Peer-to-Peer and Social Networks

Small World Graphs

The small-world model

[**Watts and Strogatz** (1998)]

They followed up on Milgram's work and reason about why there is a small degree of separation between individuals in a social network. Research originally inspired by Watt's efforts to understand the synchronization of cricket chirps, which show a high degree of coordination over long ranges, as though the insects are being guided by an invisible conductor.

Disease spreads faster over a small-world network.

Questions not answered by Milgram

Why **six** degrees of separation? Any scientific reason? What properties do these social graphs have?

Is clustering the only missing link? (Human beings prefer clustered environments). But the diameter must also be low!

Time to reverse engineer this.

What are small-world graphs

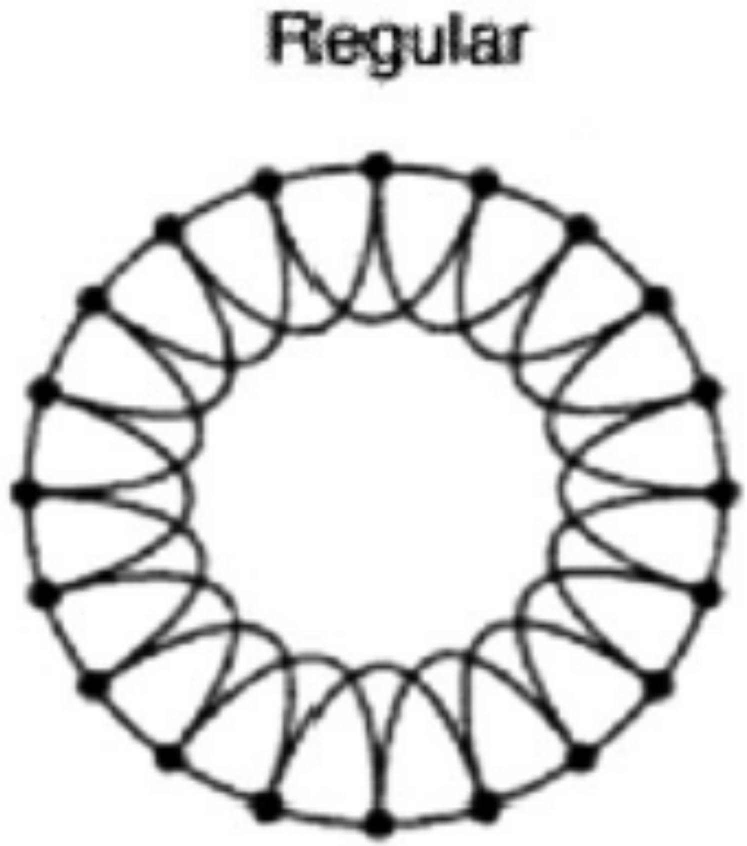
Completely regular

Small-world graphs ($n \gg k > \log n$)

