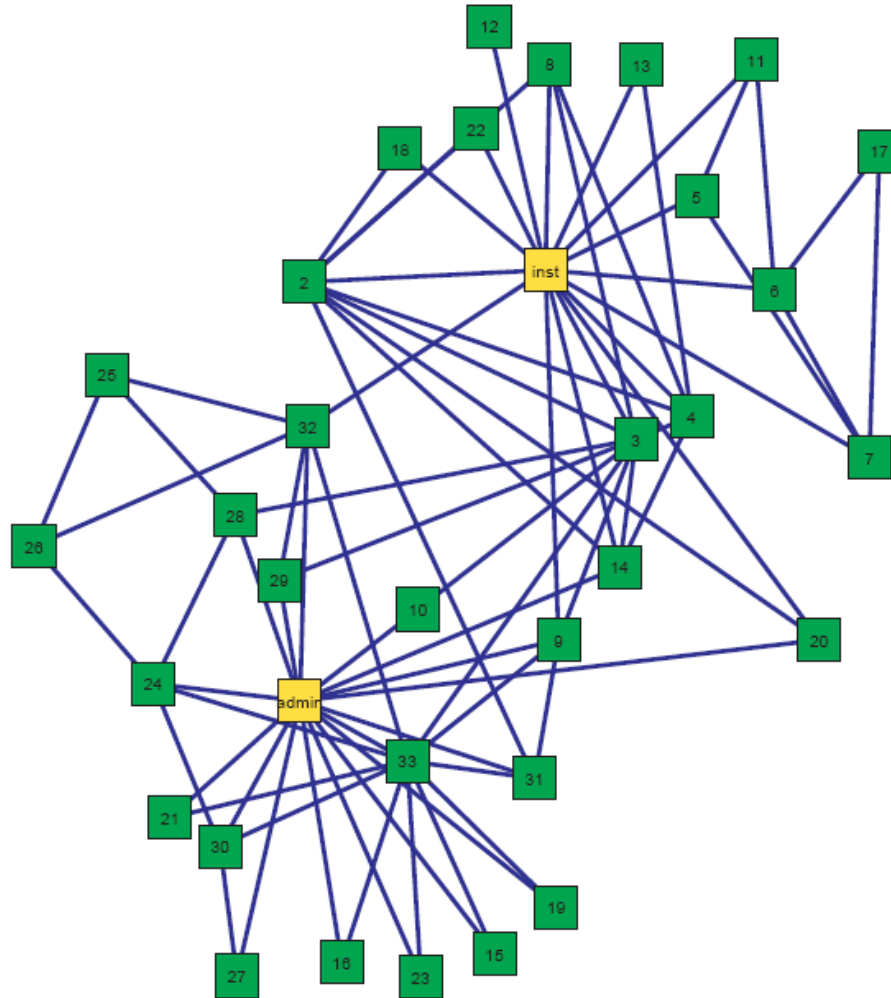


Distributed Computing over  
Communication Networks:

# Social Networks

# Social Relationships in Karate Club (1970)

---



Observations?  
Dispute caused split:  
guess where!

# Small World Networks: Phenomena, Vision, and Science

---

**Karinthy (1929): „World is shrinking as humans are more connected!“**  
(inspired by **Marconi?**)

## **Hot topic in 60s:**

- McLuhan coined „**Global Village**“
- **Milgram's** experiment: average **path length** between „random people“

## **Milgram's experiment:**

- Choose random people from US Midwest
- Tell them to send letter to some guy in Boston
- They are only allowed to forward letter to someone they know on **first name basis!**
- How many „hops“ until letters arrive? (What do you think?)

# Milgram's Experiment

How many hops until letters arrive?



# Results and Discussion

---

## Results:

- Many letters got **lost**
- But for the ones that arrived, the average hop distance was **5.5!**
- „**six-degrees-of-separation**“ / „small world“

**How to explain? Not only small diameter but also „**navigable**“!**

## Still an important research question!

- E.g., concept of „**power-law**“ **graphs**: node degrees are distributed according to power law, i.e., number of nodes with degree  $\delta$  is proportional to  $\delta^\alpha$  for some  $\alpha > 0$
- Power law graphs have been observed also in the Internet, in **biology**, in physics, etc.
- Kleinberg's explanation: regular grid with a **small number of random links!**

## Discussion

- Implications of people being closely connected?!
- Many people criticize experiment, or generalization of insight: e.g., letter **from white person to black person** needs more hops, etc.

# Implication for Computer Science?

The good properties of social networks have inspired scientists and practitioners to **build computer networks** with similar characteristics!

**Example: Jon Kleinberg at Cornell discovered navigable („greedy routable“) social networks, which inspired Clarke/Sandberg/... to build the peer-to-peer system **Freenet**.**

## Navigation in a small world

It is easier to find short chains between points in some networks than others.

