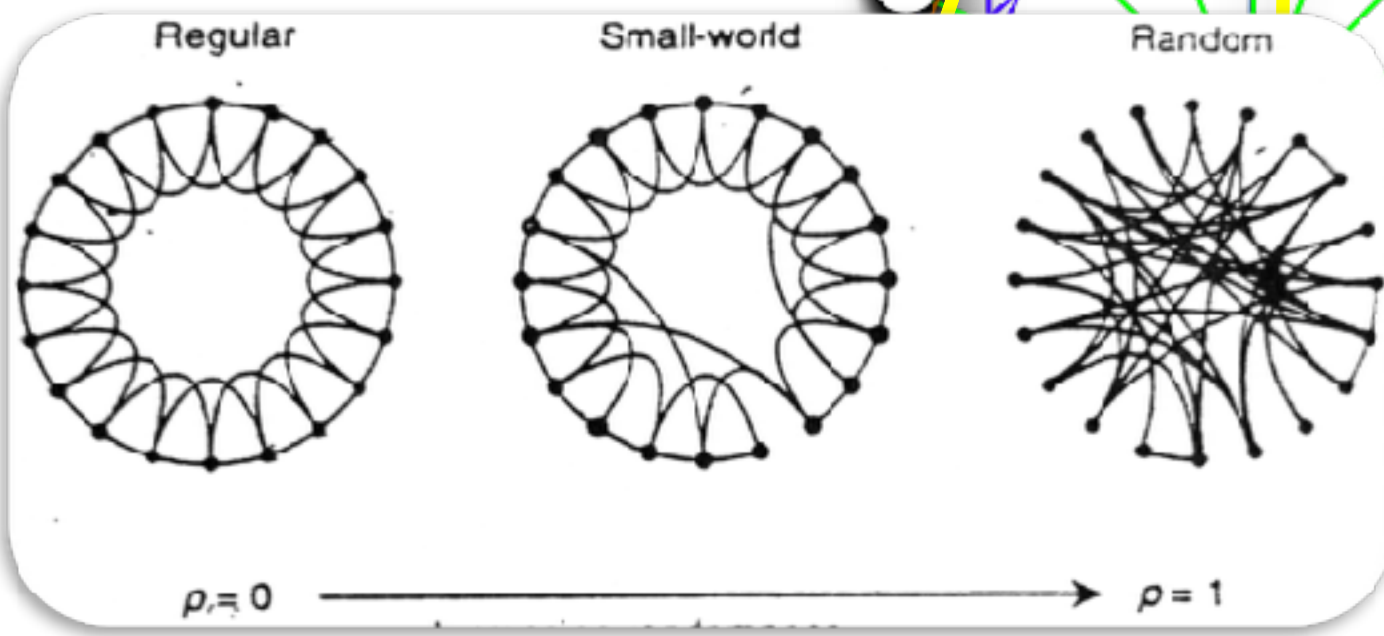
= Hash (abc.doc)

= 14

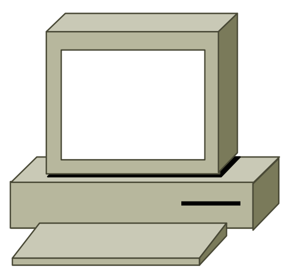
P2: How data are retrieved ?

Small World Like Topology





Hash (data) =14



2

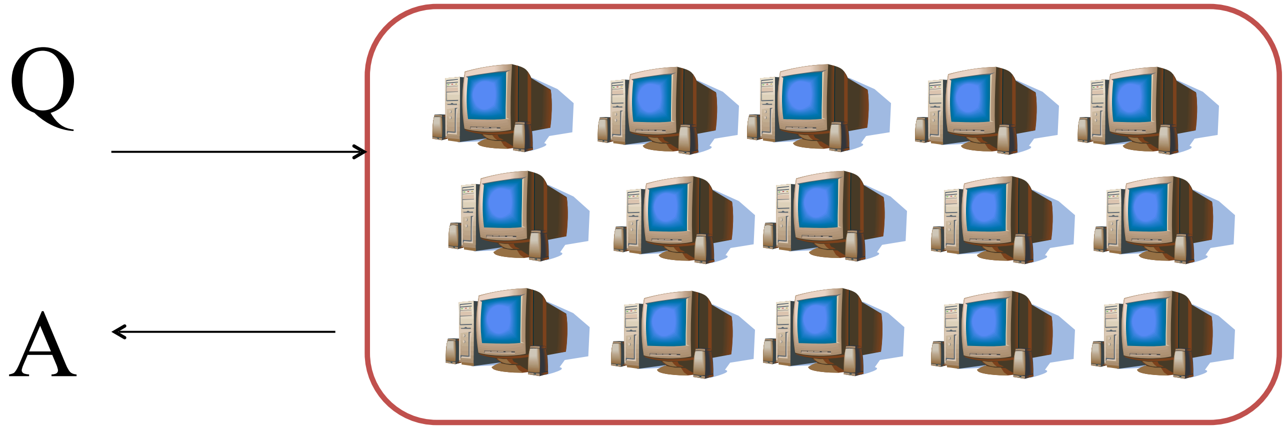
6

12

What we get from studying small world ?

Given a huge set of computers, how to utilize them to have a data storage system

We can have a robust, load balance, efficient data storage system



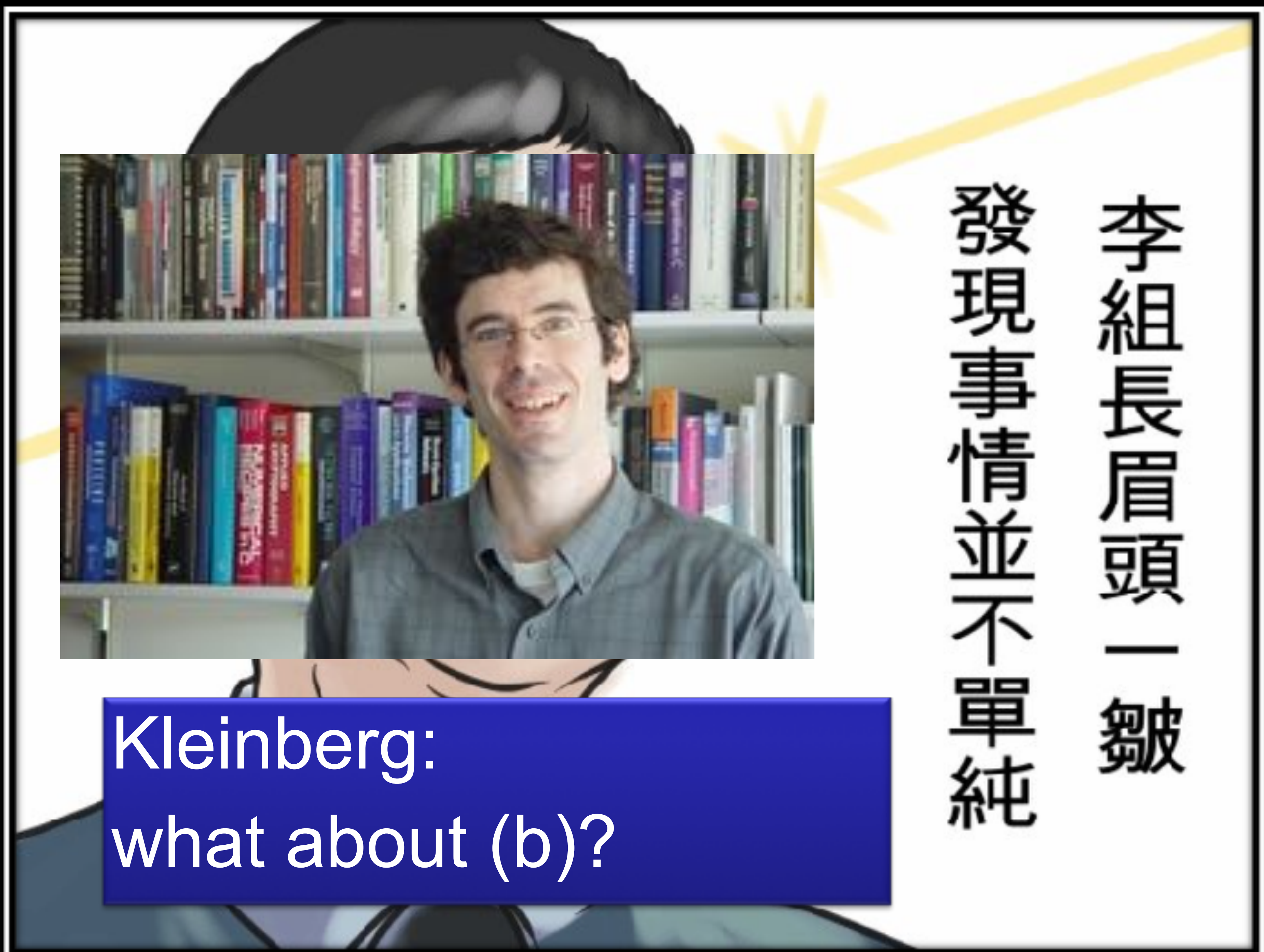
As a system

Five Models

- Regular Model
- Random Graph Model
- Small World Model/Watts-Strogatz Model
- Scale Free Model/BA Model
- Geographical Small World Model