Completely random

n = number of nodes, k= number of neighbors of each node

Completely regular



A ring lattice

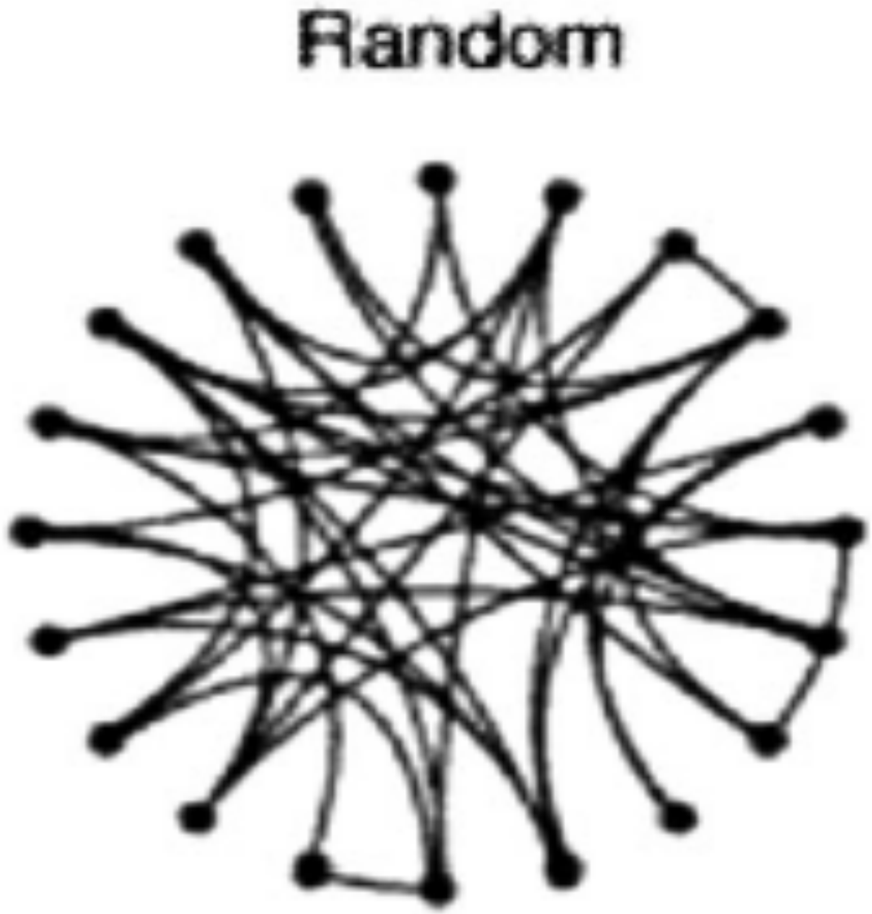
If $k = 4$ then

$$Clustering\ coefficient\ CC = 3/6 = 1/2$$

$$Diameter\ L = \frac{n}{2k}$$

The clustering coefficient is OK, but Diameter is too large!