The small-world phenomenon — the principle that most of us are linked by short chains of acquaintances — was first investigated as a question in sociology<sup>1</sup> and is a feature of a range of networks arising in nature and technology<sup>2,3</sup>. Experimental study of the phenomenon<sup>4</sup> revealed that it has two fundamental components: first, such short chains are ubiquitous, and second, individuals operating with purely local information are very adept at finding these chains. The first issue has been analysed<sup>5,6</sup>, and here I investigate the second by modelling how individuals can find short chains in a large social network.

I have found that the cues needed for discovering short chains emerge in a very simple network model. This model is based on early experiments<sup>7</sup>, in which source individuals in Nebraska attempted to transmit a letter to a target in Massachusetts, with the letter being forwarded at each step to someone the holder knew on a first-name basis. The networks underlying the model follow the 'small-world' paradigm<sup>8</sup>: they are rich in structured short-range connections and have a few random long-range connections.

Long-range connections are added to a two-dimensional lattice controlled by a clustering exponent,  $\alpha$ , that determines the probability of a connection between two nodes as a function of their lattice distance (Fig. 1a). Decentralized algorithms are studied for transmitting a message: at each step, the holder of the message must pass it across one of its short- or long-range connections; crucially, this current holder does not know the long-range connections of nodes that have not touched the message. The primary figure of merit for such an algorithm is its expected delivery time  $T$ , which represents the expected number of steps needed to forward a message between

nodes following an inverse-square distribution, there is a decentralized algorithm that achieves a very rapid delivery time;  $T$  is bounded by a function proportional to  $(\log N)^2$ . The algorithm achieving this bound is a 'greedy' heuristic: each message holder forwards the message across a con-

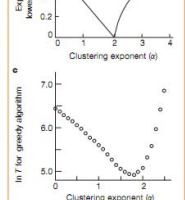
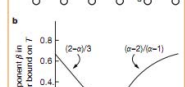
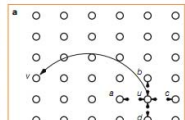
nection that brings it as close as possible to the target in lattice distance. Moreover,  $\alpha = 2$  is the only exponent at which any decentralized algorithm can achieve a delivery time bounded by any polynomial in  $\log N$ ; for every other exponent, an asymptotically much larger delivery time is required, regardless of the algorithm employed (Fig. 1b).


These results indicate that efficient navigability is a fundamental property of only some small-world structures. The results also generalize to  $d$ -dimensional lattices for any value of  $d \geq 1$ , with the critical value of the clustering exponent becoming  $\alpha = d$ . Simulations of the greedy algorithm yield results that are qualitatively consistent with the asymptotic analytical bounds (Fig. 1c).

In the areas of communication networks<sup>9</sup> and neuroanatomy<sup>10</sup>, the issue of routing without a global network organization has been considered; also in social psychology and information foraging some of the cues that individuals use to construct paths through a social network or hyper-linked environment have been discovered<sup>11</sup>. Although I have focused on a very clean model, I believe that a more general conclusion can be drawn for small-world networks — namely that the correlation between local structure and long-range connections provides critical cues for finding paths through the network.

When this correlation is near a critical threshold, the structure of the long-range connections forms a type of gradient that allows individuals to guide a message efficiently towards a target. As the correlation drops below this critical value and the social network becomes more homogeneous, these cues begin to disappear; in the limit, when long-range connections are generated uniformly at random, the result is a world





### brief communications



**Freenet**  
THE FREE NETWORK

Home Download About Help Donate Contribute

### Papers

-  [The Dark Freenet](#) (PDF)  
Detailed paper about the Freenet 0.7.5 network, as opposed to its routing algorithm, which is detailed in the below papers. Includes some new simulations. This has been submitted to PET 2010.
-  [Video of Small World talk, Berlin, December 2005](#)  
This is a video of a talk given by Ian Clarke and Oskar Sandberg at the Chaos Computer Congress in Berlin, December 2005, describing the (then) new architecture for Freenet 0.7. You can also download the [slideshow](#), and the source for the Java [demo](#) (requires Java 1.5).
-  [Searching in a Small World](#) (PDF)  
Oskar Sandberg's licentiate thesis describing a simple decentralized mechanism for constructing small world networks that is inspired by Freenet's original design. Part II of the thesis describes the basis for the new Darknet architecture.
-  [Distributed routing in Small World Networks](#) (PDF)  
A paper by Oskar Sandberg describing the theoretical basis for the new "Darknet" routing mechanism employed by Freenet 0.7.
-  [Chaos Computer Congress Talk](#) (slideshow)  
This is a [slideshow](#) for a talk given at the Chaos Computer Congress on 30th Dec 2005 in Berlin, Germany by Ian Clarke and Oskar Sandberg. It described the new "darknet" approach to be employed in Freenet 0.7. A Java demonstration to accompany the talk is [also](#) available.
-  [Switching for a small world](#) (PDF)  
A thesis by Vilhelm Verendel exploring ways to optimise the swapping algorithm.
-  [Routing in the Dark: Pitch Black](#) (citeseeq) (PDF)  
A paper describing some attacks on Freenet 0.7's location swapping algorithm.
- [Next Generation Routing Algorithm](#)  
This article describes Freenet's 'Next Generation' routing algorithm. This was a novel approach which had nodes making routing decisions based on sophisticated analysis of the time required to route previous requests. This algorithm was promising, but was eventually dropped in favour of a much simpler approach in Freenet 0.7.
-  [Protecting Freedom of Information Online with Freenet](#) (PDF)  
An IEEE Internet Computing article describing the Freenet architecture circa 2002 - probably the best introduction to the

