Small Words
Navigability

Pierre Fraigniaud

Master MPRI

26/09/06

Introduction

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.

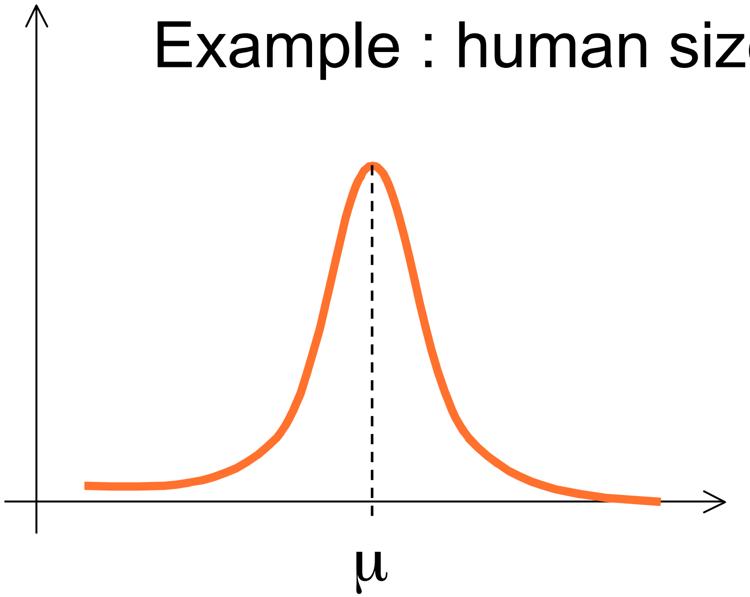
Common statistical properties

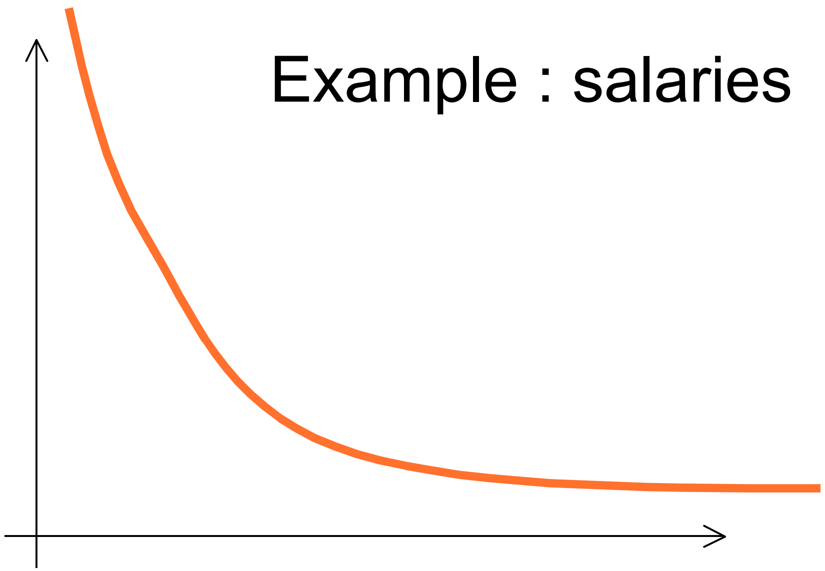
- Low density
- "Small world" properties:
  - Average distance between two nodes is small, typically O(log n)
  - The probability p that two distinct neighbors u_1 and u_2 of a same node v are neighbors is large.

p = clustering coefficient

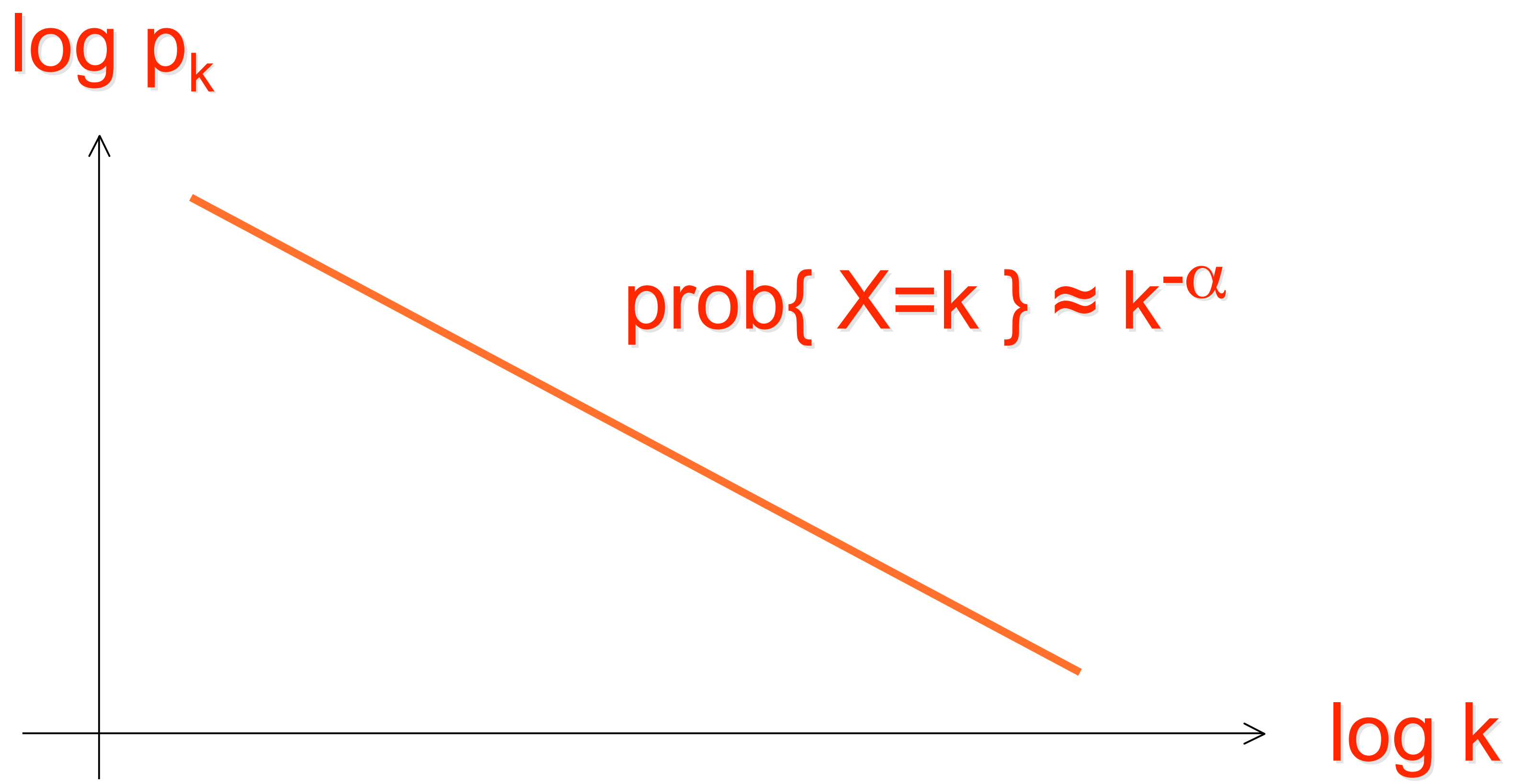
- "Scale free" properties:
  - Heavy tailed probability distributions (e.g., of the degrees)

Gaussian vs. Heavy tail





Power law



Random graphs vs. Interaction networks

- Random graphs
  - prob(e exists) ≈ 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 lecture addresses navigability

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?