Milgram's experiment revisited

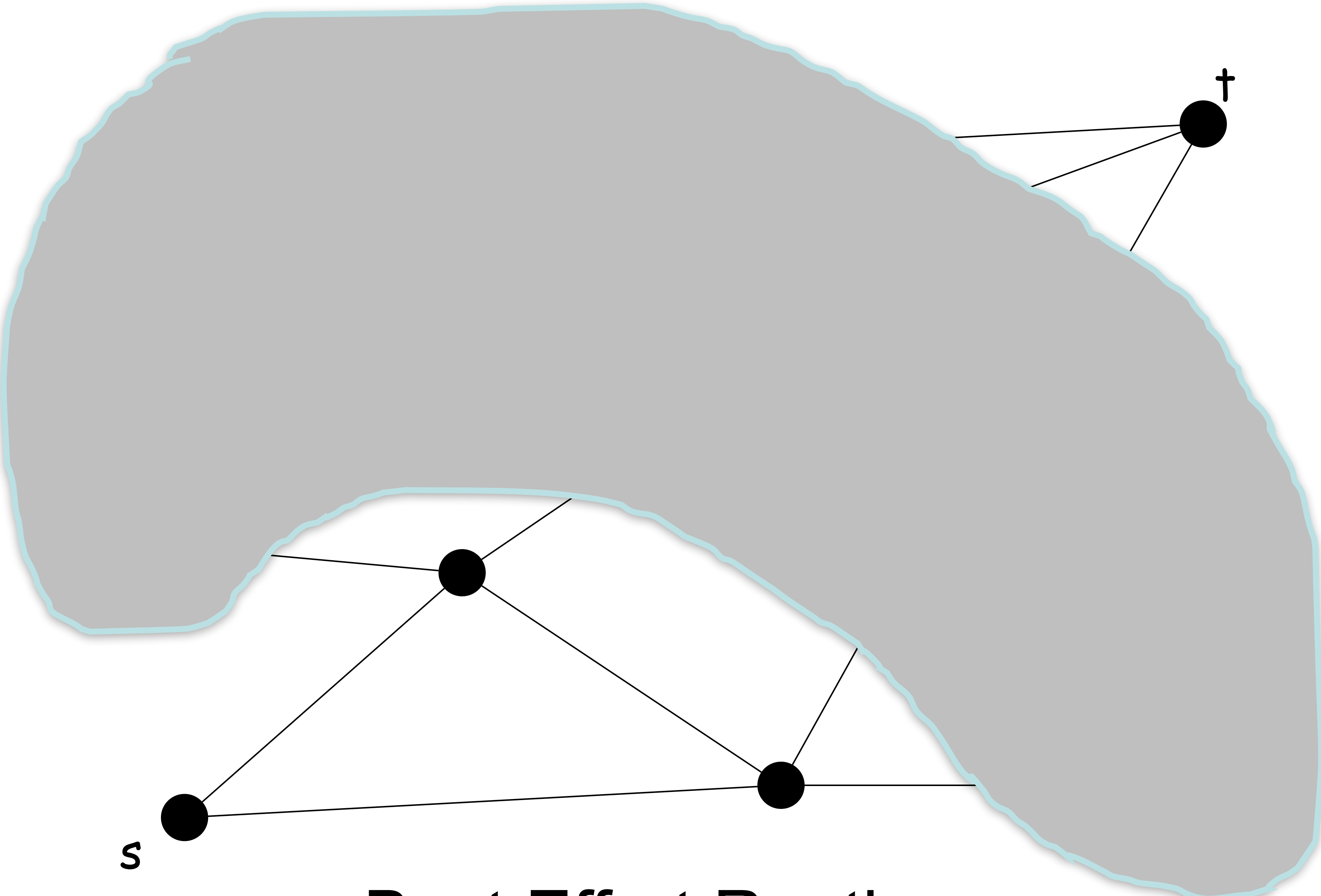
- What did Milgram's experiment show?
  - (a) There are short paths in large networks that connect individuals
  - (b) People are able to find these short paths using a simple, greedy, decentralized algorithm
- Small world models take care of (a)


李組長眉頭一皺
發現事情並不單純
Kleinberg:
what about (b)?

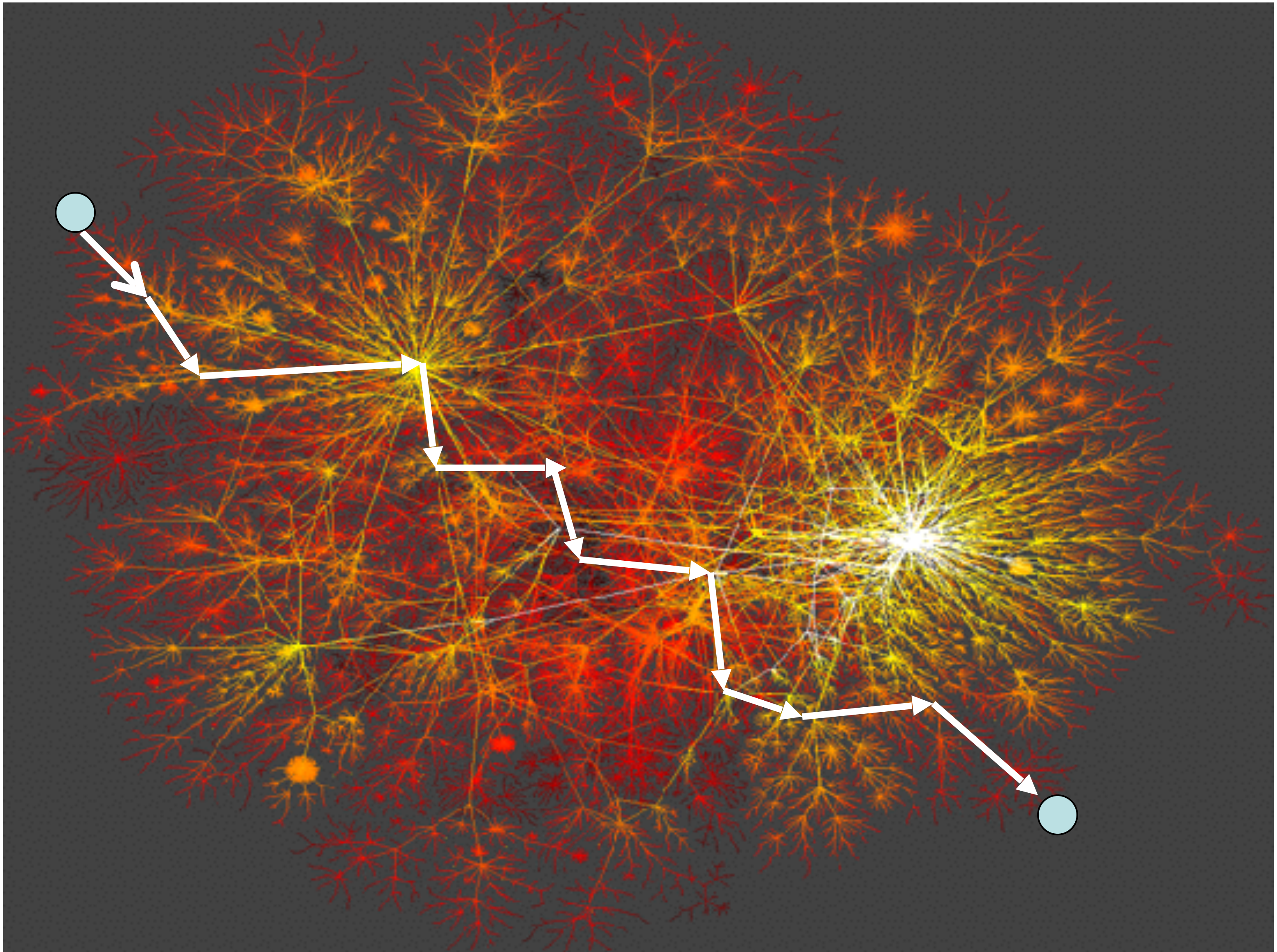
The Algorithmic Side

- Input:
  - Grid G = (V,E)
  - arbitrary nodes *s*, *t*

- Goal: Transmit a message from *s* to *t* in as few steps as possible using only locally available information



Best Effort Routing

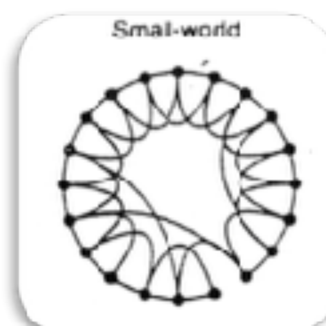


Navigation in a small world network

- Kleinberg (2000)
  - Why should arbitrary pairs of strangers, using only locally available information, be able to ***find*** short chains of acquaintances that link them together?
  - Does this occur in all small-world networks, or are there properties that must exist for this to happen?