# Watts-Strogatz Model

---

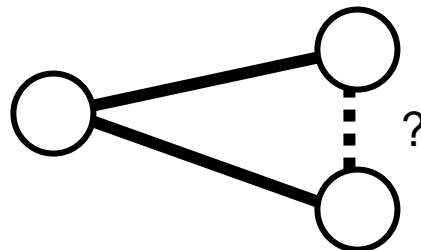
Idea: Model social network as combination of **two networks!**

- (1) As a basis network with **large cluster coefficient**,
- (2) and then add **random links** (e.g., constant number of random nodes all over the graph): **long-range acquaintances**

Important concept in social networks:

## Cluster Coefficient

Probability that two friends of a node are also friends, summed up over all nodes.

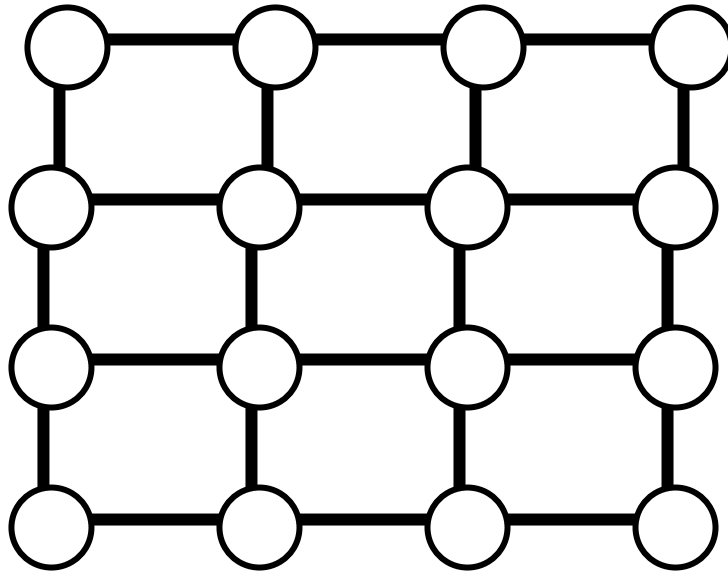


**Problem:**  
Is clustering a good measure?  
And how to navigate??

# Cluster Coefficient

---

Careful, not so useful in the „worst-case“: cluster coefficient of grid?



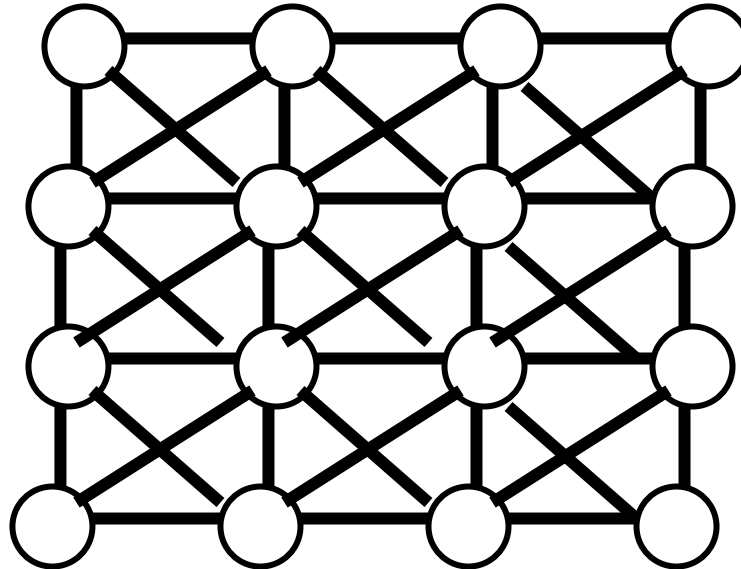
**Coefficient: 0**



# Cluster Coefficient

---

And here?