Prix Nevanlinna

- Prix récompensant une contribution majeure dans le domaine des mathématiques, dans son aspect informatique.
- Lauréats
  - 1982 - Robert Tarjan
  - 1986 - Leslie Valiant
  - 1990 - A.A. Razborov
  - 1994 - Avi Wigderson
  - 1998 - Peter Shor
  - 2002 - Madhu Sudan
  - 2006 - Jon Kleinberg

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 $dist_G()$.

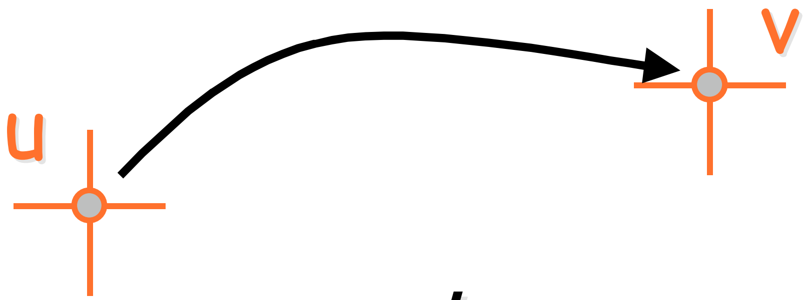
Greedy routing in augmented graphs aims at modeling the routing process performed by social entities in Milgram's experiment.

Augmented meshes

Kleinberg [STOC 2000]

d-dimensional n-node meshes

augmented with d-harmonic links



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

Harmonic distribution

- d-dimensional mesh
- $B(x,r)$ = ball centered at x of radius r
- $S(x,r)$ = sphere centered at x of radius r
- In d-dimensional meshes:

$$|B(x,r)| \approx r^d$$

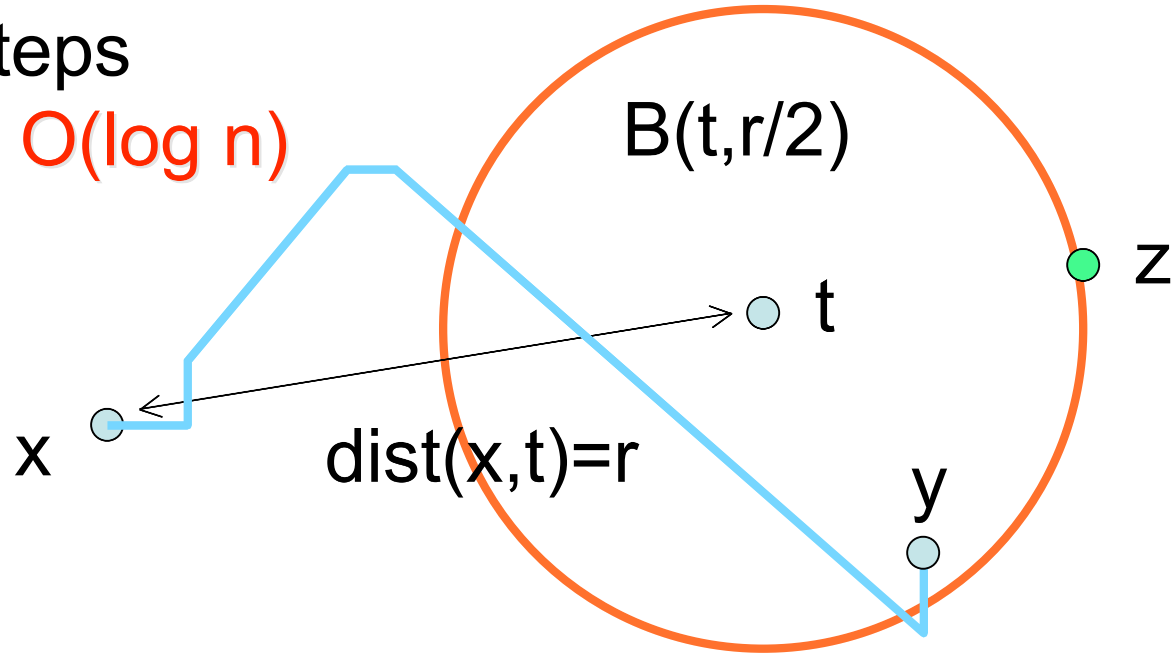
$$|S(x,r)| \approx r^{d-1}$$

$$\Sigma_{v \neq u}(1/dist(u,v)^d) = \Sigma_r |S(u,r)|/r^d$$

$$\approx \Sigma_r 1/r \approx \log n$$

Performances

Expected #steps
to enter B(t,r/2) is O(log n)



For a current node x at distance r from t,
prob{x → B(t,r/2)} is at least Ω(1/log n)