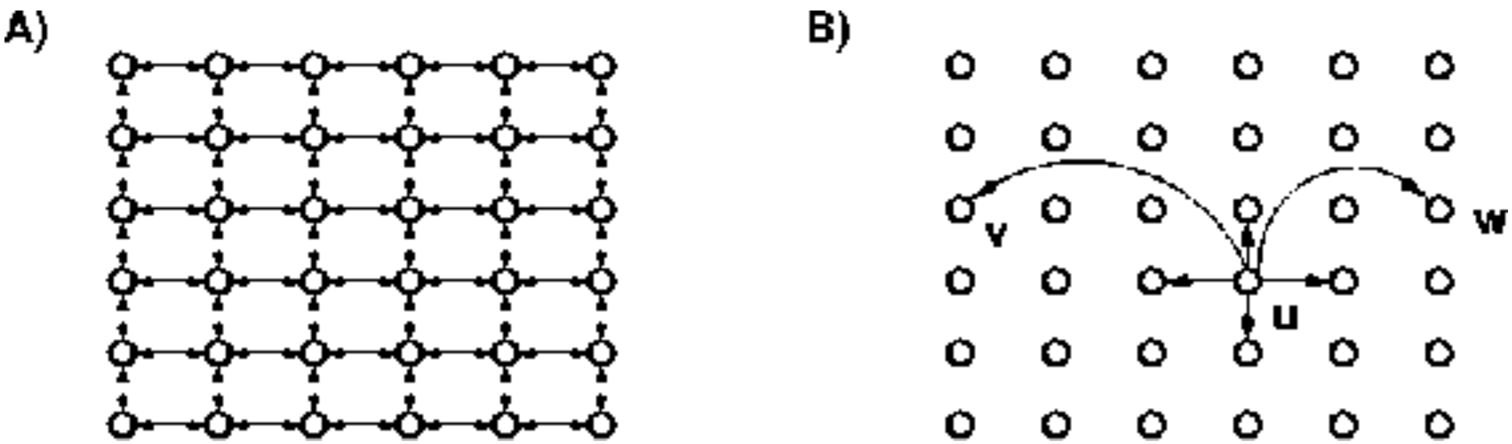
The Algorithmic Side

- Assumptions:
  - In any step, the message holder *u* knows
    - The range of <u>local</u> contacts of all nodes
    - The location on the lattice of the target *t*
    - The locations and long-range contacts of all nodes that have previously touched the message
  - *u* does not know
    - the long-range contacts of nodes that have not touched the message

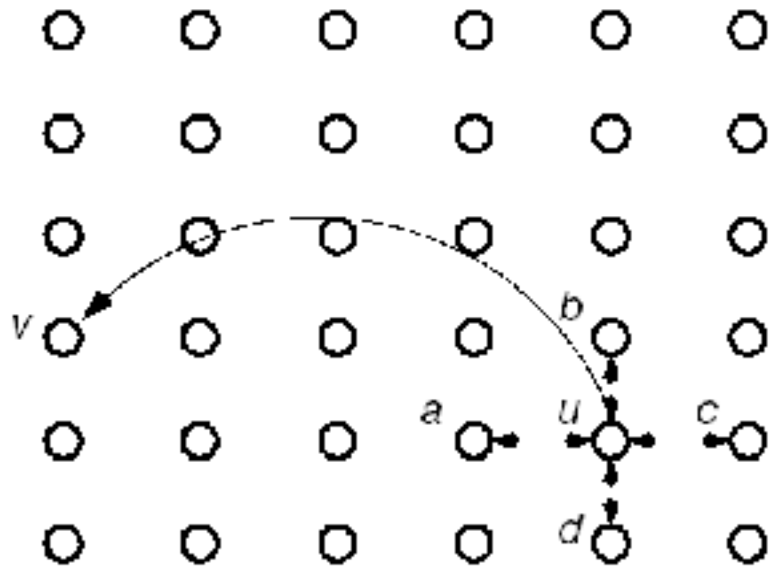


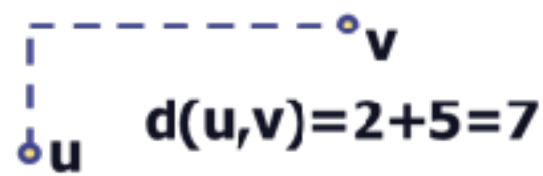
Kleinberg's model

- Consider a directed 2-dimensional lattice
- For each vertex u add q shortcuts
  - choose vertex v as the destination of the shortcut with probability proportional to $[d(u,v)]^{-r}$
  - when $r = 0$, we have uniform probabilities



Kleinberg's geographical small world model





nodes are placed on a lattice and connect to nearest neighbors

exponent that will determine navigability

additional links placed with

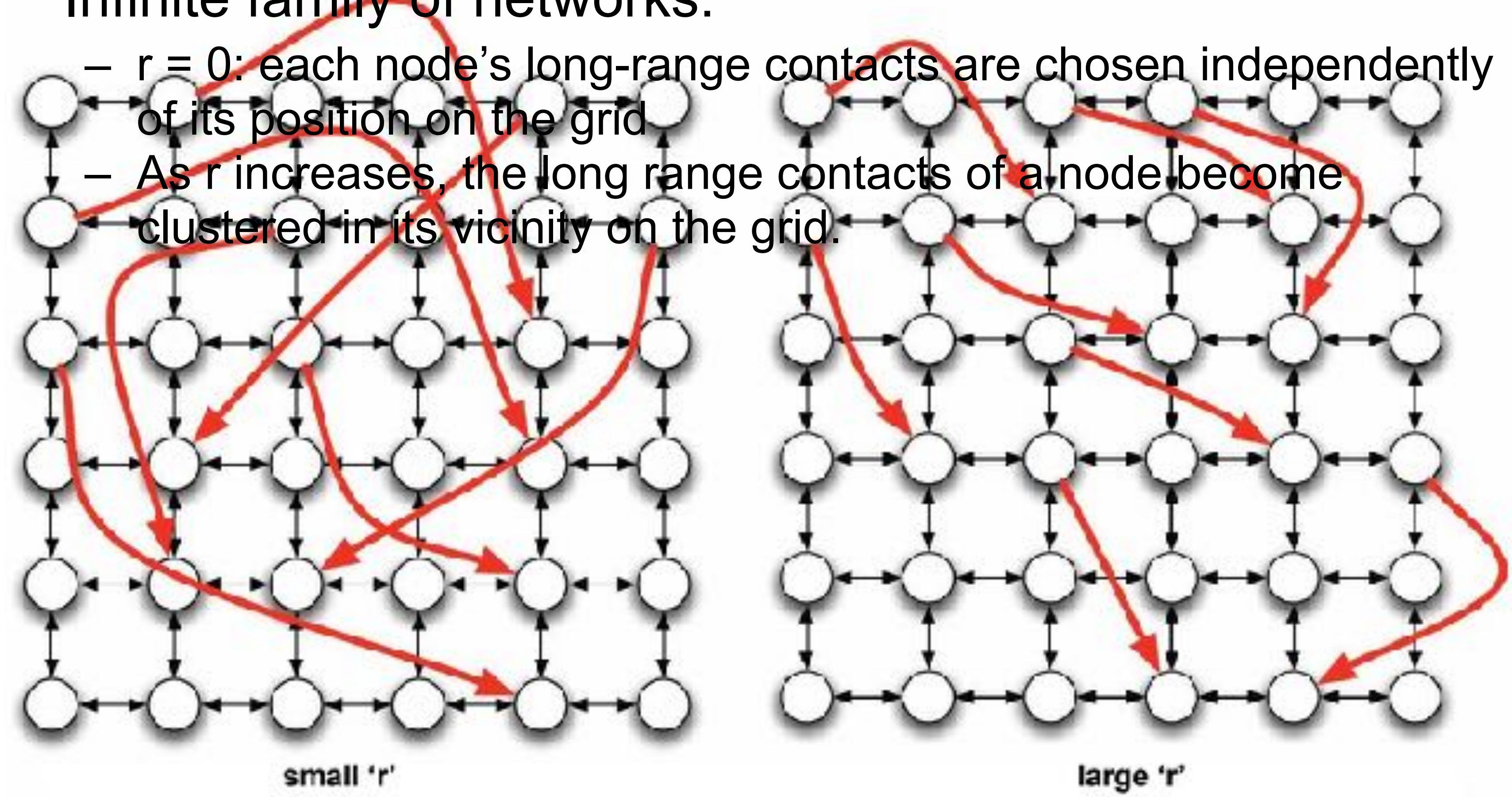
$$p(\text{link between u and v}) = (\text{distance}(u,v))^{-r}$$

