# Small Worlds Navigability

Pierre Fraigniaud
Emmanuelle Lebhar
Zvi Lotker

# Interaction Networks

- Communication networks
  - Internet
  - Ad hoc and sensor networks
- Societal networks
  - The Web
  - P2P networks (the unstructured ones)
- Social network
  - Acquaintance
  - Mail exchanges
- Biology, linguistics, etc.

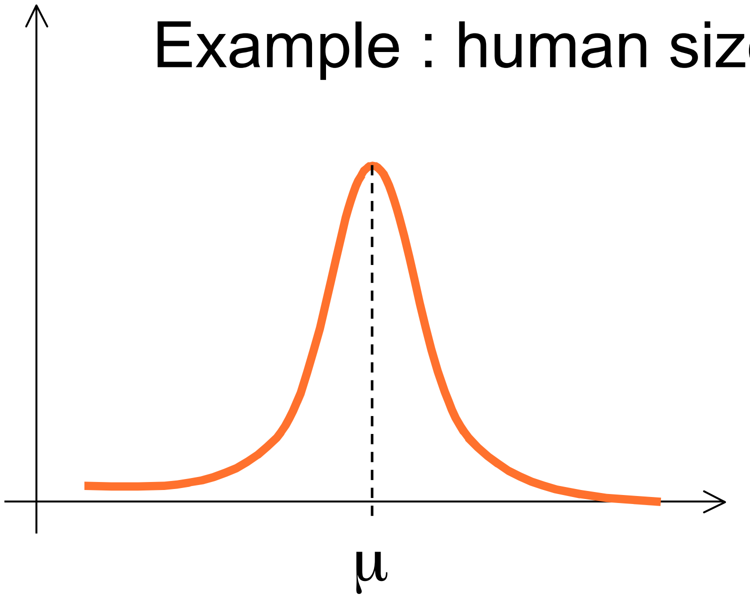
# Common statistical properties

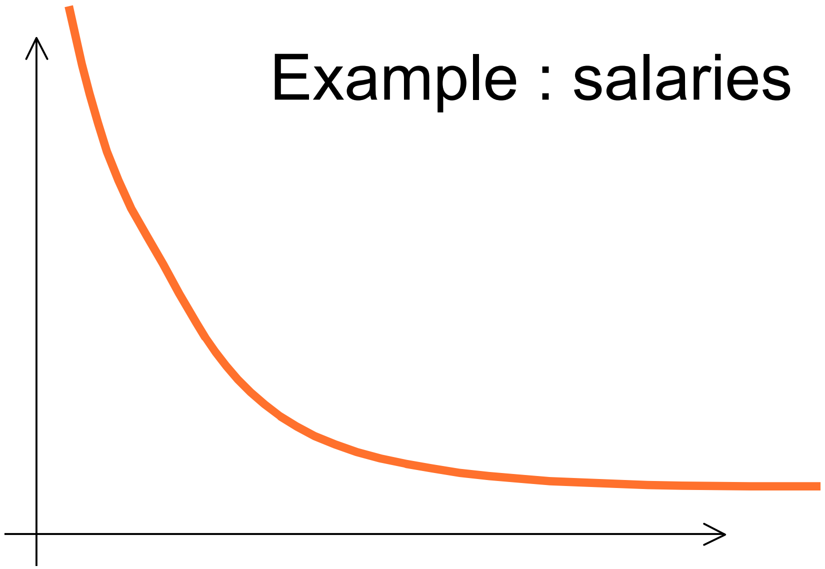
- Low density
- "Small world" properties:
  - Average distance between two nodes is small, typically $O(\log n)$
  - The probability $p$ that two distinct neighbors $u_1$ and $u_2$ of a same node $v$ are neighbors is large.

$p$ = clustering coefficient

- "Scale free" properties:
  - Heavy tailed probability distributions (e.g., of the degrees)

# Gaussian vs. Heavy tail



# Random graphs vs. Interaction networks

- Random graphs $G_{n,p}$ with $p \approx \log(n)/n$
  - low clustering coefficient
  - Gaussian distribution of the degrees
- Interaction networks
  - High clustering coefficient
  - Heavy tailed distribution of the degrees

# New problematic

- Why these networks share these properties?
- What model for
  - Performance analysis of these networks
  - Algorithm design for these networks
- Impact of the measures?
- This talk addresses navigability

# Milgram Experiment

- Source person s (e.g., in Wichita)
- Target person t (e.g., in Cambridge)
  - Name, professional occupation, city of living, etc.
- Letter transmitted via a chain of individuals related on a personal basis
- Result: "six degrees of separation"

# Navigability

- Jon Kleinberg (2000)
  - Why should there exist short chains of acquaintances linking together arbitrary pairs of strangers?
  - Why should arbitrary pairs of strangers be able to find short chains of acquaintances that link them together?
- In other words: how to navigate in a small worlds?