Completely random



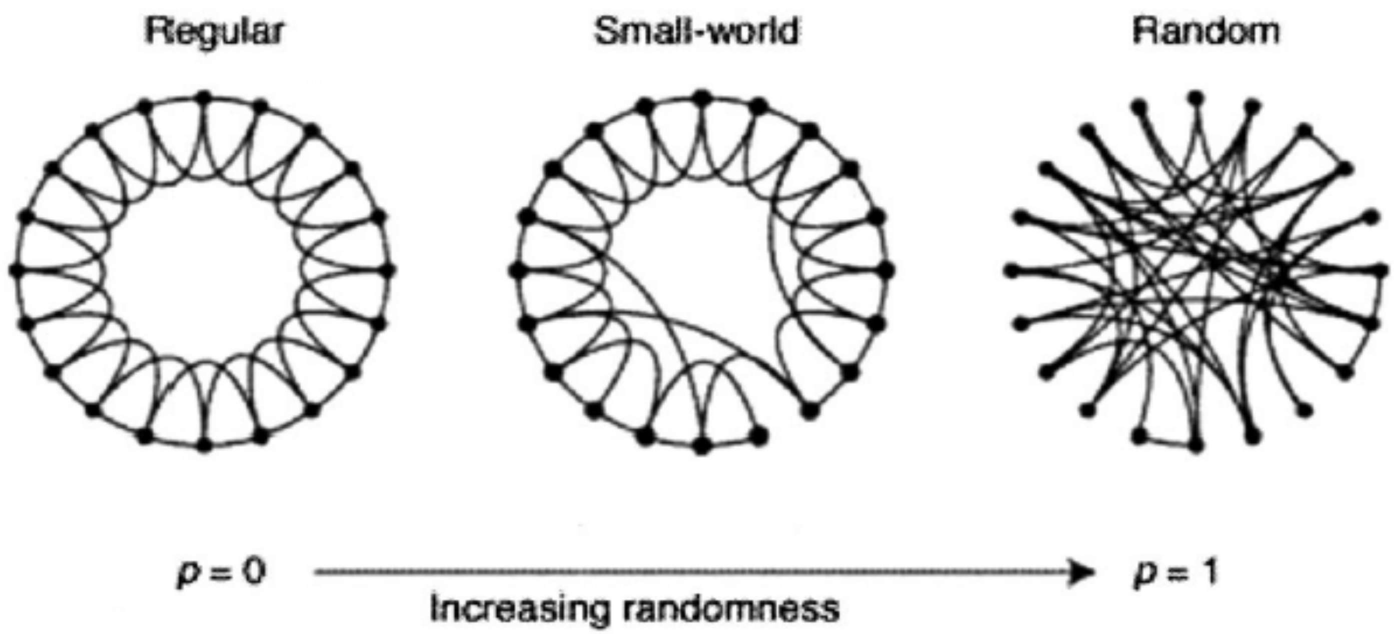
$$CC = p \simeq \frac{k}{n}$$

$$L \simeq \log_k n$$

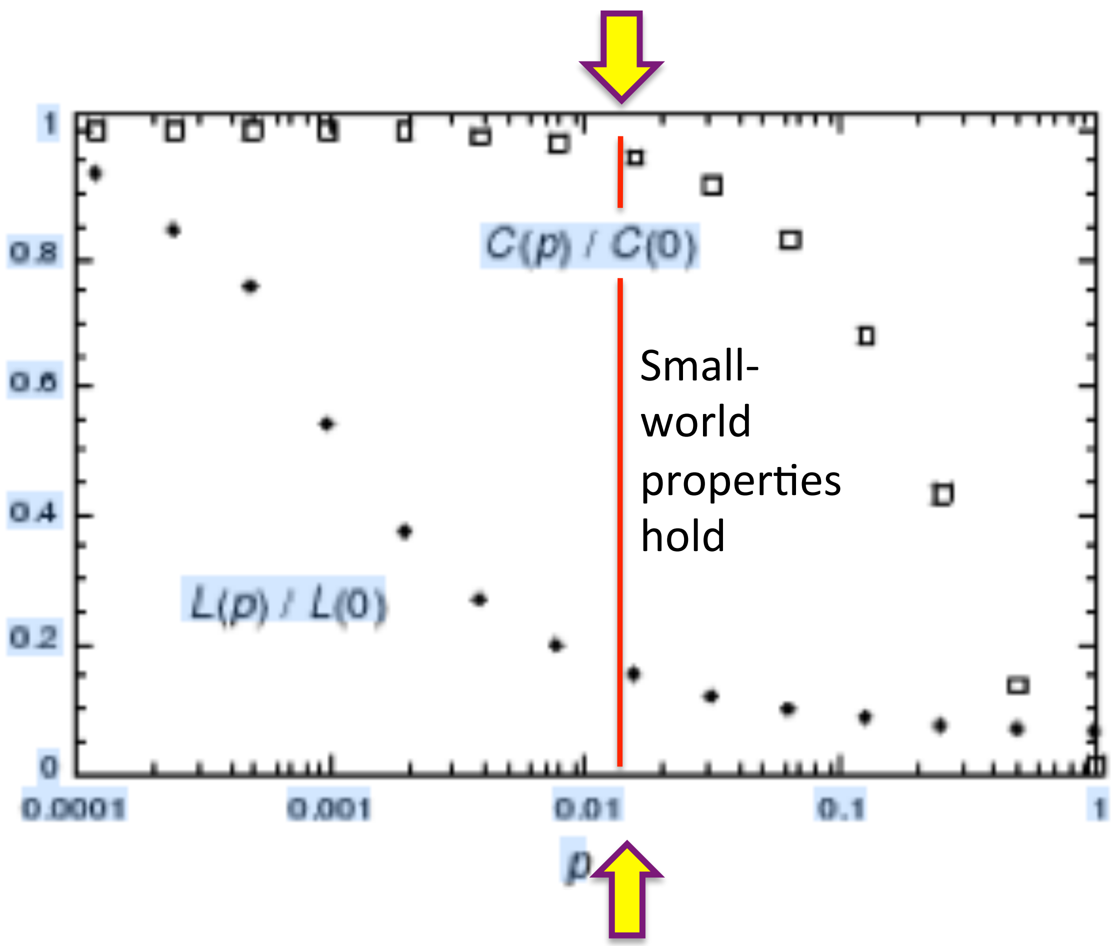
Diameter is small, but the
Clustering coefficient is too small!