Source: <u>Kleinberg, 'Navigation in a small world'</u>

Navigation in a small world network

r越大捷徑只會在
點附近出現

- Infinite family of networks:
  - r = 0: each node's long-range contacts are chosen independently of its position on the grid
  - As r increases, the long range contacts of a node become clustered in its vicinity on the grid.



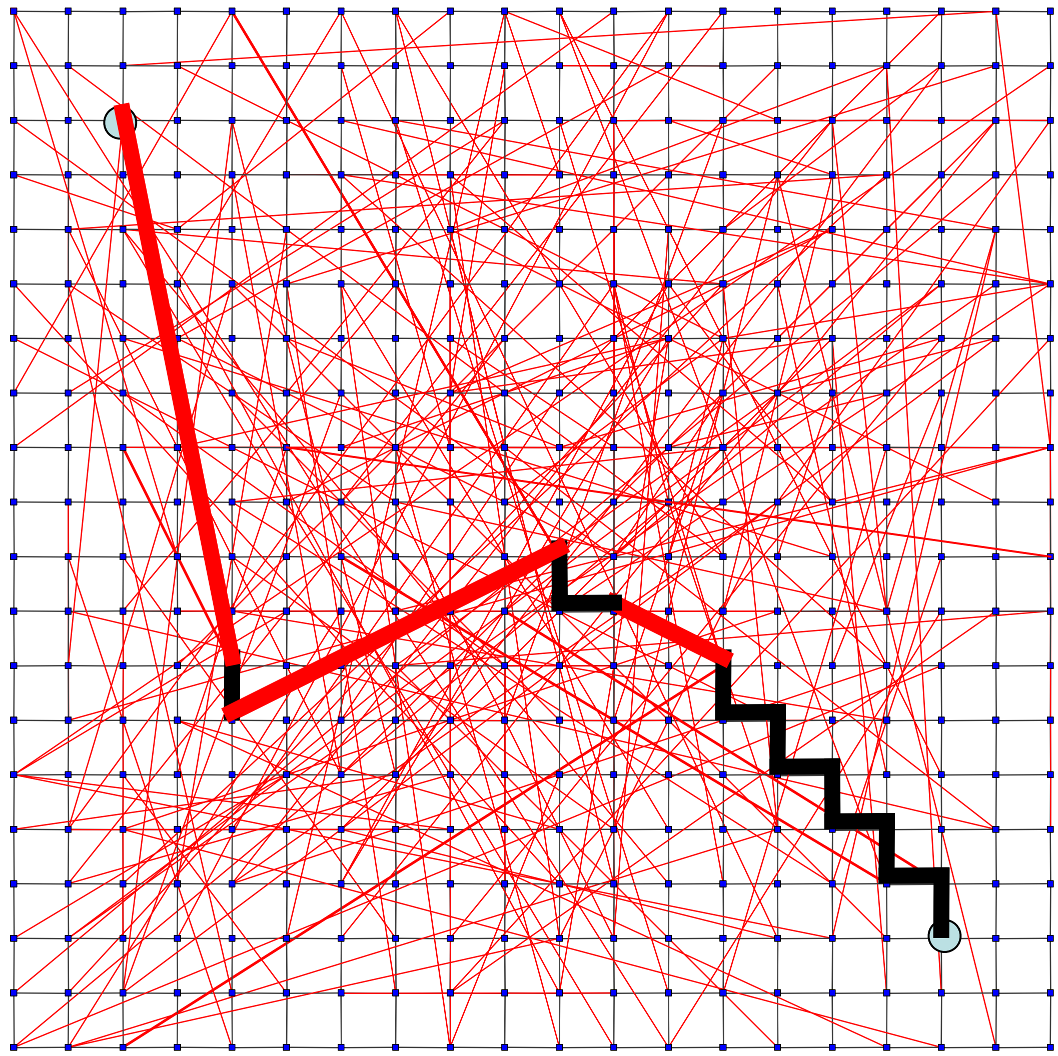
Searching in a small world

- Kleinberg proved the following
  - When $r=2$, an algorithm that uses only local information at each node can reach the destination in expected time $O(\log^2 n)$.
  - When $r<2$ a local greedy algorithm needs expected time $\Omega(n^{(2-r)/3})$.
  - When $r>2$ a local greedy algorithm needs expected time $\Omega(n^{(r-2)/(r-1)})$.
  - Generalizes for a d-dimensional lattice, when $r=d$