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

An impossibility result

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

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

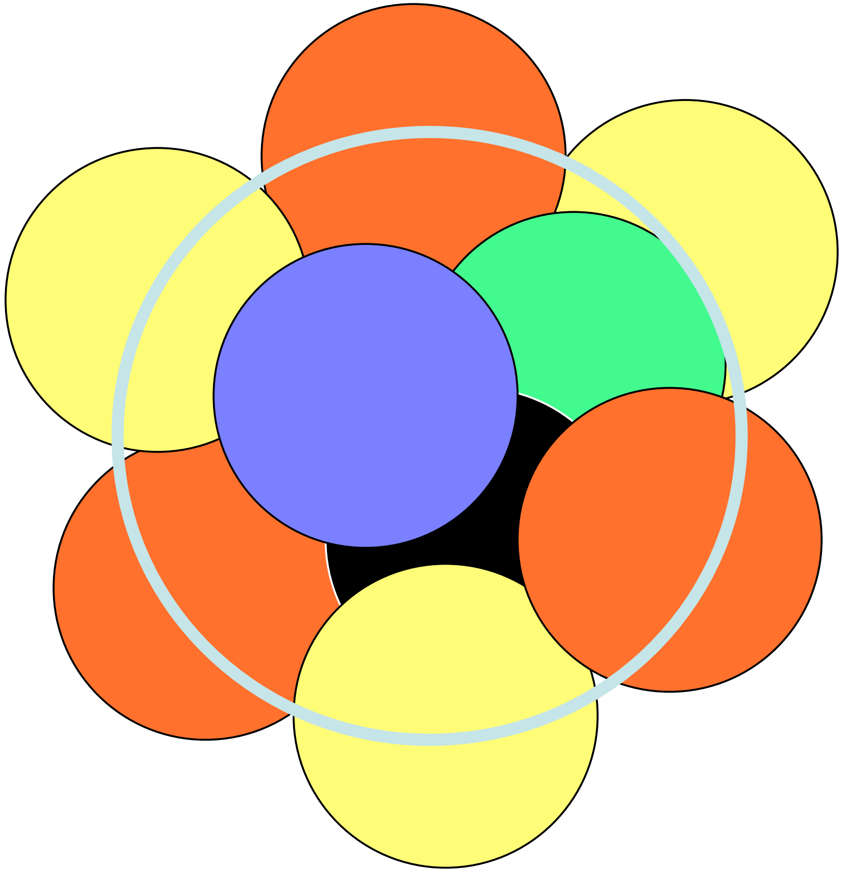
O(polylog(n))-navigable graphs

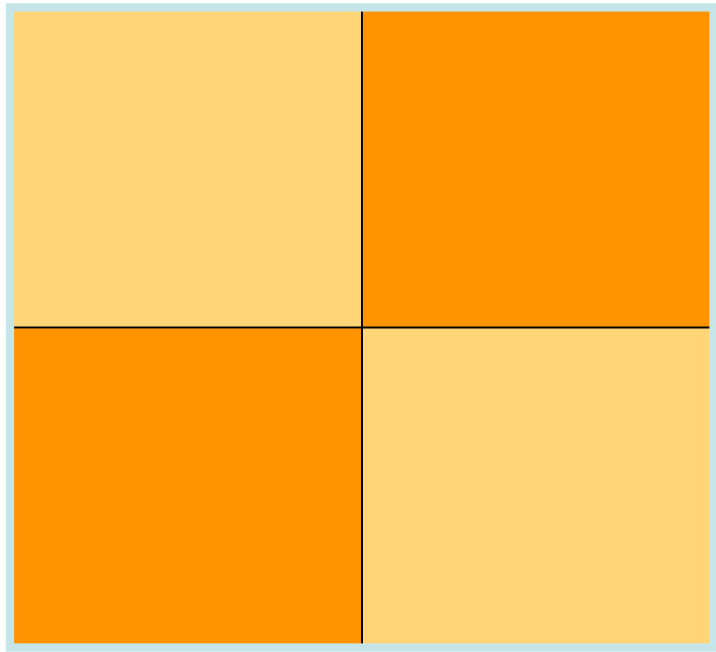
- Bounded growth graphs
  - Definition: $|B(x,2r)| \leq \rho\ |B(x,r)|$
  - Duchon, Hanusse, Lebhar, Schabanel (2005,2006)
- Bounded doubling dimension
  - Definition: DD d if every $B(x,2r)$ can be covered by at most 2^d balls of radius r
  - Slivkins (2005)
- Graphs of bounded treewidth
  - Fraigniaud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

Question

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

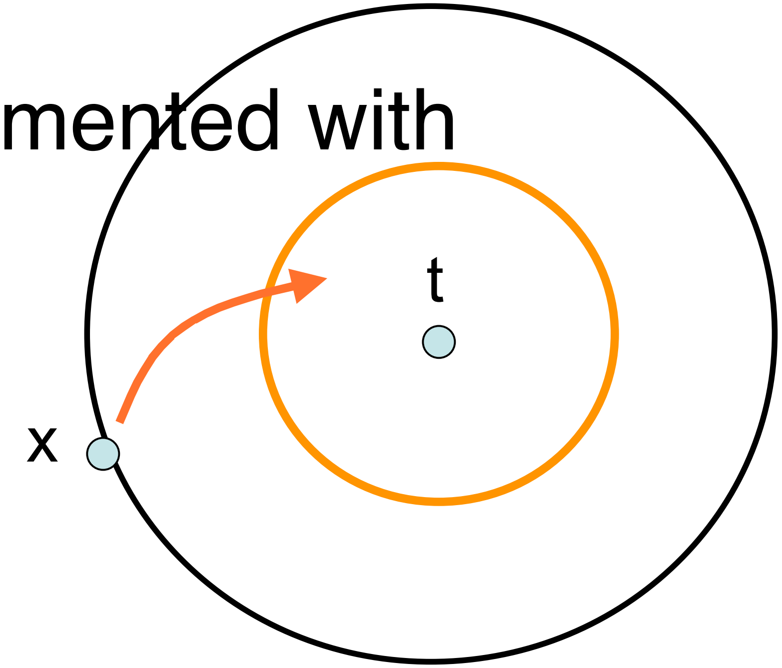
Doubling dimension





Svilkins' theorem

- Theorem: Any family of graphs with doubling dimension O(loglog n) is navigable.
- Proof: Graphs are augmented with
  - $dist_G(u,v) = r$
  - $prob(u \rightarrow v) \approx 1/|B(v,r)|$



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

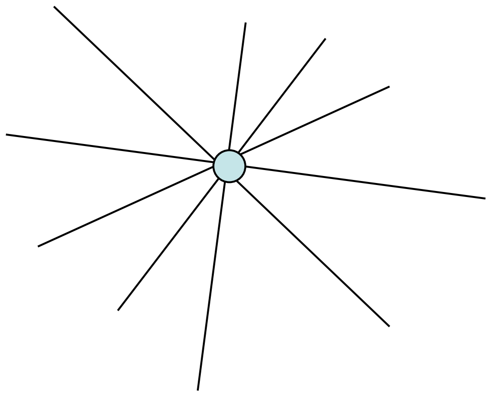
The graphs H_d with $n=p^d$ nodes

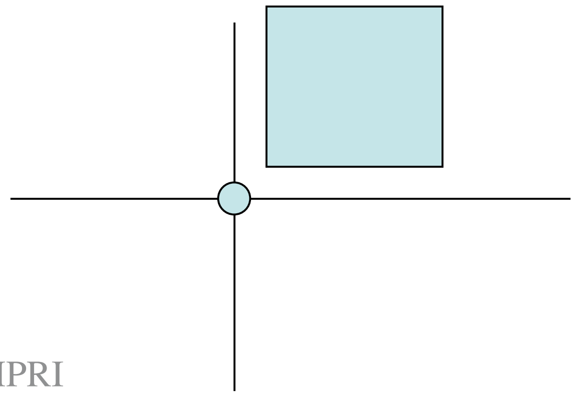
$x = x_1\, x_2 \dots x_d$
is connected to all nodes
$y = y_1\, y_2 \dots y_d$
such that $y_i = x_i + a_i$ where
$a_i \in \{-1,0,+1\}$