**Coefficient:  $3/7$**

**(out of 28 pairs of neighbors, 12 are neighbors)**

# But how to route...?!

---

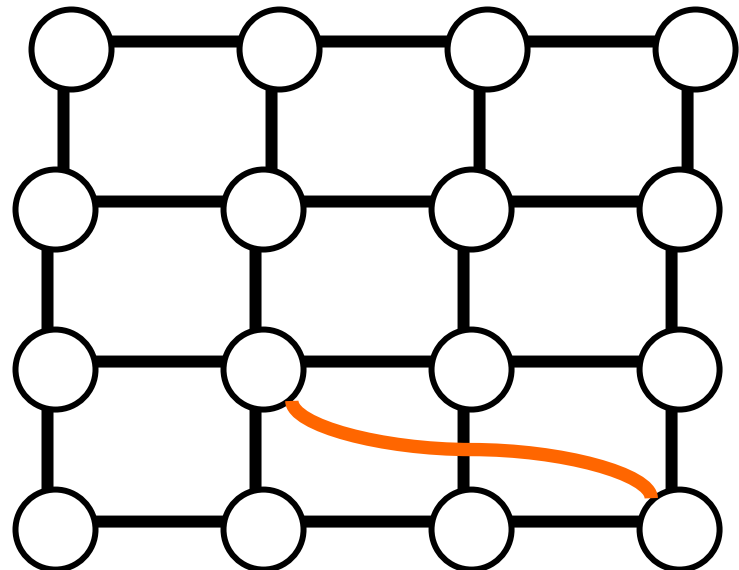
Having a graph model is good, but how did the mail arrive?!

**Jon Kleinberg**: most simple „**navigable**“ social networks

- (1) Nodes **forward** letter to neighbor, without really knowing whether neighbor is closer to destination!
- (2) Only **greedy routing** can explain phenomenon...

Jon Kleinberg's graph model:  
the **augmented  $m \times m$  grid**

A simple grid **with some additional random links** per node with **specific random distribution**.



## Augmented Grid

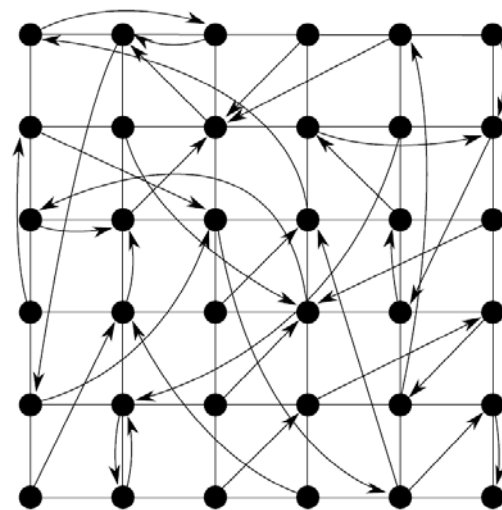
Consider an  $(m \times m)$  grid of  $n=m^2$  nodes, where each node has a directed edge to each lattice neighbor (**local contacts**). In addition, each node has an additional random link (**long-range contact**). For all  $u$  and  $v$ , the long-range contact of  $u$  to points to node  $v$  with probability  $d(u,v)^{-\alpha} / \sum_{w \in V \setminus \{u\}} d(u,w)^{-\alpha}$ , where  $d()$  is the distance in the grid and  $\alpha$  is a parameter.

**Geographic interpretation?**

**Interpretation of  $\alpha$ ?**

Parameter  $\alpha=0$  means uniform at random (indep. of distance!); larger  $\alpha$  make long-range links **shorter**.

One can show: if  $\alpha \leq 2$ , the **diameter** is **polylogarithmic**. ( $\alpha=0$  implies log diameter: proof as exercise?) But what about **routing**?



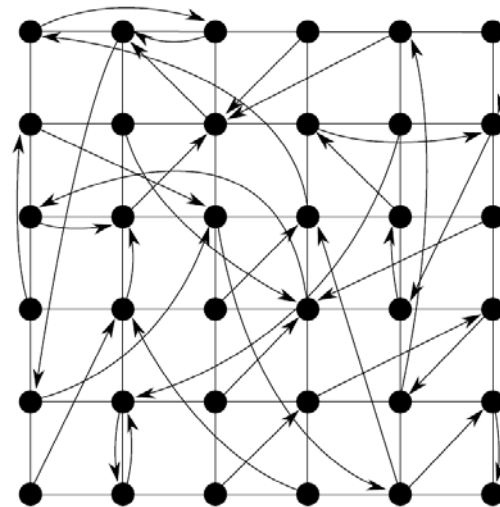
**How to navigate??**

Idea how to route on Kleinberg's graph?

## Greedy Routing

while (not at destination):  
go to neighbor which is closest to destination  
(considering **grid distance** only)

Runtime?