geographical search when network lacks locality

When **r=2**, links are randomly distributed, ASP ~ **log(n)**, n size of grid

When **r<2**, any decentralized algorithm is at least $\mathbf{a_0 n^{2/3}}$

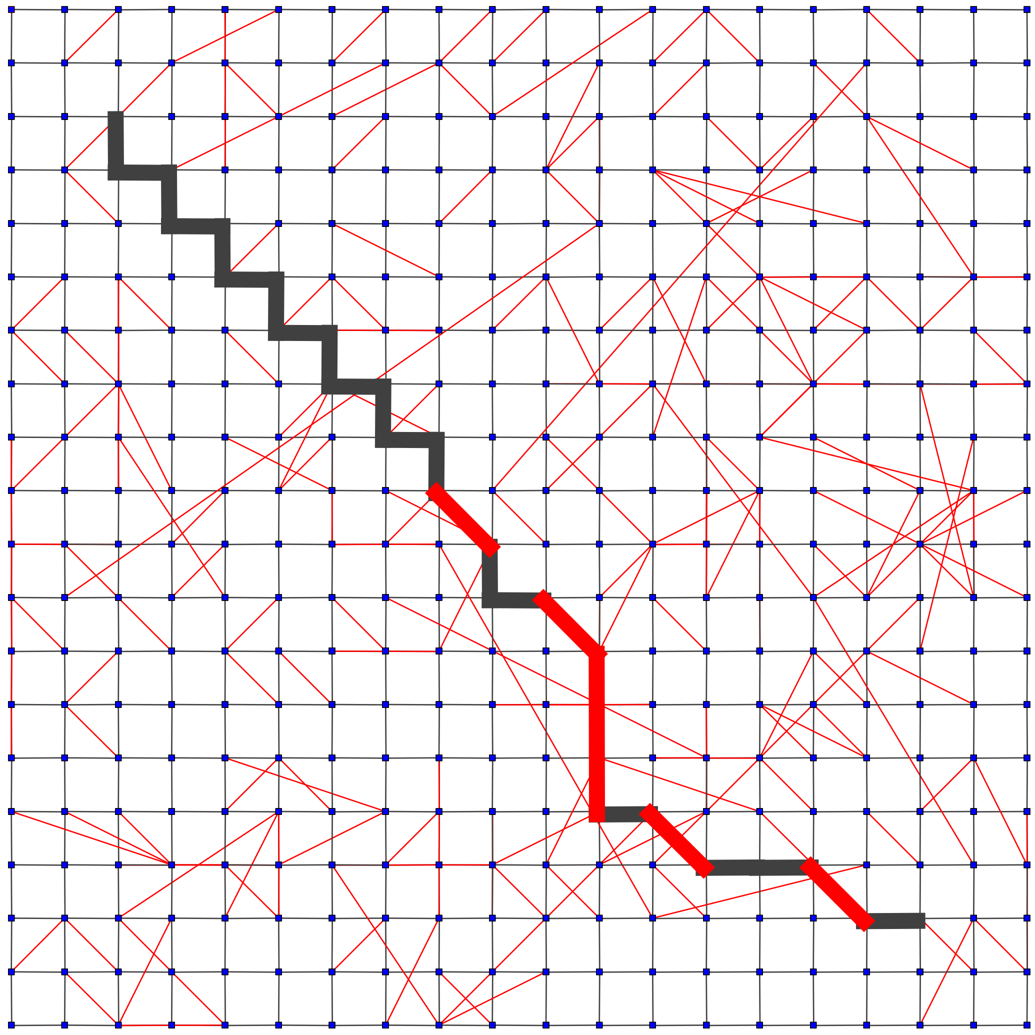


When **r<2**, expected time at least $\alpha_r n^{(2-r)/3}$

Overly localized links on a lattice

When **r>2** expected search time ~ $\mathbf{N}^{(\mathbf{r}-2)/(\mathbf{r}-1)}$

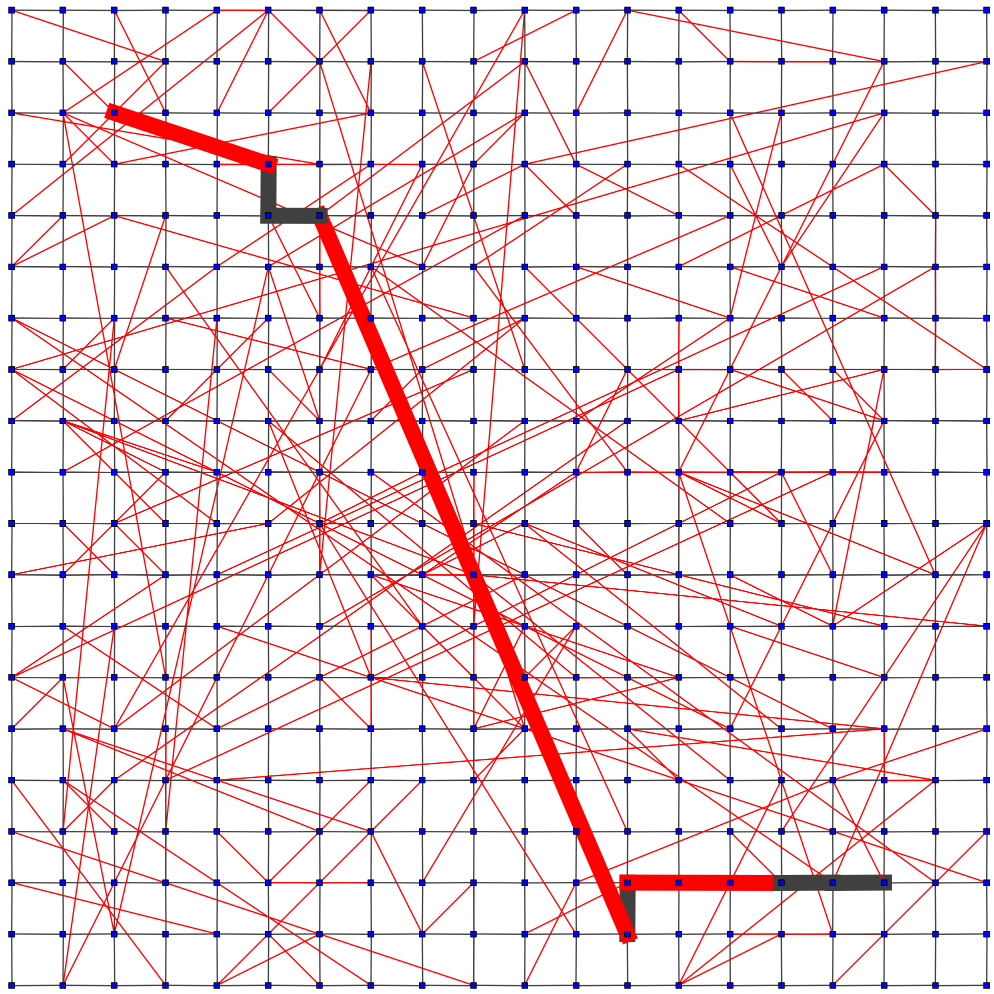
$$p \sim \frac{1}{d^4}$$



geographical small world model
Links balanced between long and short range

When **r=2,** expected time of a DA is at most **C (log N)²**

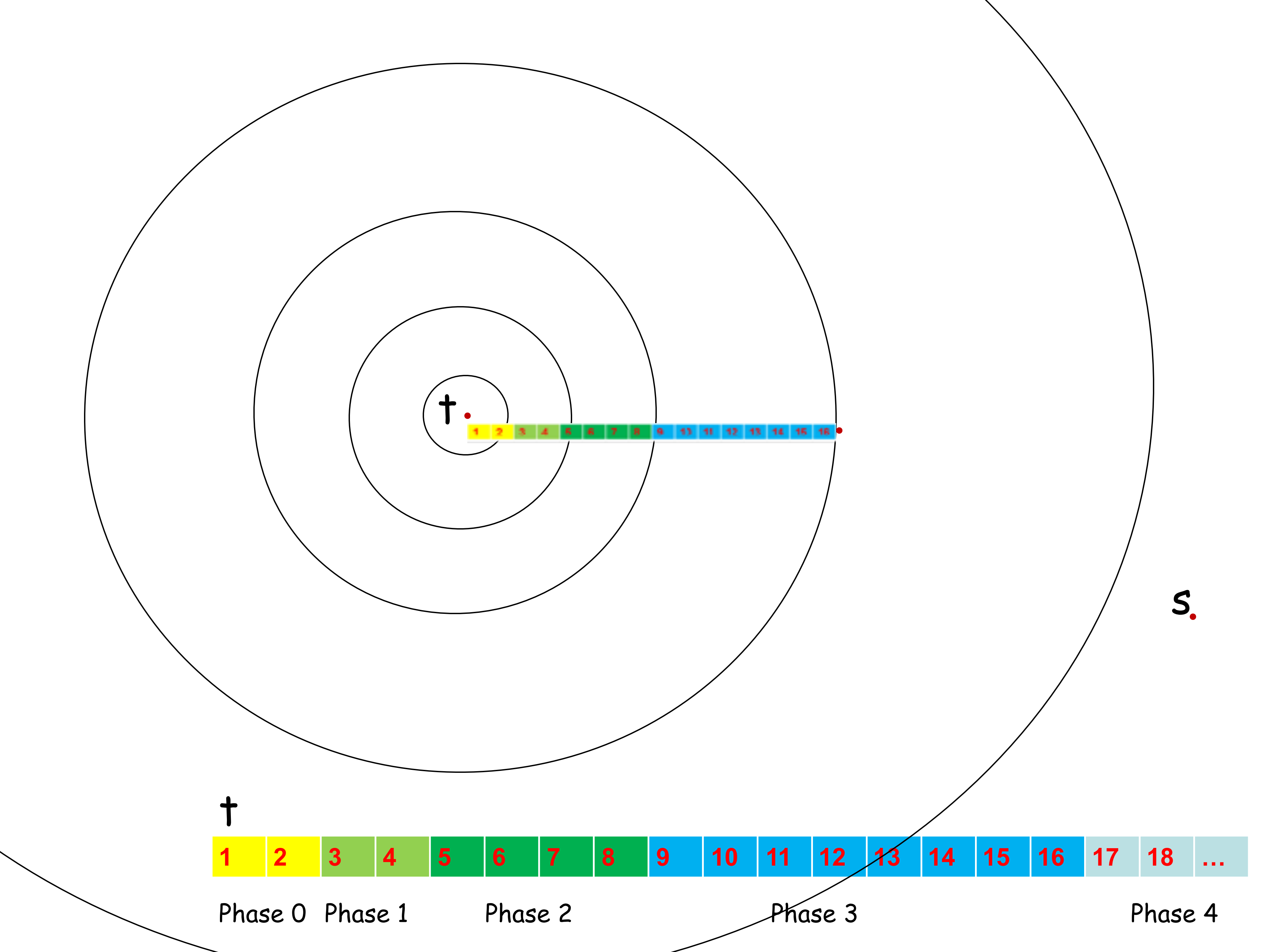
$$p \sim \frac{1}{d^2}$$



$R=2$

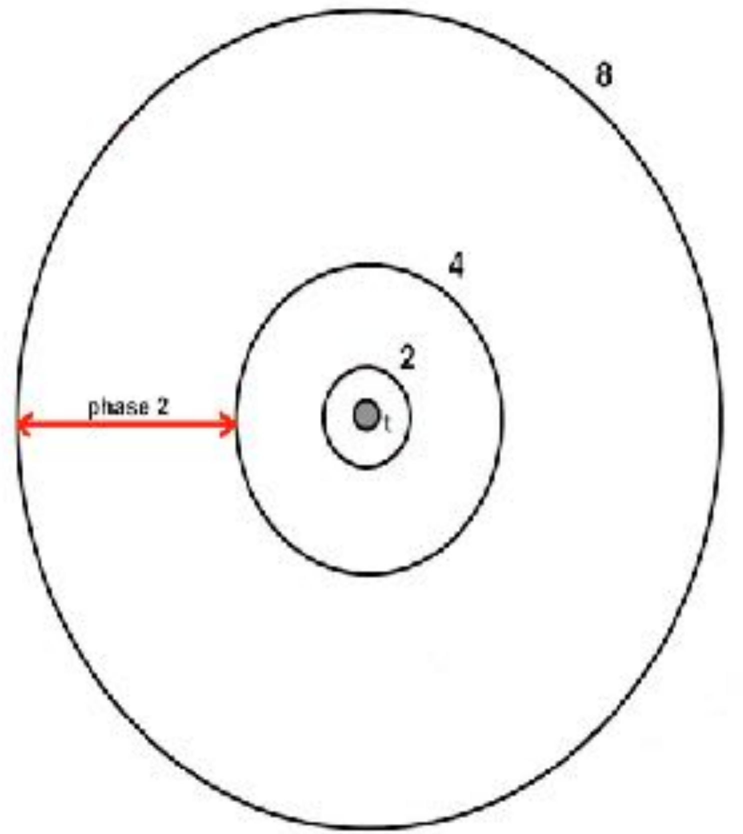
The Algorithm

- In each step the current message holder passes the message to the contact that is as close to the target as possible.


t
s
t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
...
Phase 0
Phase 1
Phase 2
Phase 3
Phase 4

Analysis

- Algorithm in phase *j*:
  - At a given step,
    $2^j < d(u,t) \leq 2^{j+1}$
  - Alg. is in phase 0 :
    - message is no more than 2 lattice steps away from the target t.
  - $j \leq \log_2 n$.



Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* in the next phase as its long range contact?

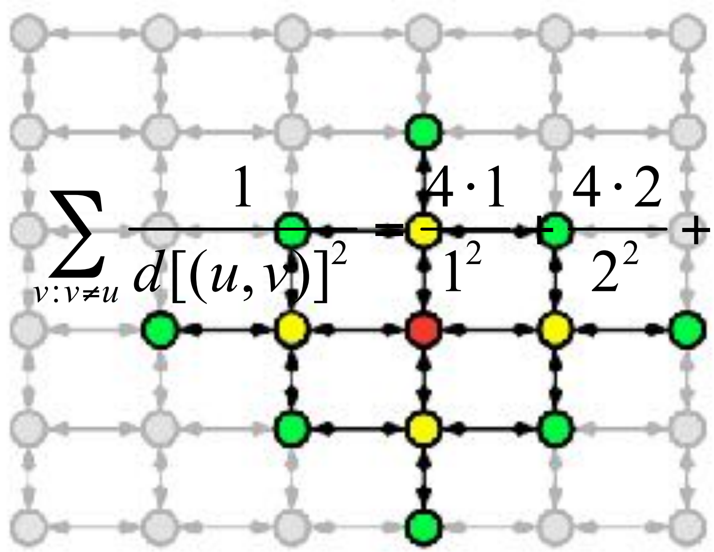
Analysis

- Pr [*u* has *v* as its long range contact] ?

$$= \frac{[d(u,v)]^{-2}}{\sum [d(u,v)]^{-2}}$$