# Augmented graphs H=G+D

- Individuals as nodes of a graph G
  - Edges of G model relations between individuals deducible from their societal positions
- A number k of “long links” are added to G at random, according to the probability distribution D
  - Long links model relations between individuals that cannot be deduced from their societal positions

# Greedy Routing in augmented graphs

- Source $s \in V(G)$
- Target $t \in V(G)$
- Current node $x$ selects among its $deg_G(x)+k$ neighbors the closest to $t$ in $G$, that is according to the distance function $dist_G()$.

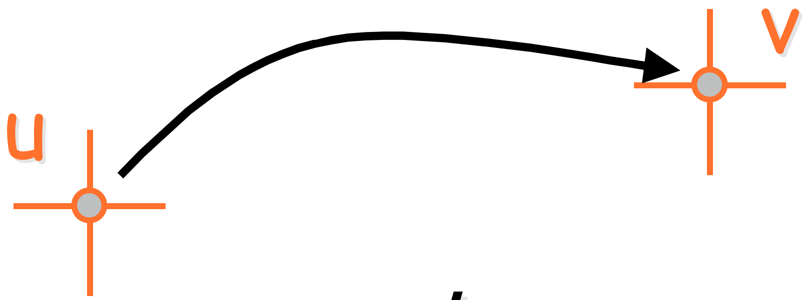
Greedy routing in augmented graphs aims at modeling the routing process performed by social entities in Milgram's experiment.

# Augmented meshes

## Kleinberg [STOC 2000]

d-dimensional n-node meshes

augmented with d-harmonic links




$$\text{prob}(u \rightarrow v) \approx 1 / (\log(n) * \text{dist}(u,v)^d)$$

# Kleinberg's theorems

- Greedy routing performs in $O(\log^2 n / k)$ expected #steps in d-dimensional meshes augmented with k links per node, chosen according to the d-harmonic distribution.
  - Note: $k = \log n \Rightarrow O(\log n)$ expect. #steps
- Greedy routing in d-dimensional meshes augmented with a h-harmonic distribution, $h \neq d$, performs in $\Omega(n^{\varepsilon})$ expected #steps.

# Extensions

- Two-step greedy routing: $O(\log n / \log\log n)$
  - Coppersmith, Gamarnik, Sviridenko (2002)
    - Percolation theory
  - Manku, Naor, Wieder (2004)
    - NoN routing
- Routing with partial knowledge: $O(\log^{1+1/d} n)$
  - Martel, Nguyen (2004)
    - Non-oblivious routing
  - Fraigniaud, Gavoille, Paul (2004)
    - Oblivious routing
- Decentralized routing: $O(\log n * \log^2 \log n)$
  - Lebhar, Schabanel (2004)
    - $O(\log^2 n)$ expected #steps to find the route

# Navigable graphs

- Let $f : \mathbf{N} \rightarrow \mathbf{R}$ be a function
- An $n$-node graph $G$ is $f$-navigable if there exists an augmentation $D$ for $G$ such that greedy routing in $G+D$ performs in at most $f(n)$ expected #steps.
- I.e., for any two nodes $u,v$ we have