## Greedy Routing

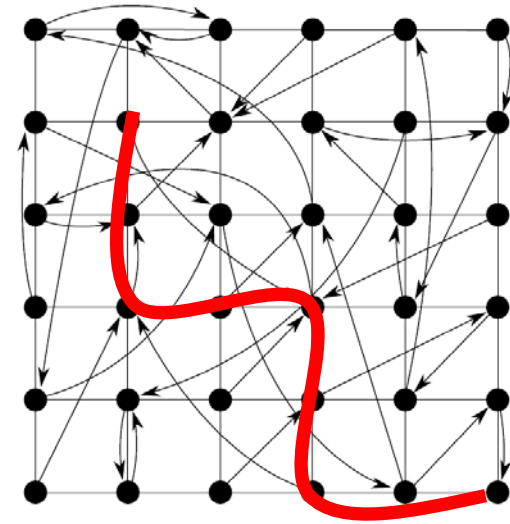
The greedy algorithm finds a routing path of length at most  $O(\sqrt{n})$ .

Proof idea?

There are always neighbors closer to the destination. We can reduce the distance in at least one **grid dimension**...

**QED**

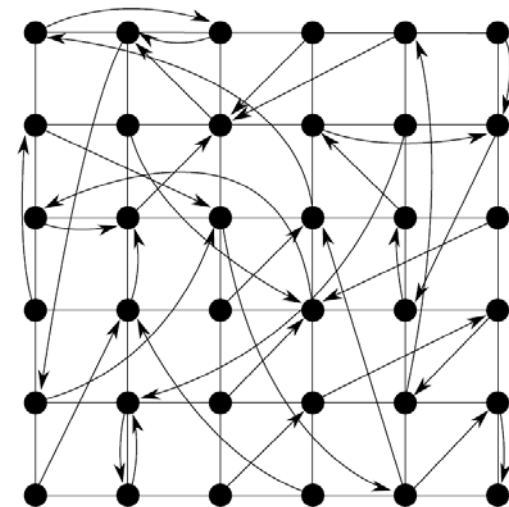
**But Milgram promises more!  
How do random links speed it up?**



## Random Links

A node  $u$ 's random link leads to a node  $v$  with probability

$$\begin{aligned} \Theta(1/(d(u,v)^\alpha m^{2-\alpha})) & \quad \text{if } \alpha < 2 \\ \Theta(1/(d(u,v)^2 \log n)) & \quad \text{if } \alpha = 2 \\ \Theta(1/d(u,v)^\alpha) & \quad \text{if } \alpha > 2 \end{aligned}$$



Proof?

since  $d(\cdot, \cdot) \geq 1$ , and there are at most  $O(r)$  new nodes at distance  $r$

$$\text{For } \alpha \neq 2: \sum_{w \in V \setminus \{u\}} \frac{1}{d(u, w)^\alpha} \in \sum_{r=1}^m \frac{\Theta(r)}{r^\alpha} = \Theta \left( \int_{r=1}^m \frac{1}{r^{\alpha-1}} dr \right) = \Theta \left( \left[ \frac{r^{2-\alpha}}{2-\alpha} \right]_1^m \right)$$

For  $\alpha < 2$  this is  $\Theta(m^{2-\alpha})$ , for  $\alpha > 2$  this is  $\Theta(1)$ .

$$\text{For } \alpha = 2 \text{ we have: } \sum_{w \in V \setminus \{u\}} \frac{1}{d(u, w)^\alpha} \in \sum_{r=1}^m \frac{\Theta(r)}{r^2} = \Theta(1) \cdot \sum_{r=1}^m \frac{1}{r} = \Theta(\log m) = \Theta(\log n)$$

QED

# Discussion

## What does it mean?

$\Theta(1/(d(u,v)^\alpha m^{2-\alpha}))$  if  $\alpha < 2$  ☹️: for small  $\alpha$ , likely long paths...

$\Theta(1/(d(u,v)^2 \log(n)))$  if  $\alpha = 2$  😊?

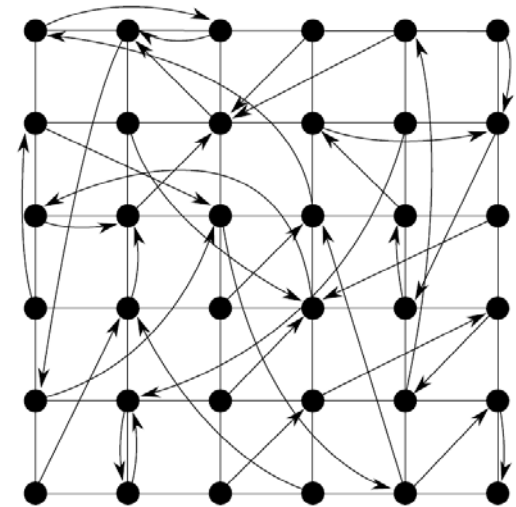
$\Theta(1/d(u,v)^\alpha)$  if  $\alpha > 2$  ☹️: for large  $\alpha$ , likely long paths...