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?



$$\sum_{v:v\neq u} \frac{1}{d[(u,v)]^2} = \frac{4\cdot 1}{1^2} + \frac{4\cdot 2}{2^2} + \frac{4\cdot 3}{3^2} \ldots \le \sum_{j=1}^{2n-2} \frac{4j}{j^2}$$

Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?

- Pr[*u* has *v* as its long range contact]?

$$\boxed{\sum_{v:v\neq u}[d(u,v)]^{-2} \leq \sum_{j=1}^{2n-2}\frac{4j}{j^2} = 4\sum_{j=1}^{2n-2} 1/j \leq 4[1+\ln(2n-2)] \leq 4\ln(6n)}$$

- Thus *u* has *v* as its long-range contact with probability

$$\geq \frac{[d(u,v)]^{-2}}{4\ln(6n)}$$

$$\geq \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?

$$\geq \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

- In any given step, Pr[phase j ends in this step]?
  - Phase j ends in this step if the message enters the set B_j of nodes within distance 2^j of t. Let v_f be the node in B_j that is farthest from *u*.

$$\Pr[\text{phase } j \text{ ends in this step}] = \sum_{v\in B_j} \Pr\left[u\, is\; friends\; with\; v \in B_j\right]$$

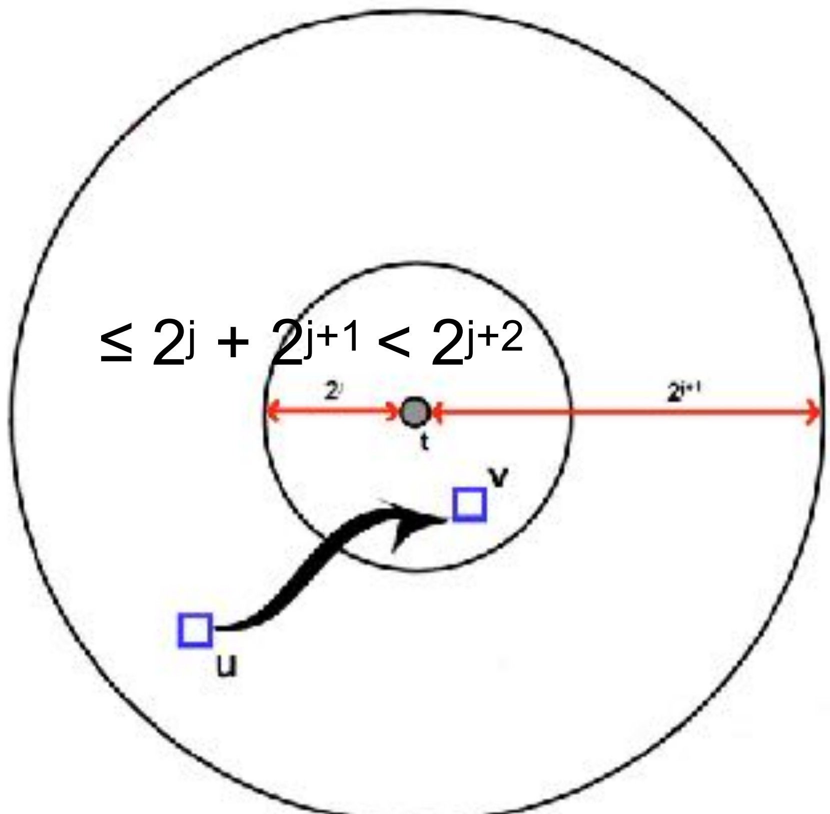
$$\geq |B_j| \cdot \left(\frac{1}{4\ln(6n)\cdot[d(u,v_f)]^2}\right)$$

Analysis

Questions:
- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?

$$\geq \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

- $\Pr[\text{phase } j \text{ ends in this step}] \geq |B_j| \cdot \left(\frac{1}{4\ln(6n)\cdot[d(u,v_f)]^2}\right)$
  - What is d[(u,v_f)]?