$$\mathbf{E}_D(\#\mathrm{steps}_{u \rightarrow v}) \leq f(n)$$

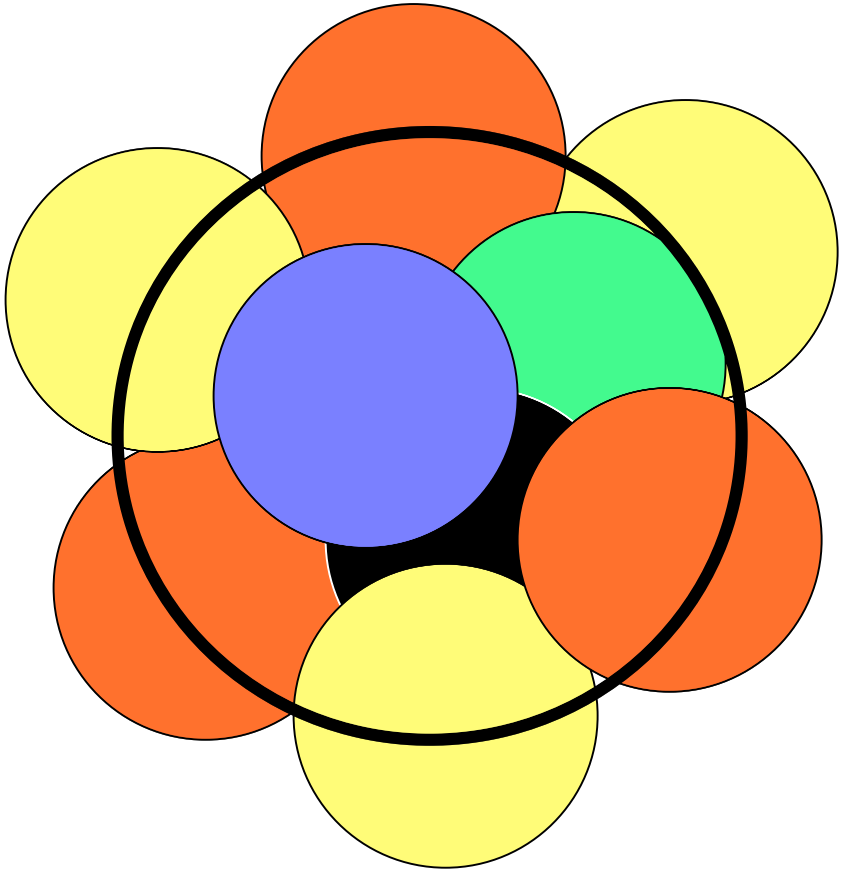
# O(polylog(n))-navigable graphs

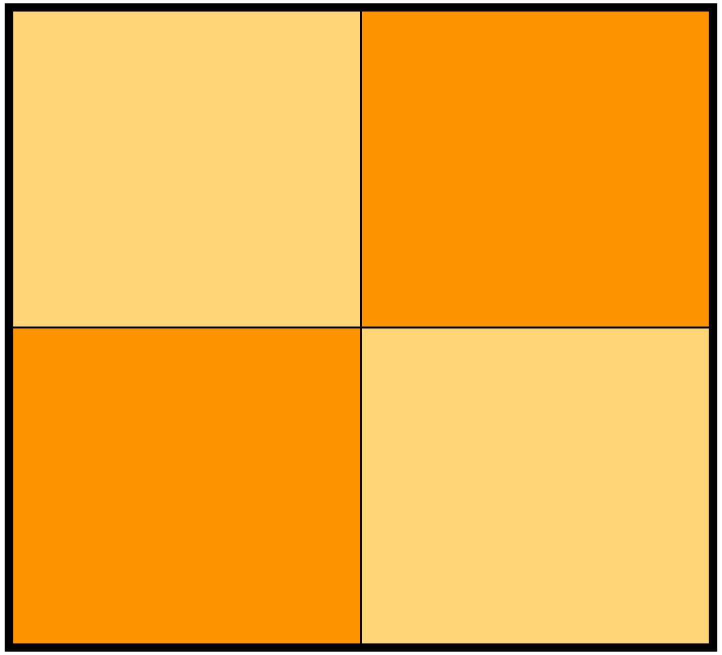
- Bounded growth graphs
  - Definition: $|B(x,2r)| \leq \rho\ |B(x,r)|$
  - Duchon, Hanusse, Lebhar, Schabanel (2005,2006)
- Bounded doubling dimension
  - Definition: Every $B(x,2r)$ can be covered by at most $2^d$ balls of radius $r$
  - Slivkins (2005)
- Graphs of bounded treewidth
  - Fraigniaud (2005)
- Graphs excluding a fixed minor
  - Abraham, Gavoille (2006)

# Question

Are all graphs O(polylog(n))-navigable?

# Doubling dimension





# Svilkins' theorem

Any family of graphs with doubling dimension O(loglog n) is navigable.

# Impossibility result

Theorem

Let d such that

$$\lim_{n \to +\infty} \log\log n / d(n) = 0$$

There exists an infinite family of n-node graphs with doubling dimension at most d(n) that are not O(polylog(n))-navigable.

Consequences:

1. Slivkins' result is tight
2. Not all graphs are O(polylog(n))-navigable

# Proof of non-navigability

The graphs $H_d$ with $n=p^d$ nodes

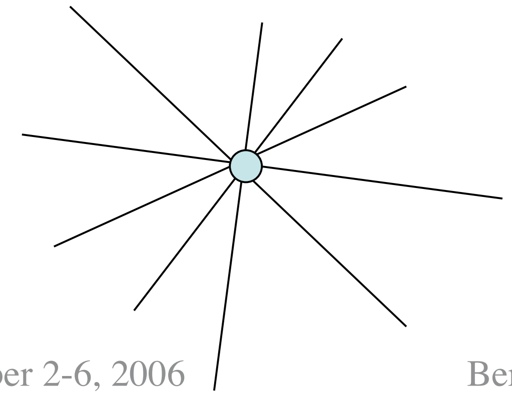


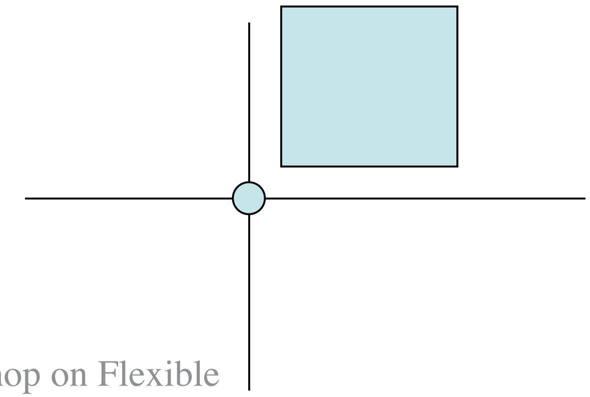
$x = x_1 x_2 \dots x_d$
is connected to all nodes
$y = y_1 y_2 \dots y_d$
such that $y_i = x_i + a_i$ where
$a_i \in \{-1,0,+1\}$