Small-world graphs

Start with the regular graph, and with probability *p* **rewire** each link to a randomly selected node. It results in a graph that has high clustering coefficient but low diameter …



Small-world graphs



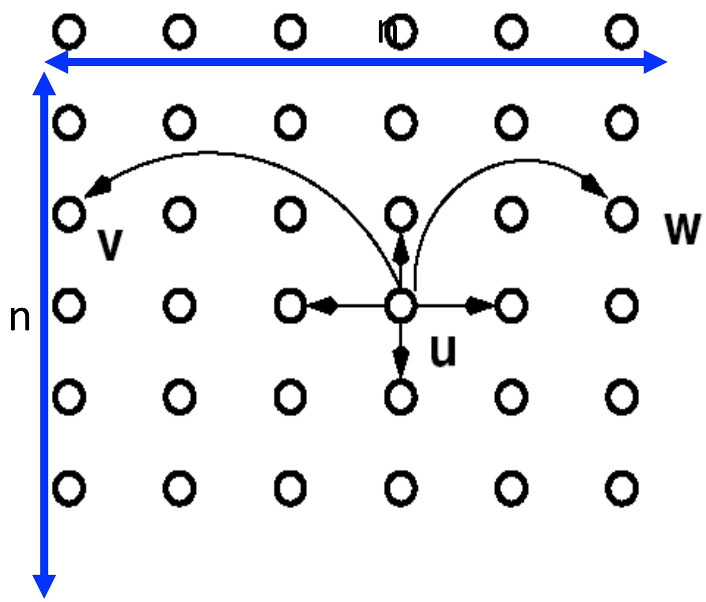
Limitation of Watts-Strogatz model

Jon Kleinberg argues …

Watts-Strogatz small-world model illustrates the existence of short paths between pairs of nodes. But it does not give any clue about how those short paths will be discovered. A greedy search for the destination will not lead to the discovery of these short paths.

Kleinberg's Small-World Model

Consider an $(n \times n)$ grid. Each node has a link to every node at lattice distance p (short range neighbors) & q long range links. Choose long-range links at lattice distance d with a probability proportional to d^{-r} (**See note below)