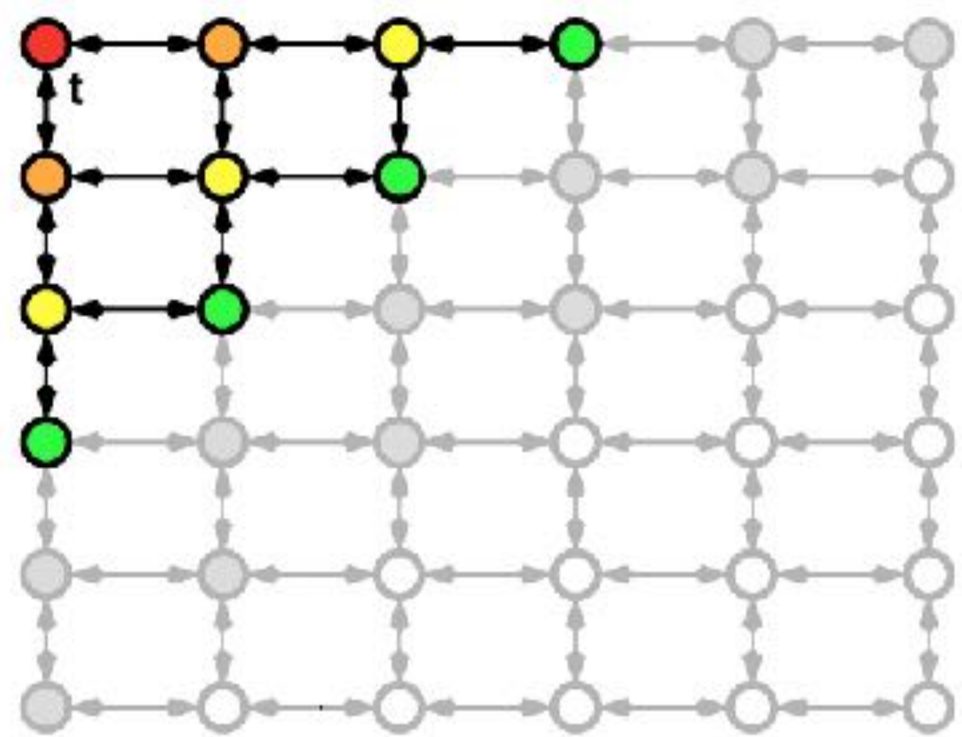
Analysis

- $$\Pr[\text{phase } j \text{ ends in this step}] \geq |B_j| \cdot \left(\frac{1}{4\ln(6n) \cdot 2^{2j+4}} \right)$$
- How many nodes are in B_j?

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?

$$\geq \frac{1}{4\ln(6n) \cdot [d(u,v)]^2}$$



Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?
- What is the probability that node *u* has a node *v* as its long range contact?

$$\geq \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

- In any given step, Pr[phase j ends in this step]?
  - Pr[*u* has a long-range contact in B_j]?

$$\geq \#\ of\ nodes\ in\ B_j \cdot (probability\ u\ is\ friends\ with\ farthest\ v \in B_j)$$

$$\geq 2^{2j-1}\left(\frac{1}{4\ln(6n)\cdot 2^{2j+4}}\right) = \frac{2^{2j-1}}{4\ln(6n)\cdot 2^{2j+4}} = \frac{1}{128\ln(6n)}$$

Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?

$$\geq \frac{1}{128 \ln(6n)}$$

- What is the probability that node *u* has a node *v* as its long range contact?

$$\geq \frac{1}{4 \ln(6n) \cdot [d(u,v)]^2}$$

- How many steps will we spend in phase *j*?
  - Let X_j be a random variable denoting the number of steps spent in phase *j*.
    - X_j is a random variable with a probability of success at least

Analysis

- How many steps will we spend in phase *j*?
  - Since X_j is a geometric random variable, we know that

$$E[X_j] = \frac{1}{p} \le \frac{1}{1/128\ln(6n)} = 128\ln(6n)$$

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?
- In a given step, with what probability will phase *j* end in this step?

$$\ge \frac{1}{128\ln(6n)}$$

- What is the probability that node *u* has a node *v* as its long range contact?

$$\ge \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$$

Analysis

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?

  $\leq 128\ln(6n)$

- In a given step, with what probability will phase *j* end in this step?

  $\geq \frac{1}{128\ln(6n)}$

- What is the probability that node *u* has a node *v* as its long range contact? $\geq \frac{1}{4\ln(6n)\cdot[d(u,v)]^2}$

- How many steps does the algorithm take?
  - Let X be a random variable denoting the number of steps taken by the algorithm.
  - By Linearity of Expectation we have

$$E[X] \leq (1+\log n)(128\ln(6n)) = O(\log n)^2$$

Analysis

- When r = 2, expected delivery time is

$$O(\log n)^2$$

Questions:

- How many steps will the algorithm take?
- How many steps will we spend in phase *j*?

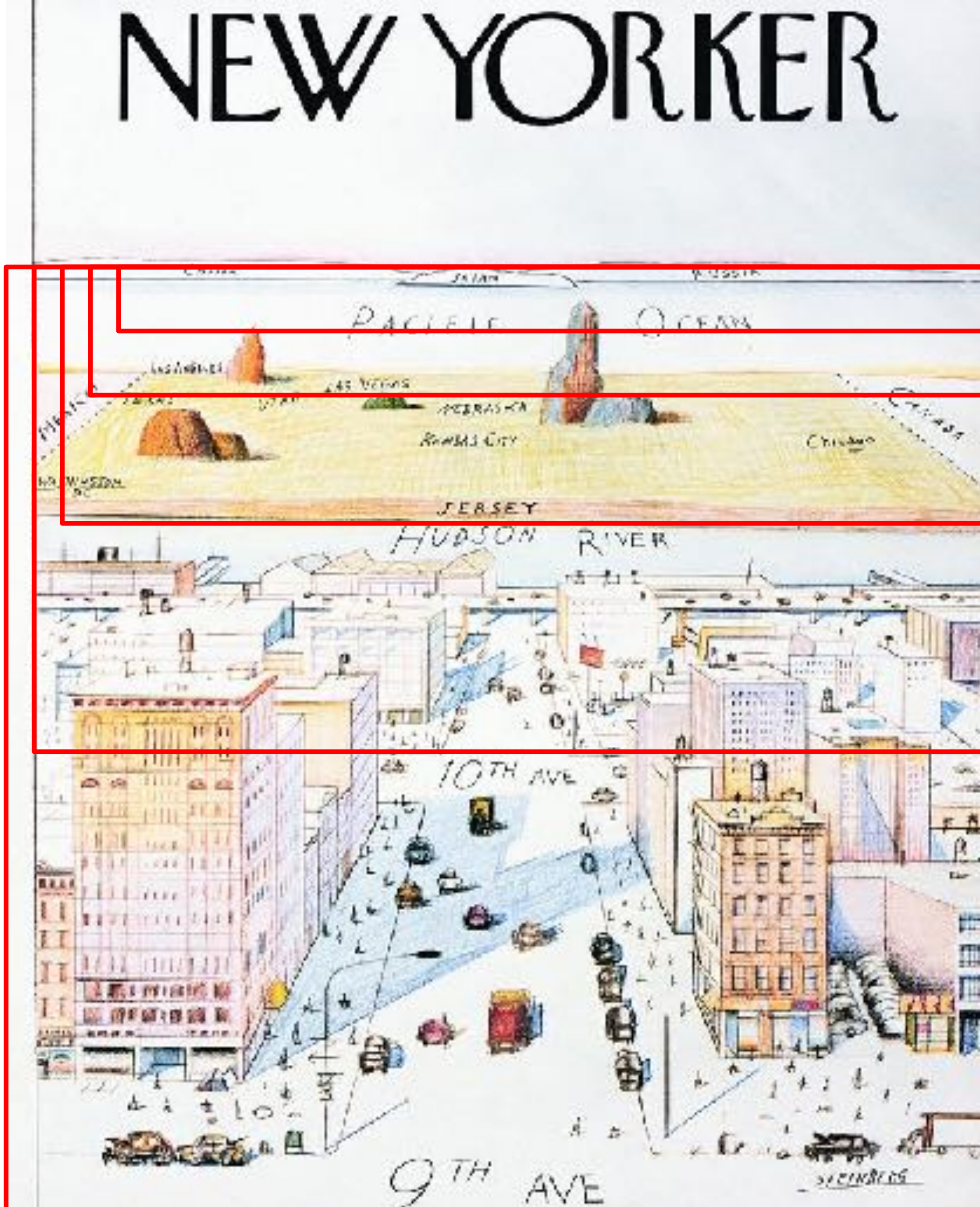
  $\leq 128 \ln(6n)$

- In a given step, with what probability will phase *j* end in this step?