$H_d$ has doubling dimension $d$

# Intuitive approach

- Large doubling dimension d
  $\Rightarrow$ every nodes $x \in H_d$ has choices over exponentially many directions
- The underlying metric of $H_d$ is $L_\infty$

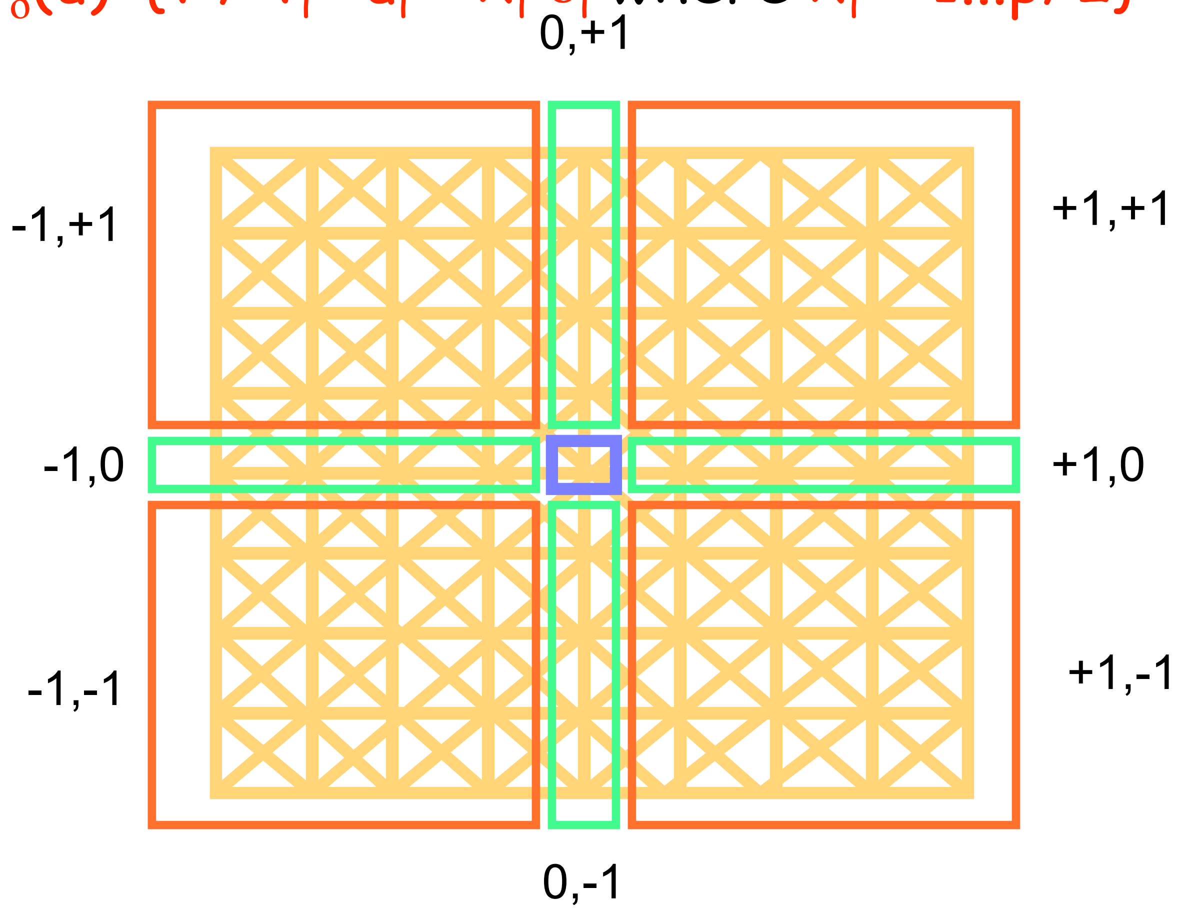




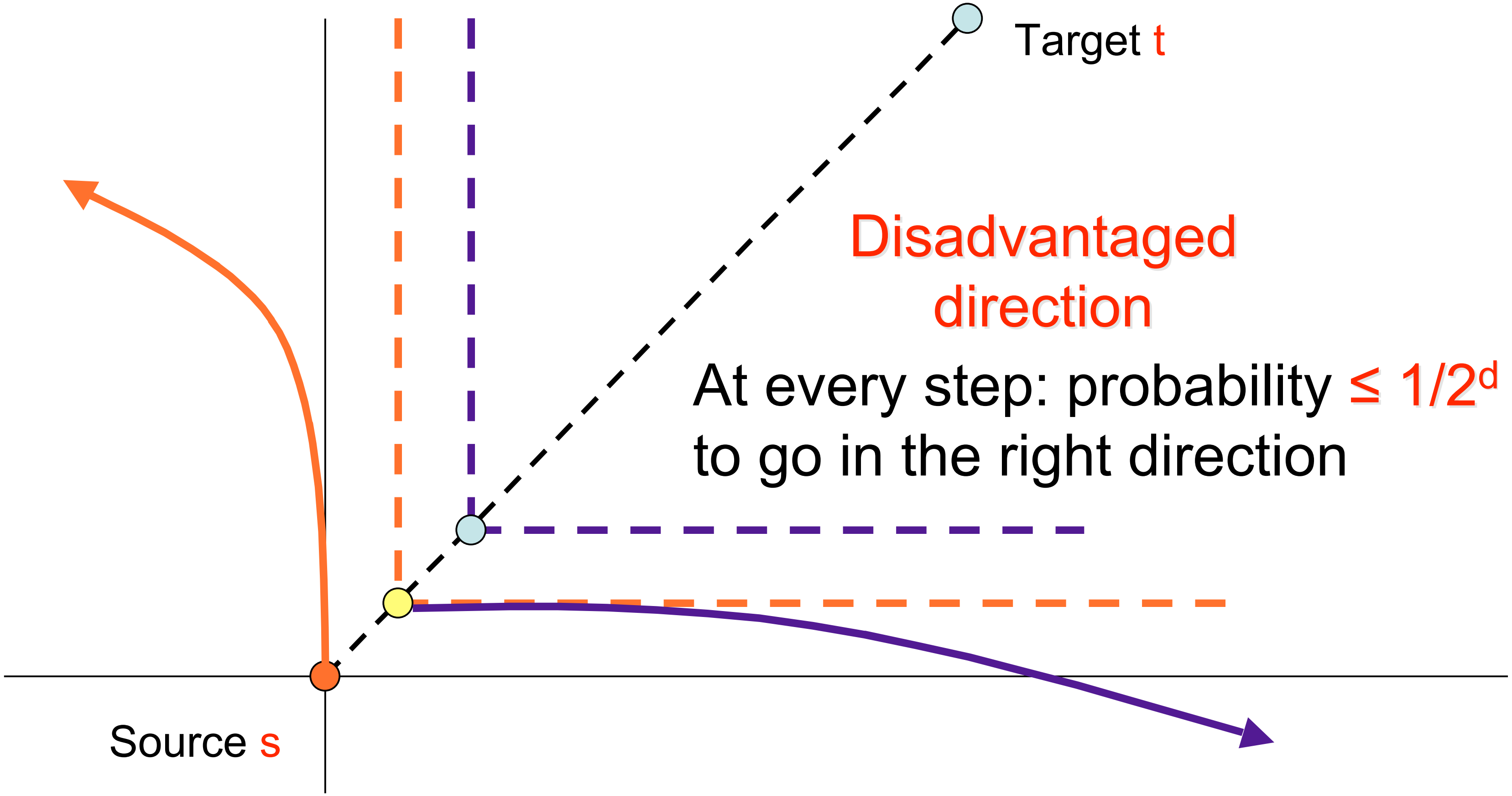
# Directions

$\delta = (\delta_1, \ldots, \delta_d)$ where $\delta_i \in \{-1,0,+1\}$

$Dir_\delta(u)=\{v \,/\, v_i = u_i + x_i\, \delta_i$ where $x_i = 1 \ldots p/2\}$