p = 1, q = 2

r = 2

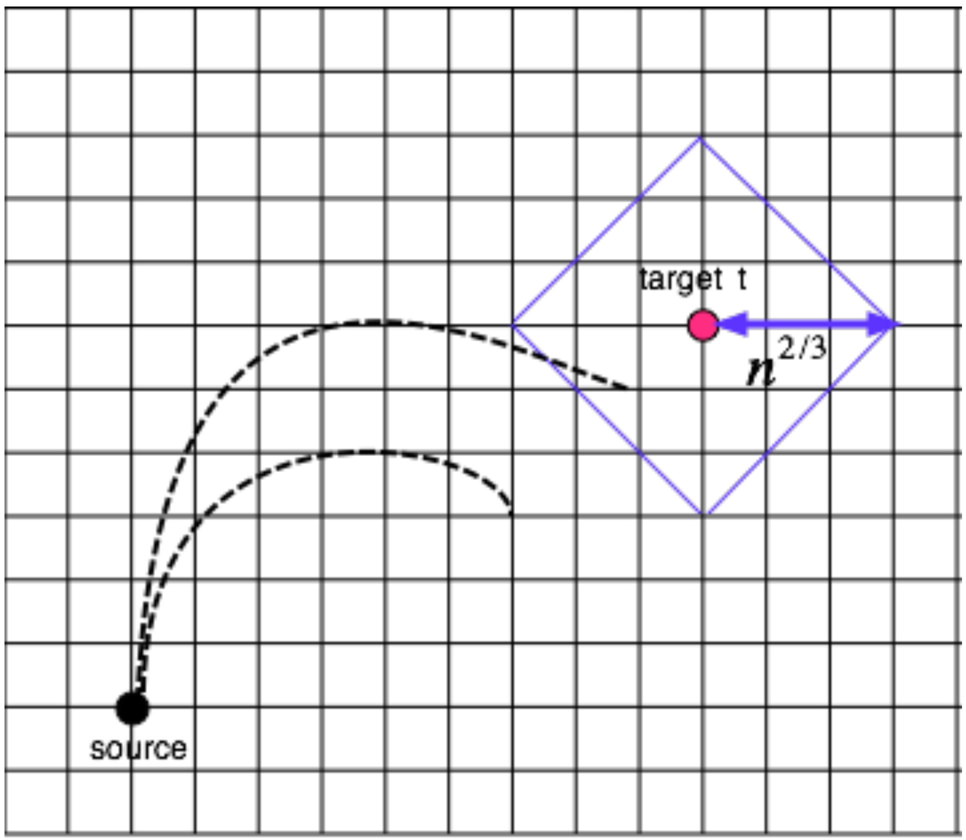
**Here r denotes the dimension of the space. Since we are considering a 2D grid, r=2

Results

Theorem 1. There is a constant $\alpha 0$ (depending on p and q but independent of n), so that when $r = 0$, the expected delivery time of any decentralized algorithm is at least $\alpha 0.\ n^{2/3}$

** The above result is valid for a 2D grid only. For a 1D space like a Linear topology of a ring, the expected time will be different

Proof of theorem 1



$n \times n$

Probability to reach within a lattice distance $n^{2/3}$ from the target is

$$\frac{2.n^{4/3}}{n^2} = 2.n^{-2/3}$$

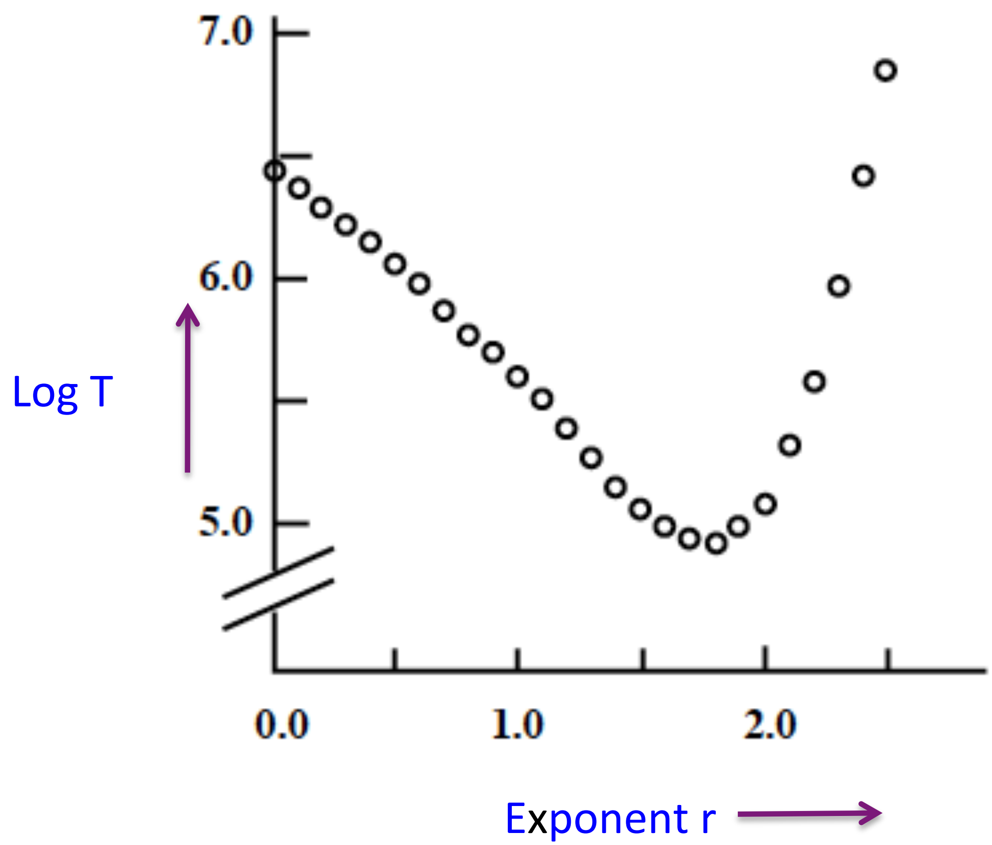
So, it will take an expected $O(n^{2/3})$ steps to reach the target.