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_∞

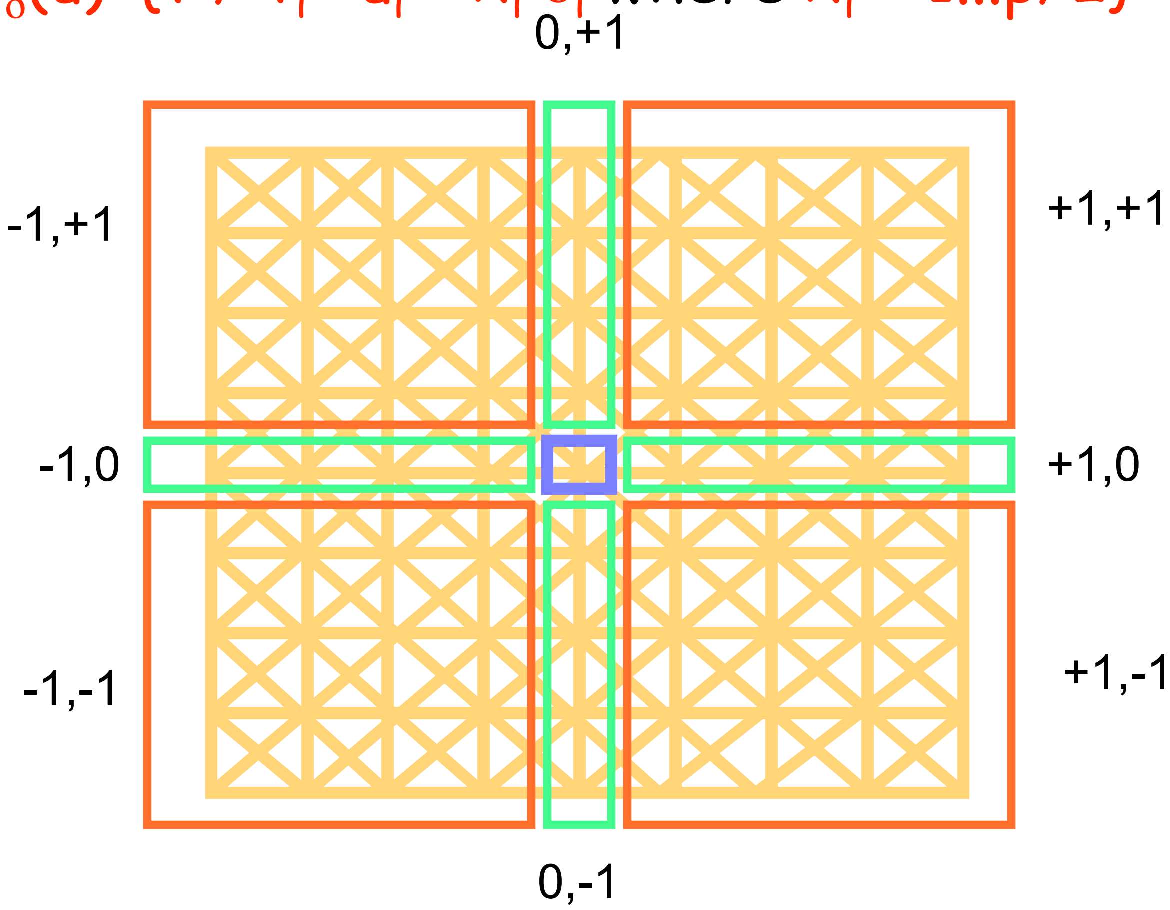




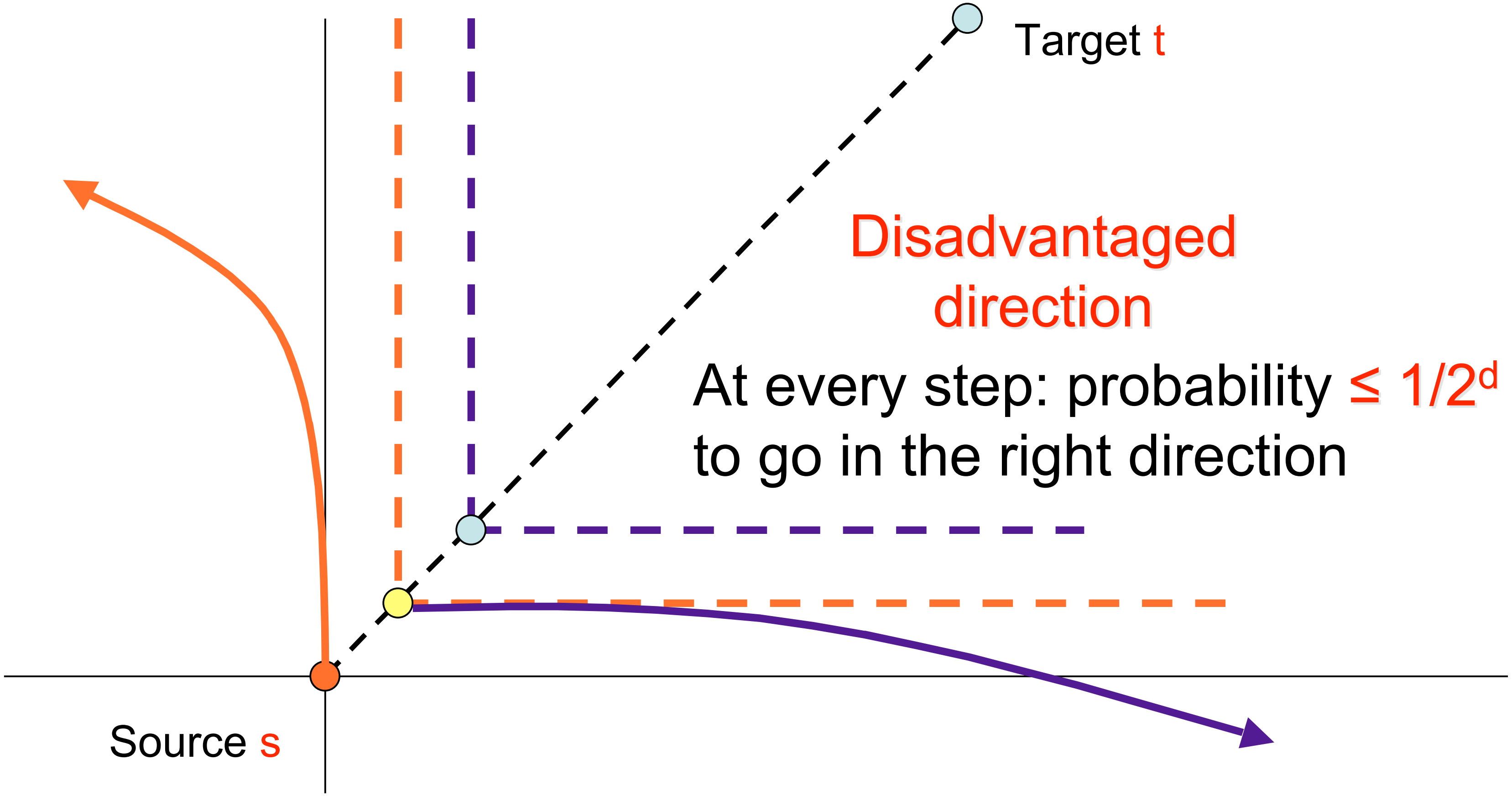
Directions

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