$m = \sqrt{n}$ ...

exponential in  $\alpha$

$\alpha < 2$ : Links have roughly same distance. We have to go far until finding a link that points close to destination („too random“)! Until then, we **walk on grid**...

$\alpha > 2$ : Mostly short links only, and we have to go far until finding a link that **reaches far** („too focused“)! Until then, we don't make much progress wrt distance...



# Navigability: $\alpha=2$

---

So let's study the case  $\alpha=2$ !

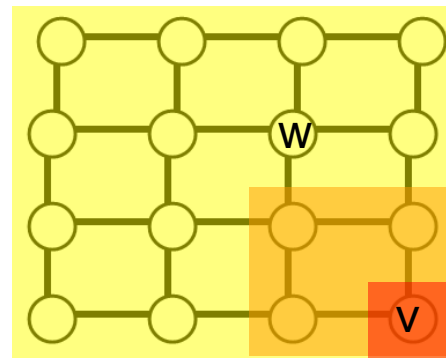
## Phase

Consider routing from node  $u$  to  $v$ , and assume we are at some intermediate node  $w$ . We say we are in phase  $j$  at node  $w$  if the lattice distance  $d(w,v)$  to the target  $v$  is between  $2^j < d(w,v) \leq 2^{j+1}$ .

So we **count-down phases**!

How many phases are there at most?

Logarithmic in max routing distance,  
i.e.,  $O(\log m) = O(\log n)$ .



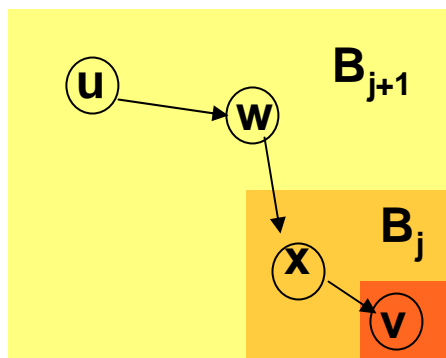


## Lemma

Assume we are **at node  $w$**  in phase  $j$  (on the way from  $u$  to  $v$ ). The probability of getting to phase  $j-1$  **in one step** is at least  $\Omega(1/\log n)$ .

### Proof idea?

Let  $B_j$  denote the nodes with  $d(x,v) \leq 2^j$ . We get **to phase  $j-1$  from phase  $j$**  if the **long-range contact** of  $w$  points to some node in  $B_j$ . Since we have not been at  $w$  before, its link points to a random node, **independent** of path to  $w$ .



# Navigability: $\alpha=2$

We know that for all nodes  $x \in B_j$ ,  
 $d(w,x) \leq d(w,v) + d(v,x) \leq 2^{j+1} + 2^j < 2^{j+2}$   
(**triangle inequality**).

So for each  $x \in B_j$ , the probability that  
 $w$  points to  $x$  is  
 $\Omega(1/(2^{2j+4} \log n))$ .

Why?

See **our lemma**...

So probability to hit one of them?

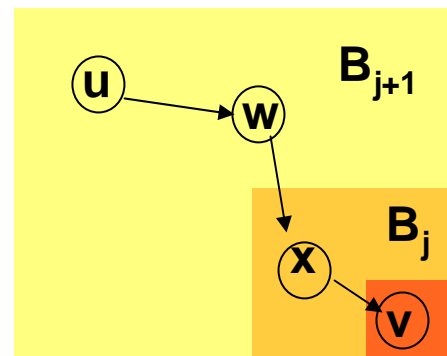
**Sum over all nodes** that are in  $B_j$ ...

And  $B_j$  (number of nodes at distance) grows **quadratic**, so at least  $(2^j)/2$ .

So:

$$\Omega \left( |B_j| \cdot \frac{1}{2^{2j+4} \log n} \right) = \Omega \left( \frac{2^{2j-1}}{2^{2j+4} \log n} \right) = \Omega \left( \frac{1}{\log n} \right)$$

**QED**



## Theorem

The expected path length is  $O(\log^2 n)$ .

### Proof idea?

We have  $O(\log n)$  phases. We proceed from one phase to the next with probability at least  $\Omega(1/\log n)$ .

And linearity of expectation... 😊

**QED**

**Yay, it's polylogarithmic!**  
**So does it explain Milgram? 😊**

# Propagation Studies

---

**Where to put ice cream stand?**



In the middle...