More results

Theorem 2. There is a decentralized algorithm A and a constant $\alpha 2$ (dependent on p and q) but independent of n, such that when r=2 and $p = q = 1$, the expected delivery time of A is at most $\alpha 2 . \log^2 n$

For a one-dimensional search space, the same result will hold for $(r = 1)$ i.e the expected delivery time is $O(\log^2 n)$ when long-range links at distance d are chosen with probability proportional to d^{-1}

Variation of search time with r



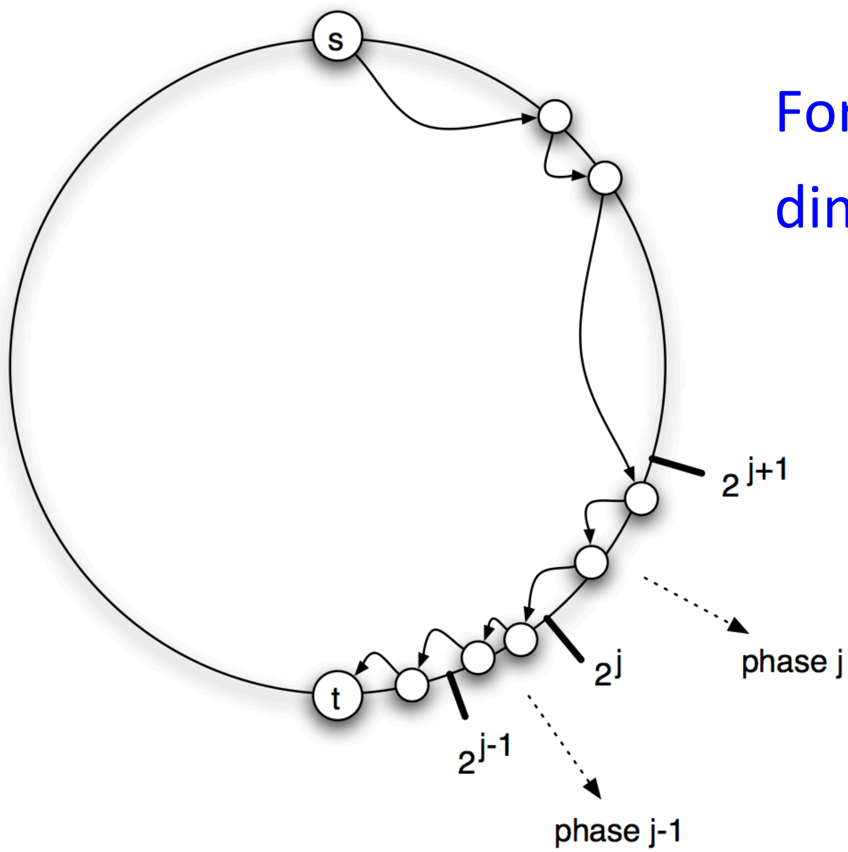
This is for a 2D topology

Proof of Theorem 2

Main idea.

We show that **in phase j**, the expected time before the current message holder has a long-range contact within lattice distance 2^j from t is O(log n); at this point, phase j will come to an end. As there are at most **log n phases**, a bound proportional to $\log^2 n$ follows.