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

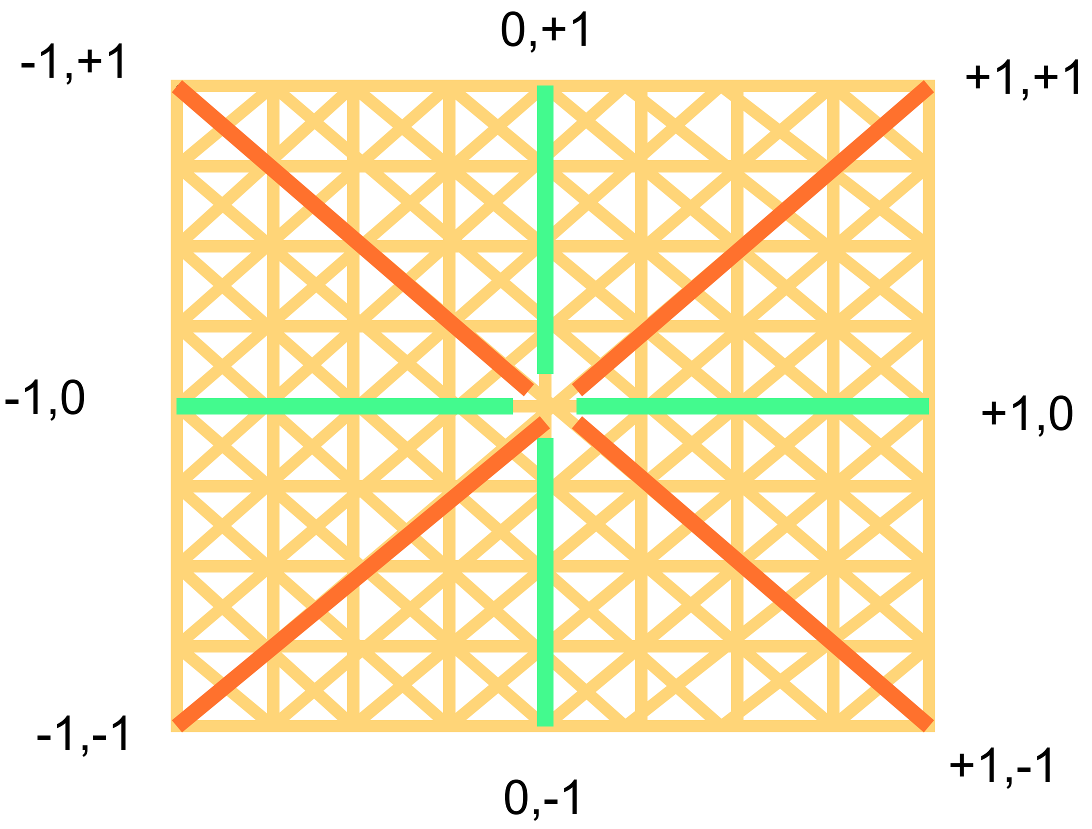


Case of symmetric distribution

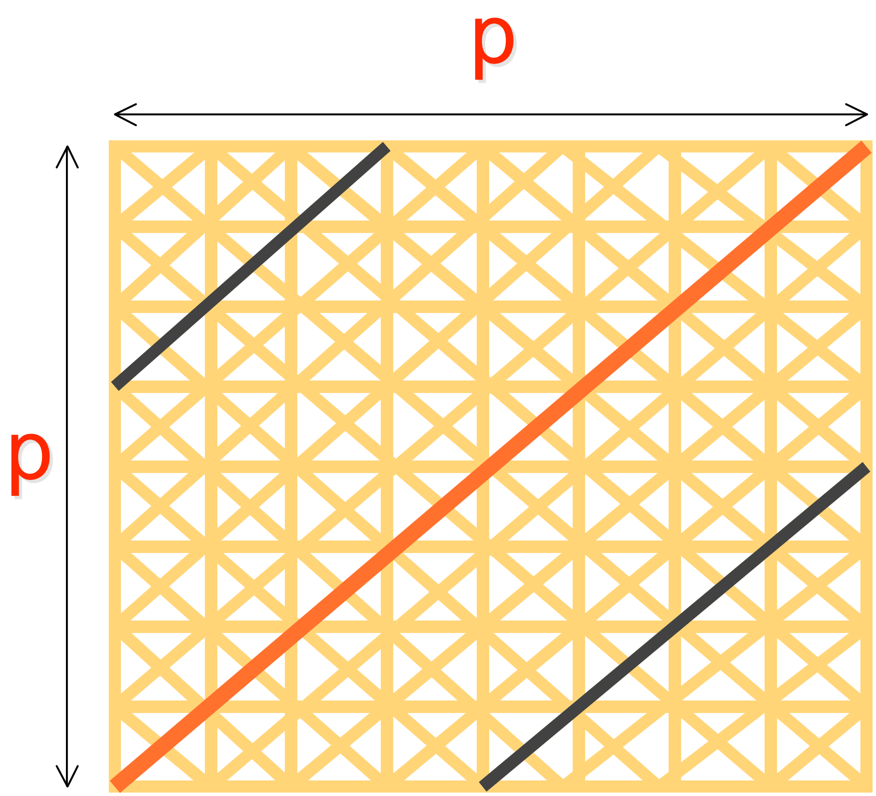


-- General case --

Diagonals