  $\geq \frac{1}{128 \ln(6n)}$

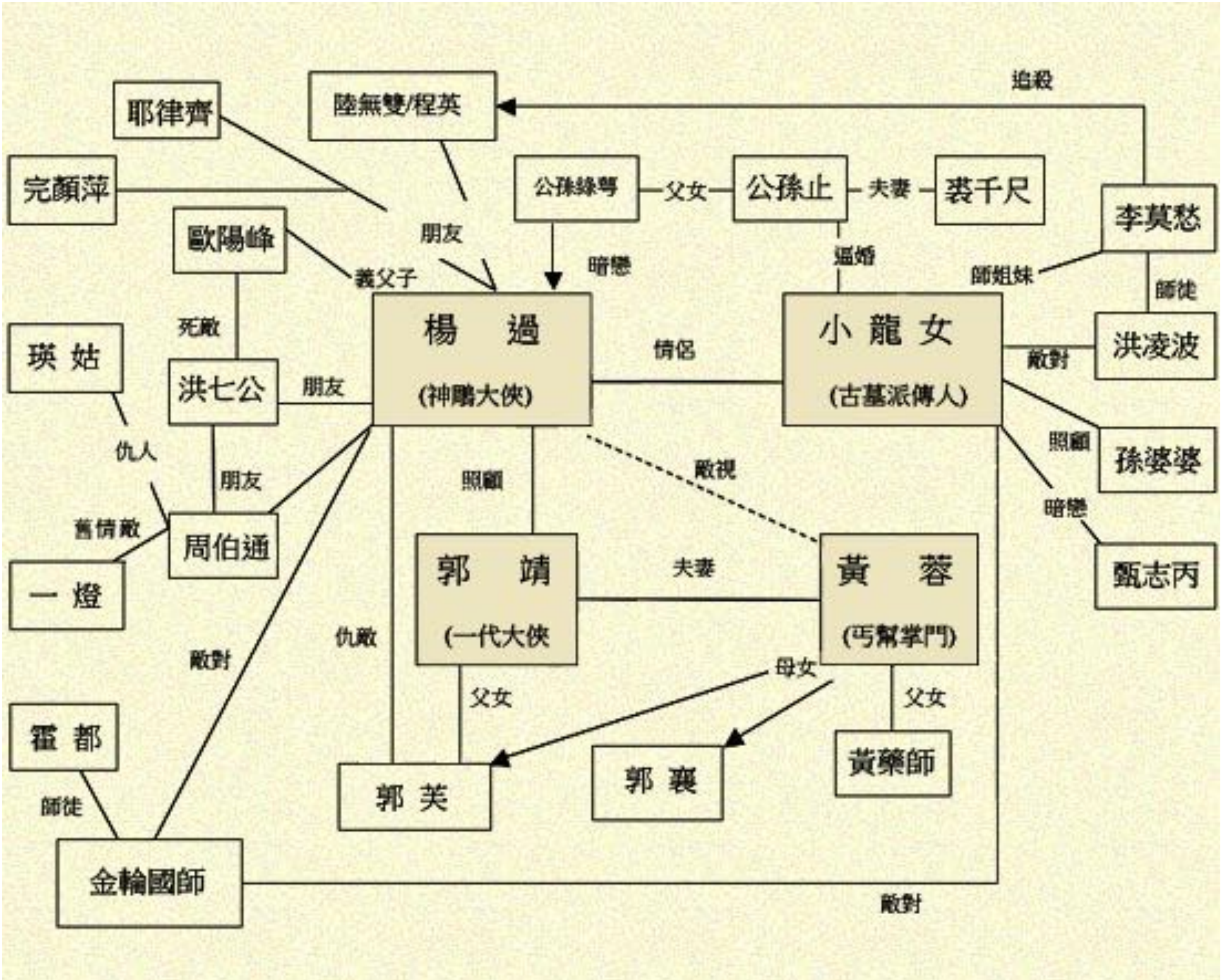
- What is the probability that node *u* has a node *v* as its long range contact?

  $\geq \frac{1}{4 \ln(6n) \cdot [d(u,v)]^2}$

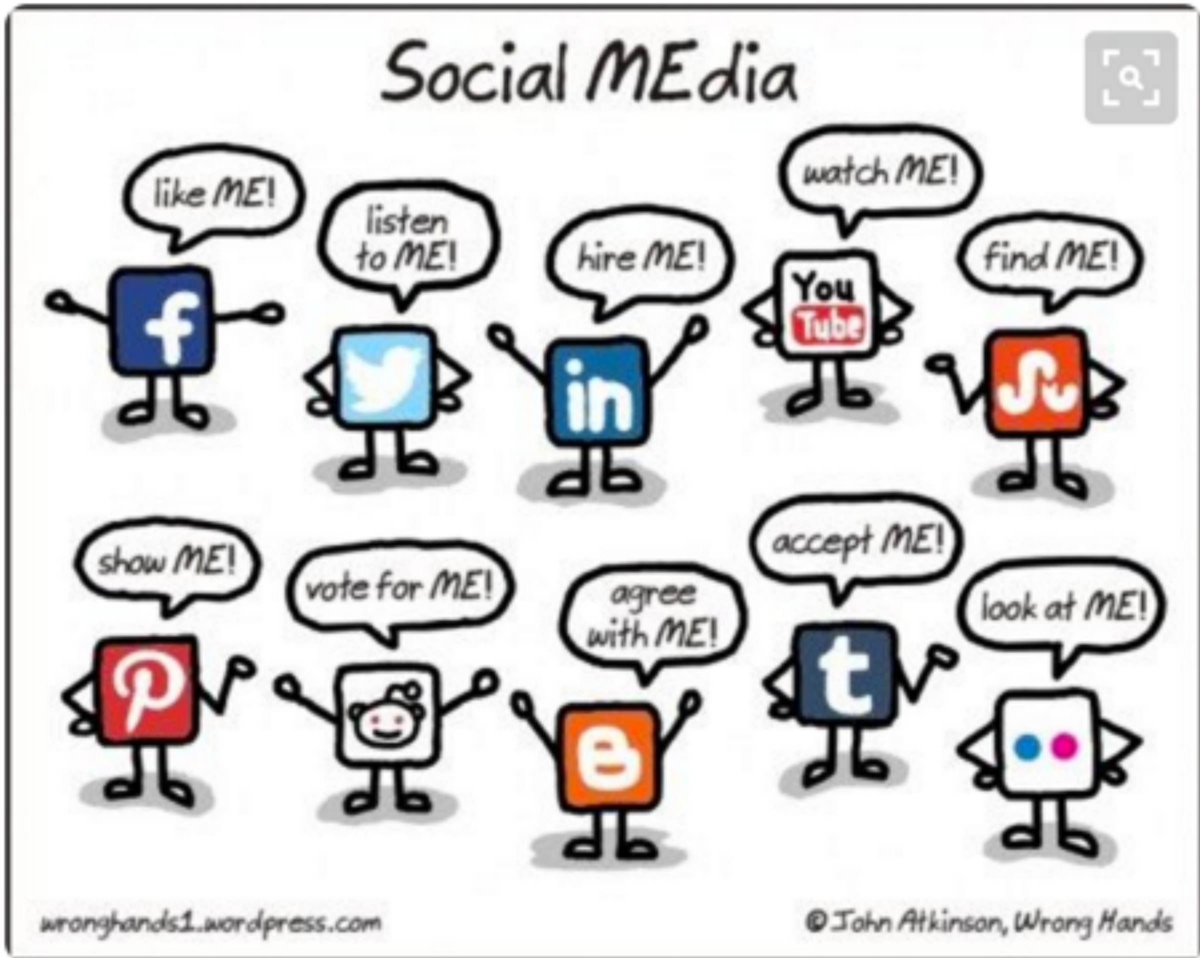

NEW YORKER
RUSSIA
PACIFIC OCEAN
LAS VEGAS
NEBRASKA
KANSAS CITY
CHICAGO
CANADA
JERSEY
HUDSON RIVER
10TH AVE
9TH AVE
STEINBERG

Assignment 2: Social Network Visualization

- Learn to Use **Social Network Visualization** Tool
- Draw a picture for the social network of the characters in the following novel:
  - 天龍八部
  - 笑傲江湖
  - 絕代雙驕
  - 神鵰俠侶+射鵰英雄傳
  - 倚天屠龍記
  - 三國演義
- A node stands for a character
- An edge between two characters stands for the relationship that the characters co-appear in a paragraph
- **Web based Presentation Only! Deadline 3/26**



追殺
耶律齊
陸無雙/程英
完顏萍
公孫綠萼
父女
公孫止
夫妻
裘千尺
李莫愁
朋友
歐陽峰
暗戀
逼婚
師姐妹
義父子
師徒
死敵
楊 過
(神鵰大俠)
情侶
小 龍 女
(古墓派傳人)
洪凌波
瑛 姑
洪七公
朋友
敵對
照顧
孫婆婆
仇人
朋友
照顧
敵視
暗戀
舊情敵
周伯通
郭 靖
(一代大俠)
夫妻
黃 蓉
(丐幫掌門)
甄志丙
一 燈
仇敵
敵對
母女
父女
父女
霍 都
師徒
郭 芙
郭 襄
黃藥師
金輪國師
敵對


Social MEdia
like ME!
listen to ME!
hire ME!
watch ME!
You Tube
find ME!
show ME!
vote for ME!
agree with ME!
accept ME!
look at ME!
wronghands1.wordpress.com
© John Atkinson, Wrong Hands