Proof of Kleinberg's theorem



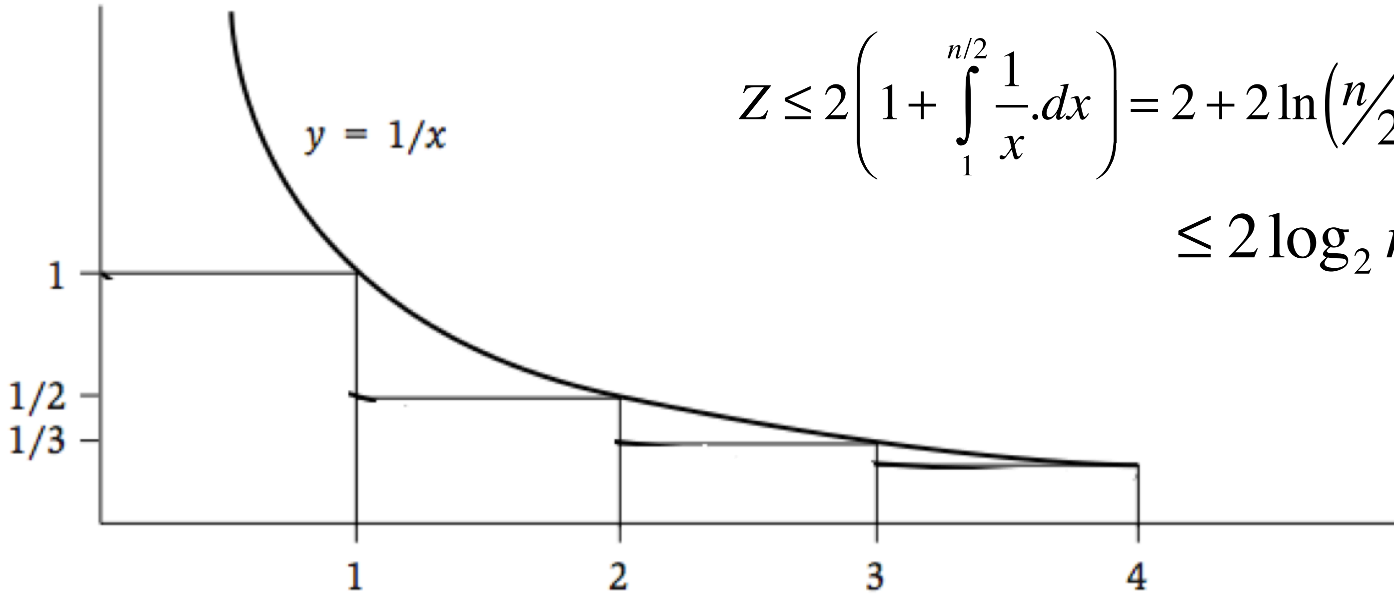
For simplicity we prove it for a one dimensional ring topology, so r =1

Probability (u linking to v) = $\frac{1}{Z} \cdot \frac{1}{d(v,u)}$

Since $\sum_{d(u,v)=1}^{d(u,v)=n/2} \frac{1}{Z} \cdot \frac{1}{d(v,u)} \geq 1/2$

$$Z \leq 2\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots+\frac{1}{n/2}\right).$$

Proof continued



$$Z \le 2\left(1 + \int_1^{n/2} \frac{1}{x}.dx\right) = 2 + 2\ln\left(n/2\right)$$

$$\le 2\log_2 n$$

An upper bound of the normalizing constant is the area under the curve which is $Z \le 2\log n$

Proof continued

Thus, probability that a link (v,w) exists is = $\frac{1}{Z} d(v,w)^{-1} \geq \frac{1}{\log n} d(v,w)^{-1}$

We now calculate the time taken in **one phase** (implies that the distance to the **target becomes less than d/2**.