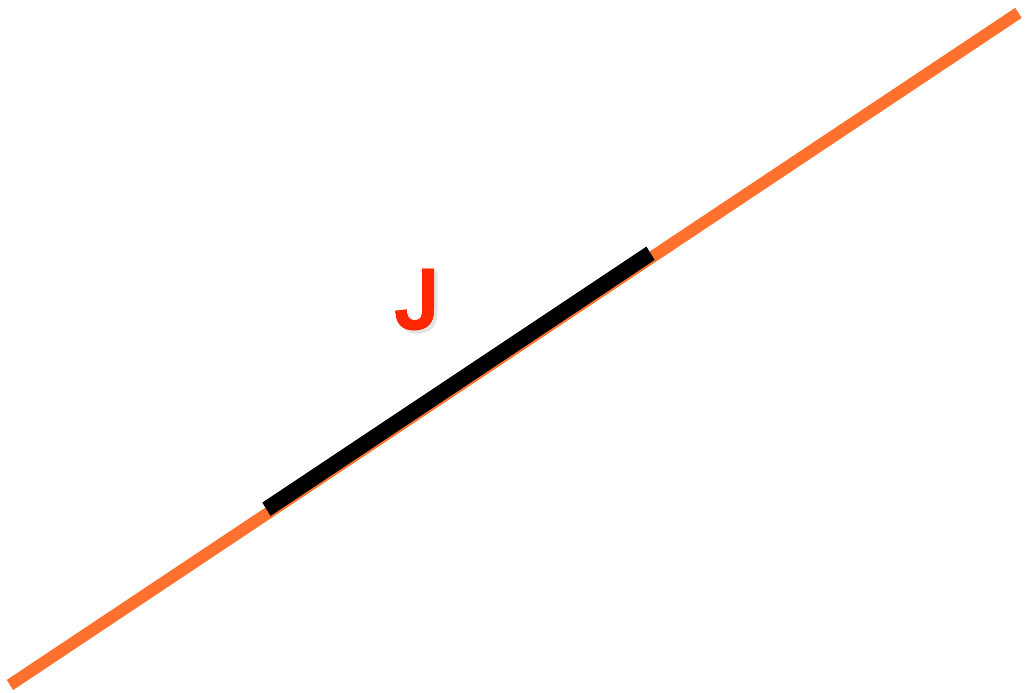


Lines

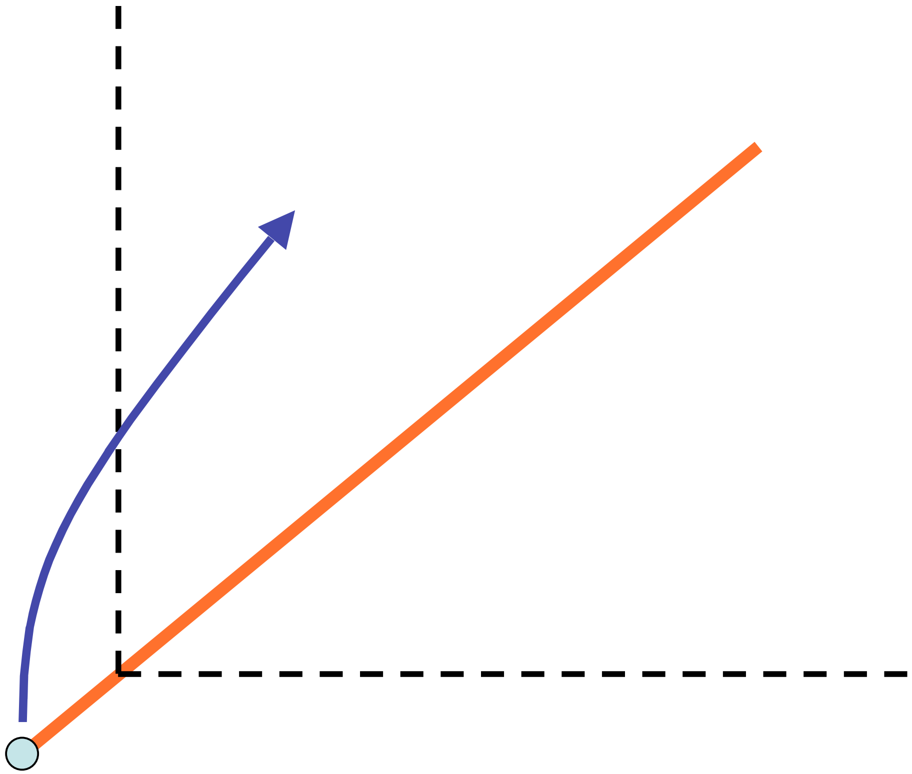


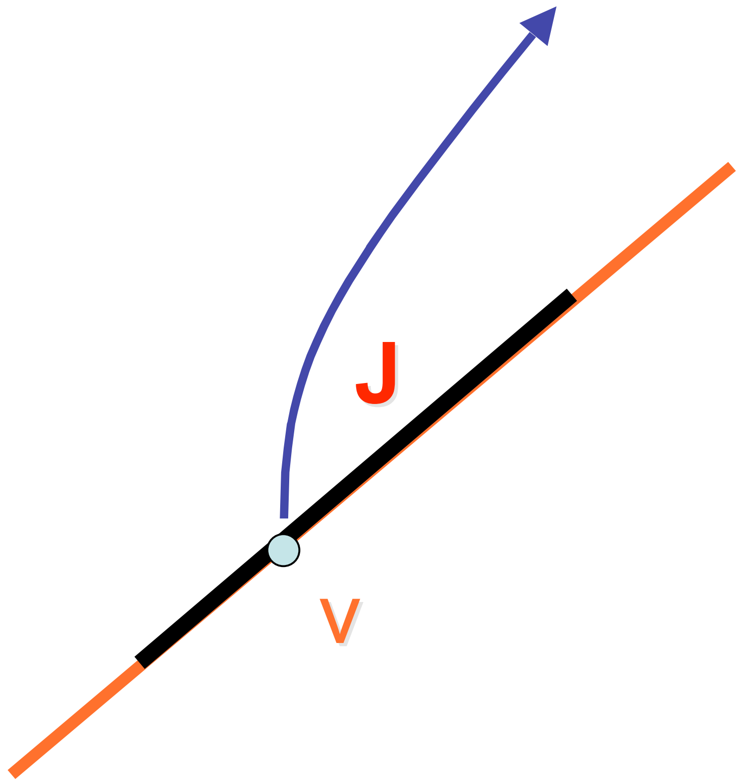
p lines in each direction

Intervals



Certificates