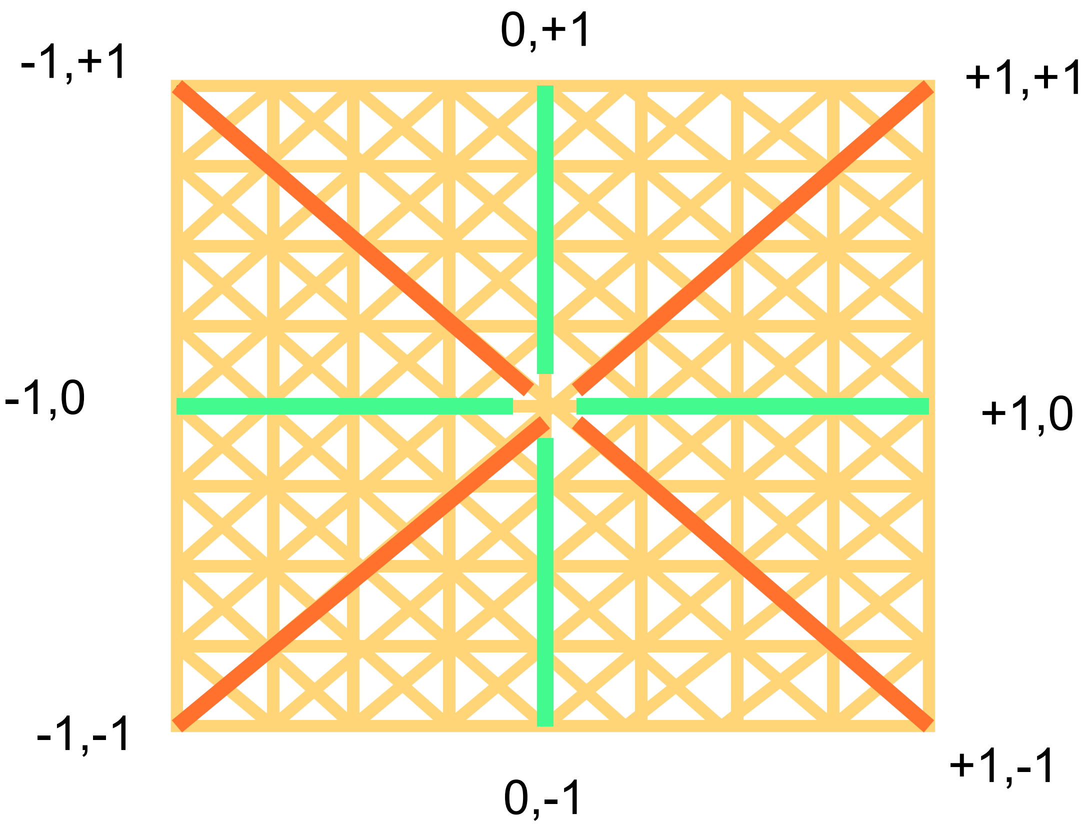


# Case of symmetric distribution

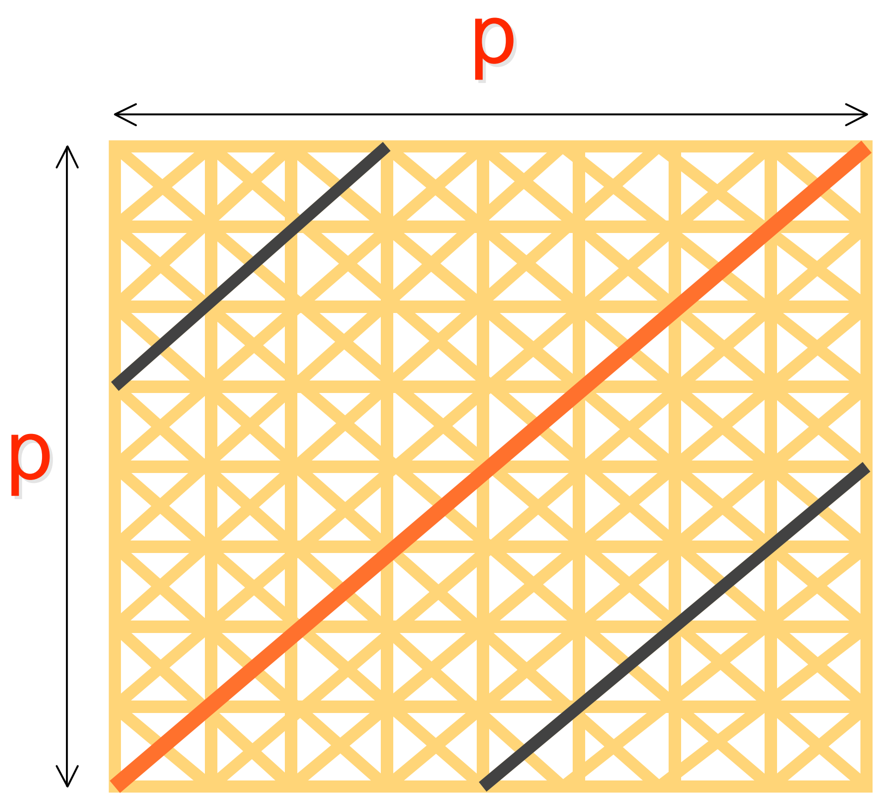


# -- General case --

# Diagonals

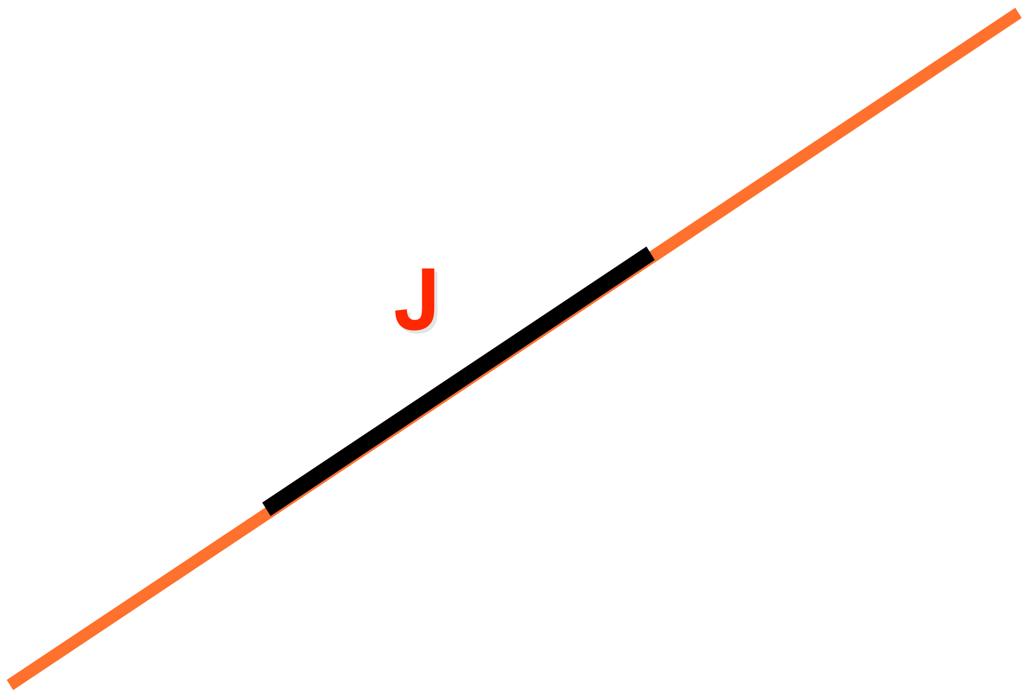


# Lines

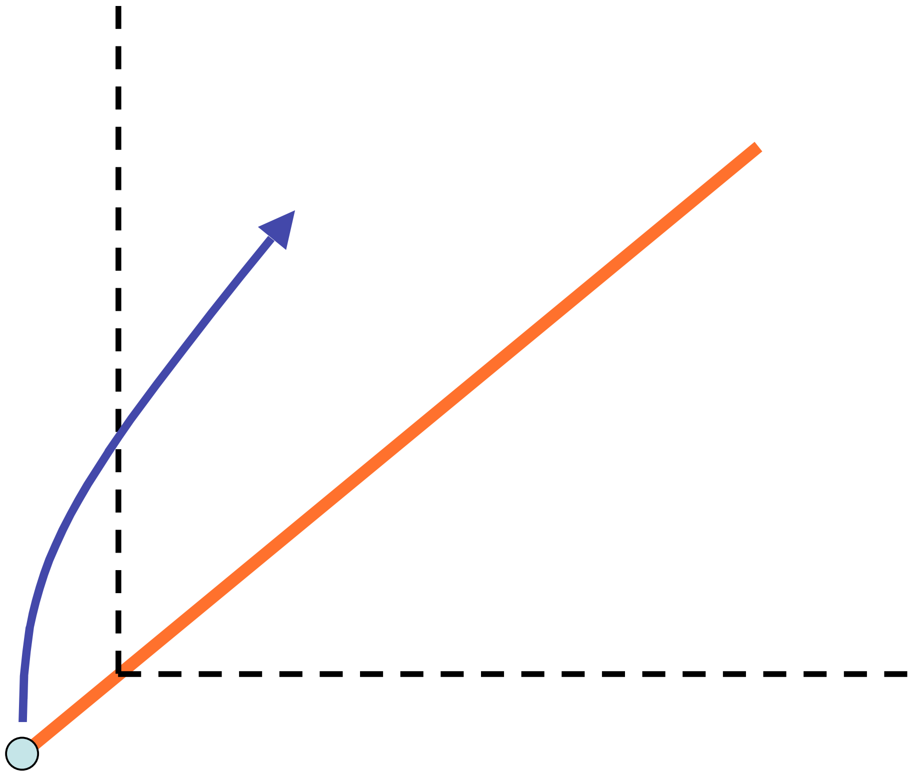


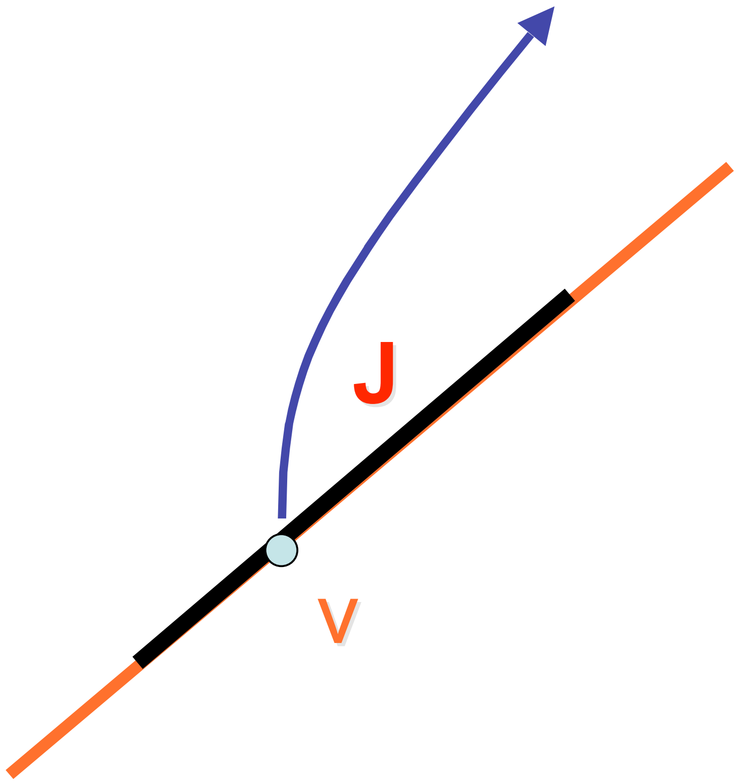

p lines in each direction

# Intervals



# Certificates



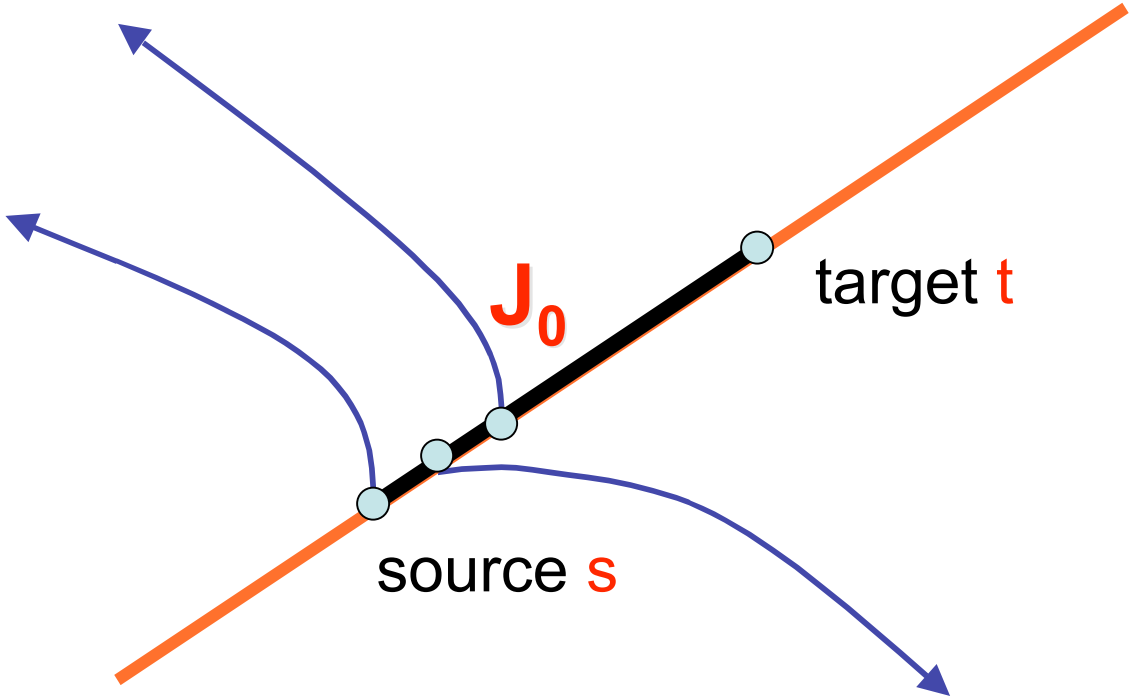



v is a certificate for J

# Counting argument

- $2^d$ directions
- Lines are split in intervals of length $L$
- $n/L \times 2^d$ intervals in total
- Every node belongs to many intervals, but can be the certificate of at most one interval
- If $L<2^d$ there is one interval $J_0$ without certificate

# L-1 steps from s to t



# In expectation…

- $n/L \times 2^d - n$ intervals without certificate
- $L = 2^{d-1} \Rightarrow n$ of the $2n$ intervals are without certificate
- This is true for any trial of the long links
- Hence $E = \mathbf{E}_D$(#interval without certificate) $\geq n$
- On the other hand:

$$E = \sum_J \Pr(J \text{ has no certificate})$$

- Hence there is an interval $J_0=[s,t]$ such that

$$\Pr(J_0 \text{ has no certificate}) \geq 1/2$$

- Hence $\mathbf{E}_D(\#steps_{s\to t}) \geq (L-1)/2$ **QED**

Remark: The proof still holds even if the long links are not set pairwise independently.

# Open Problem

- We have considered the worst case:

$$\max_{u,v} \mathbf{E}_D(\#steps_{u \to v})$$

- What about the average case?

$$\Sigma_{u,v} \mathbf{E}_D(\#steps_{u \to v}) / n^2$$