**Where to put the second stand?**

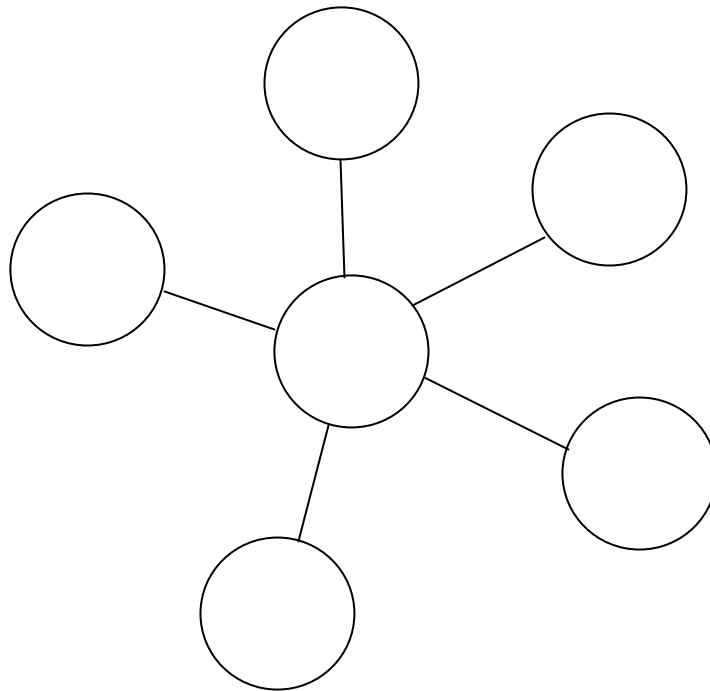
Right next to it?

**Does the first player always have an advantage (cover more customers)?**

# Propagation Studies

---

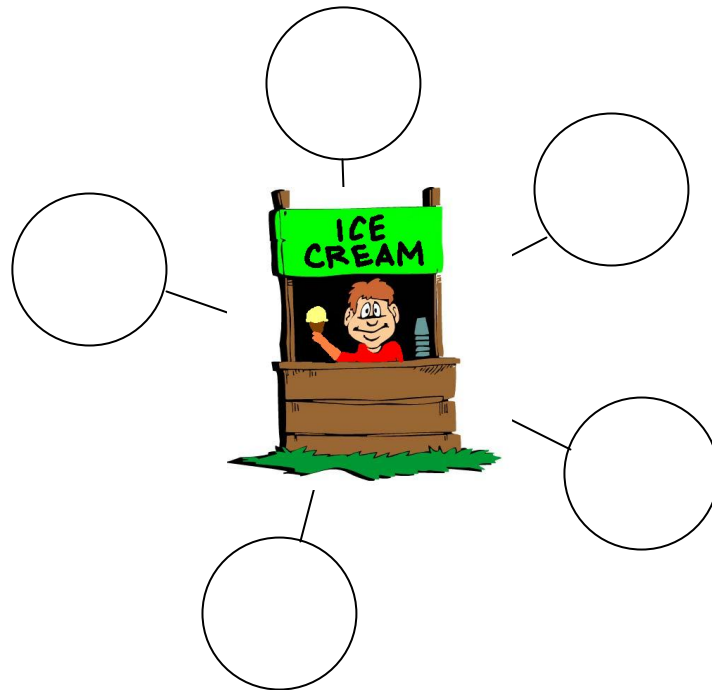
**Here?**



# Always first mover advantage?

---

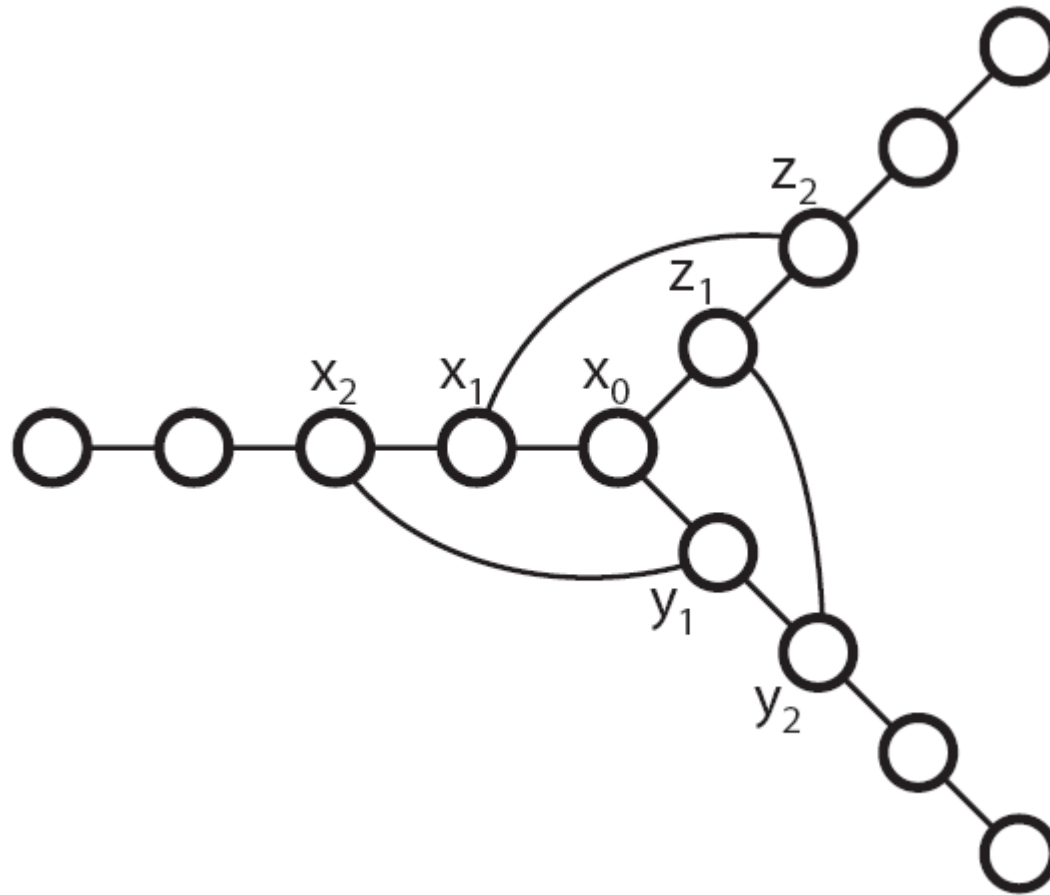
Here yes:



# Always first mover advantage?

---

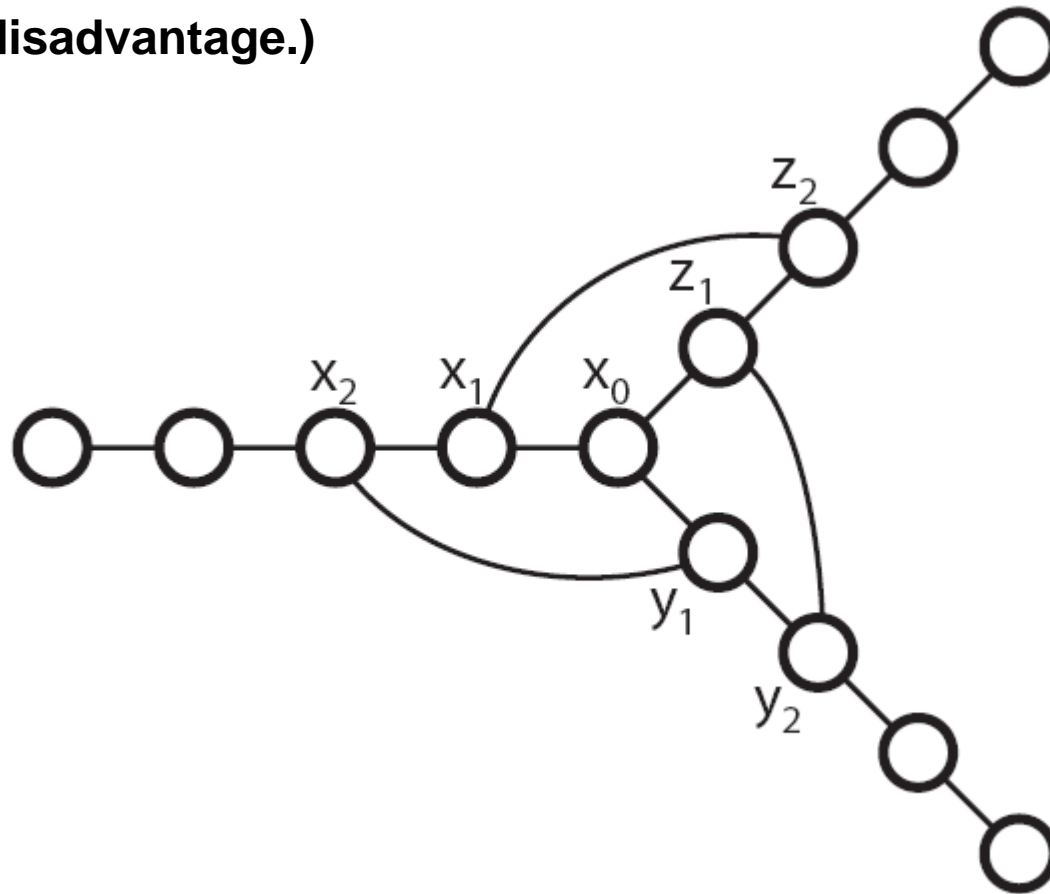
Here?



# Always first mover advantage?

---

**No advantage!**  
**(But also no disadvantage.)**





Literature for further reading:

- not Peleg's book 😊

End of lecture