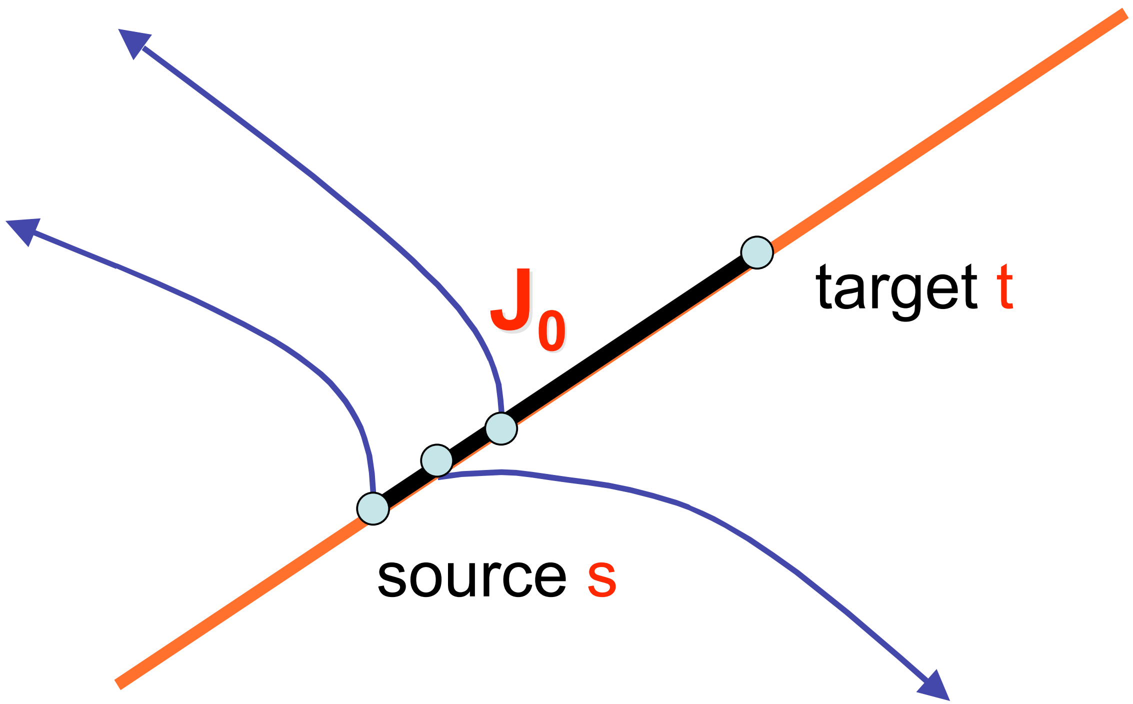


v is a certificate for J

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

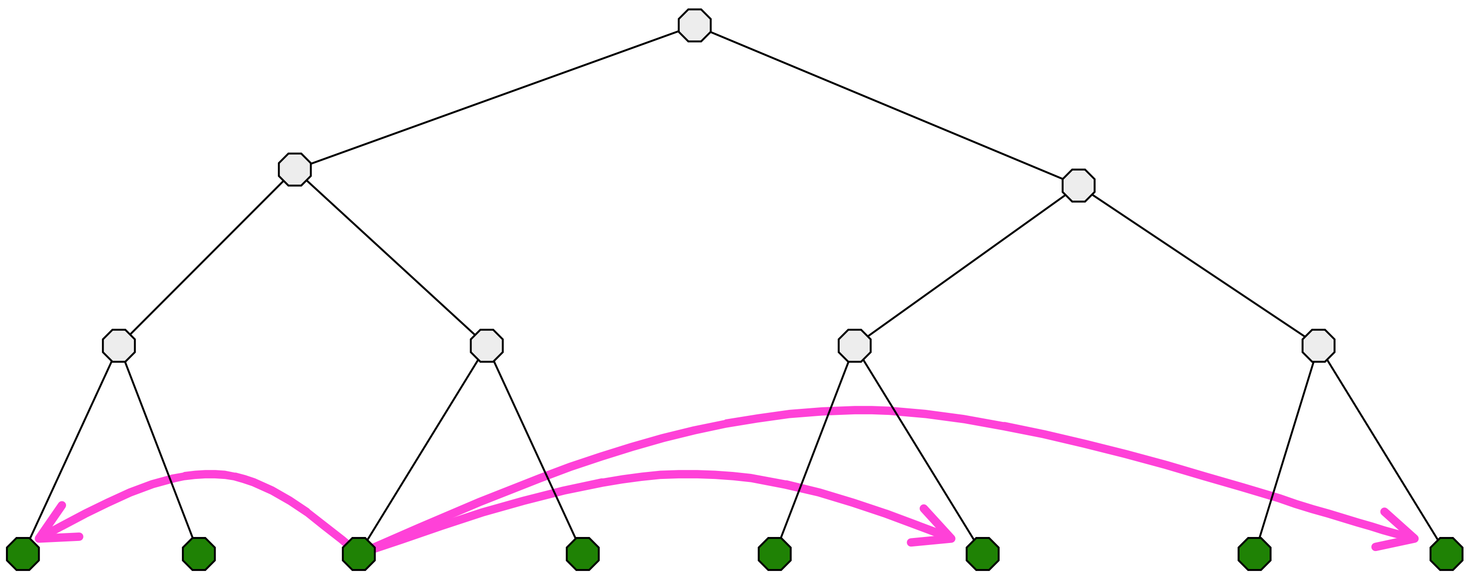
$$\Pr(J_0 \text{ has no certificate}) \geq 1/2$$