Probability in one step the search reaches a given node in the target zone ≥

$$\frac{1}{\log n} d(v,w)^{-1} \geq \frac{1}{\log n} \cdot \frac{1}{3d/2} = \frac{2}{3d \log n}$$

↑ Why?

Probability that in one step the search reaches some node within distance d/2 ≥

$$d \cdot \frac{2}{3d \log n} = \frac{2}{3 \log n}$$

Proof continued

How can this continue? Let X_j be the number of steps in phase j

The probability that this phase continues for at least i steps ≤

$$\Pr[X_j \geq i] \leq \left(1 - \frac{2}{3\log n}\right)^{i-1}$$

The expected number of steps to complete phase j is

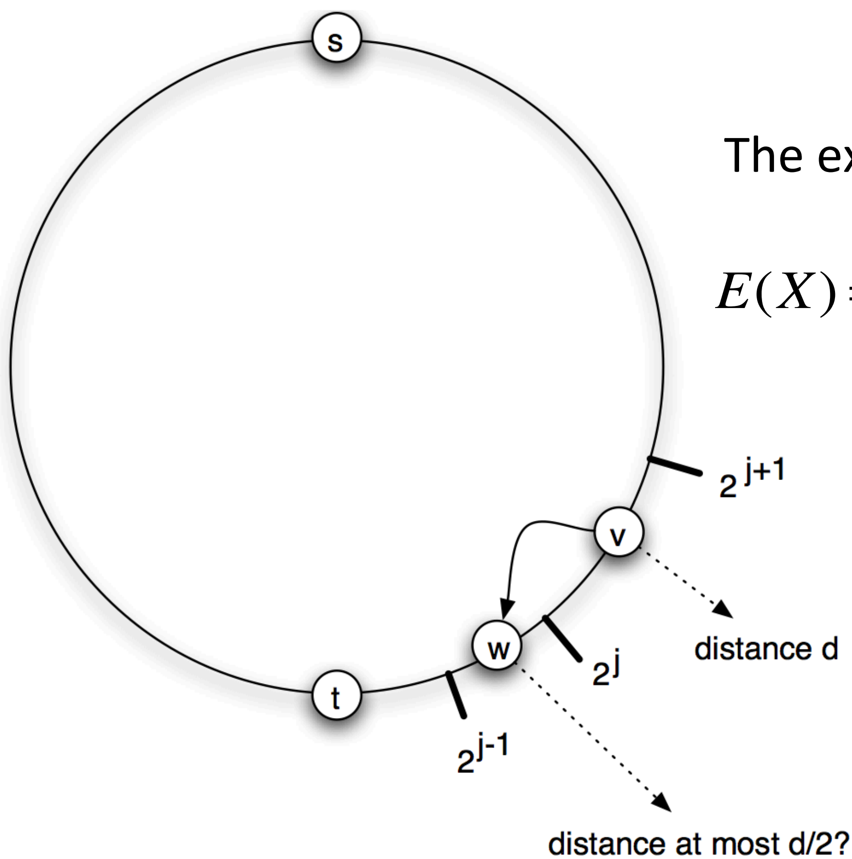
$$E[X_j] = 1 \cdot \Pr[X_j = 1] + 2 \cdot \Pr[X_j = 2] + 3 \cdot \Pr[X_j = 3] + \cdots$$

$$= \Pr[X_j \geq 1] + \Pr[X_j \geq 2] + \Pr[X_j \geq 3] + \cdots$$

So, $$E[X_j] \leq 1 + \left(1 - \frac{2}{3\log n}\right) + \left(1 - \frac{2}{3\log n}\right)^2 + \left(1 - \frac{2}{3\log n}\right)^3 + \cdots$$

This leads to $$E[X_j] \leq \frac{3}{2}\log n.$$

Proof continued



The expected number of steps for the total search

$$E(X) = E(X_1) + E(X_2) + \ldots + E(X_{\log n}) \leq \frac{3}{2}(\log n)^2$$

$$i.e. \quad O\left(\log^2 n\right)$$

□