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

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

Hierarchical Models

Kleinberg's Hierarchical Model



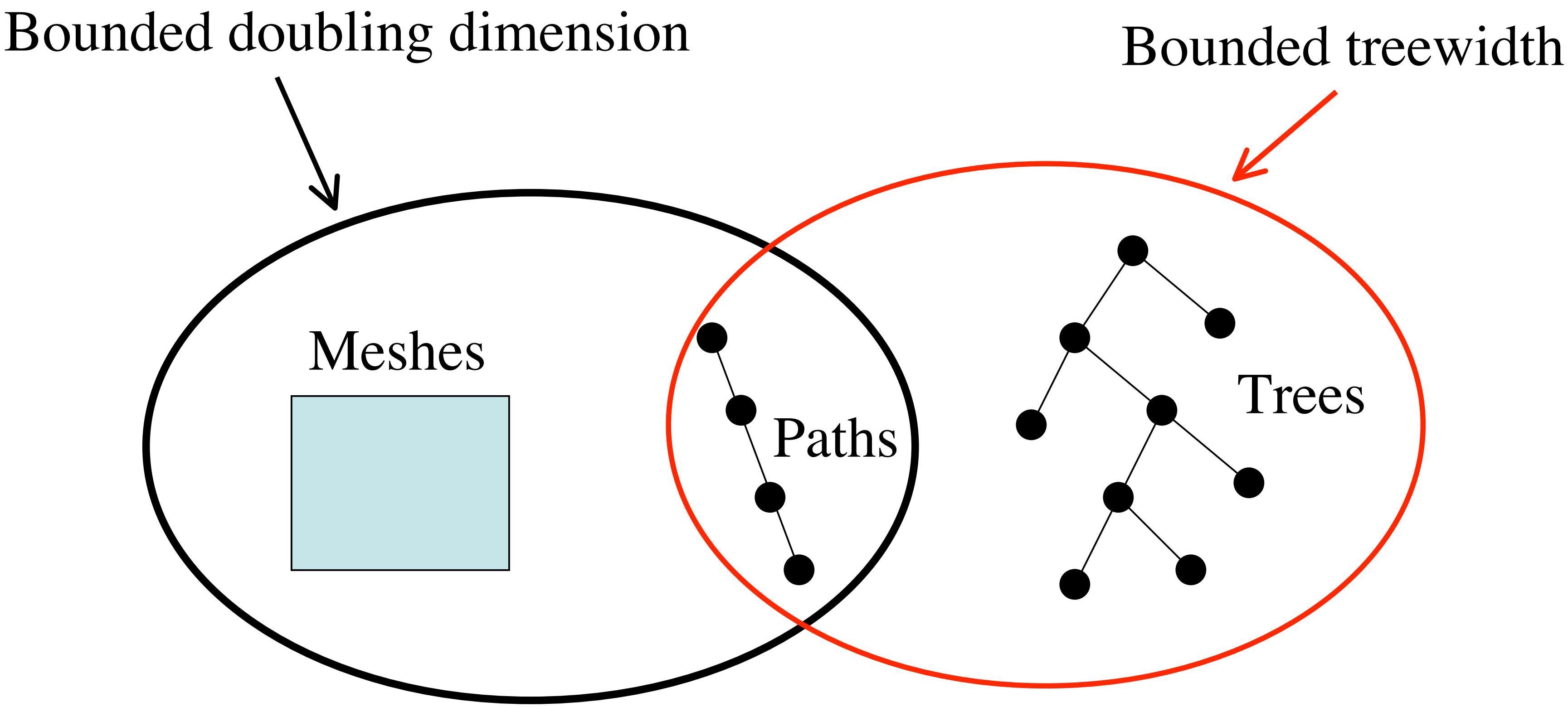
$\Theta(\log n)$ long links per node

Prob(x→y) ≈ height of their lowest common ancestor

Interleaved Hierarchies

- Many hierarchies:
  - place of living
  - professional activity
  - recreative activity
  - etc.
- Can we extract a “global” hierarchy reflecting all these interleaved hierarchies?

Graph classes

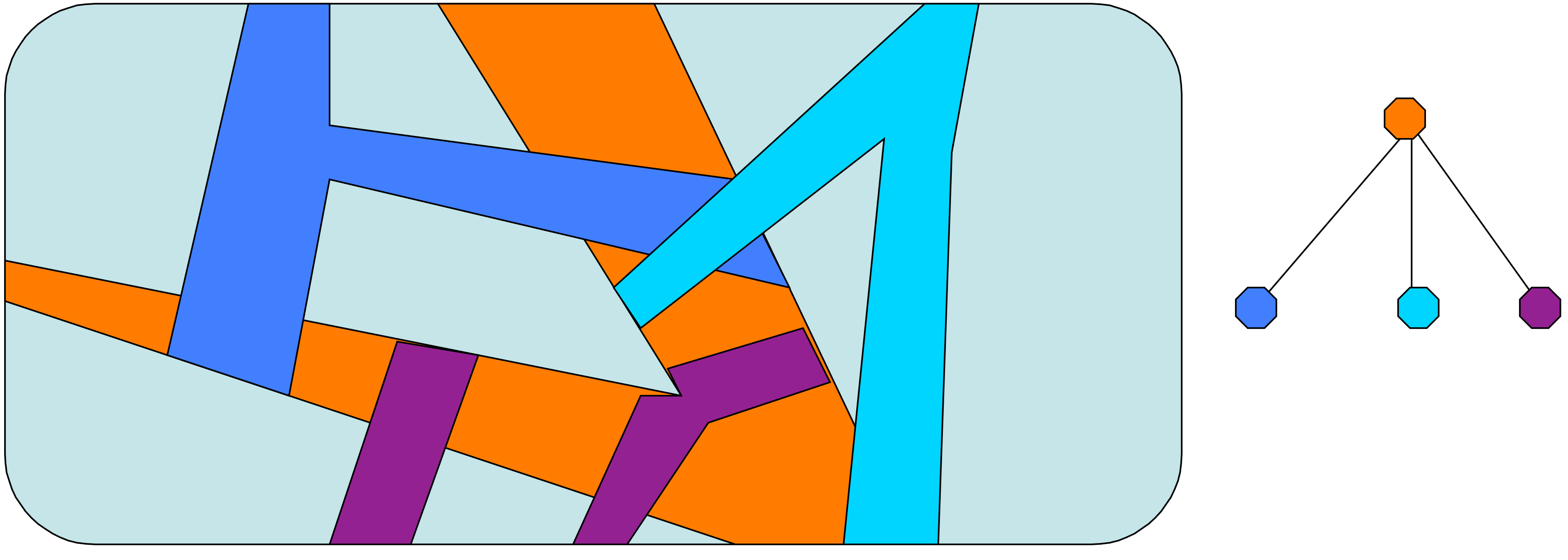


Tree-Decomposition

Definition: A tree-decomposition of a graph G is a pair (T,X) where T is a tree of node set I and X is a collection $\{X_i \subseteq V(G), i \in I\}$ such that

- $\bigcup_{i \in I} X_i = V(G)$
- $\forall\ e=\{x,y\} \in E(G), \exists\ i \in I\ /\ \{x,y\} \subseteq X_i$
- if $k \in I$ in on the path between i and j in T, then $X_i \cap X_j \subseteq X_k$

Recursive Separators



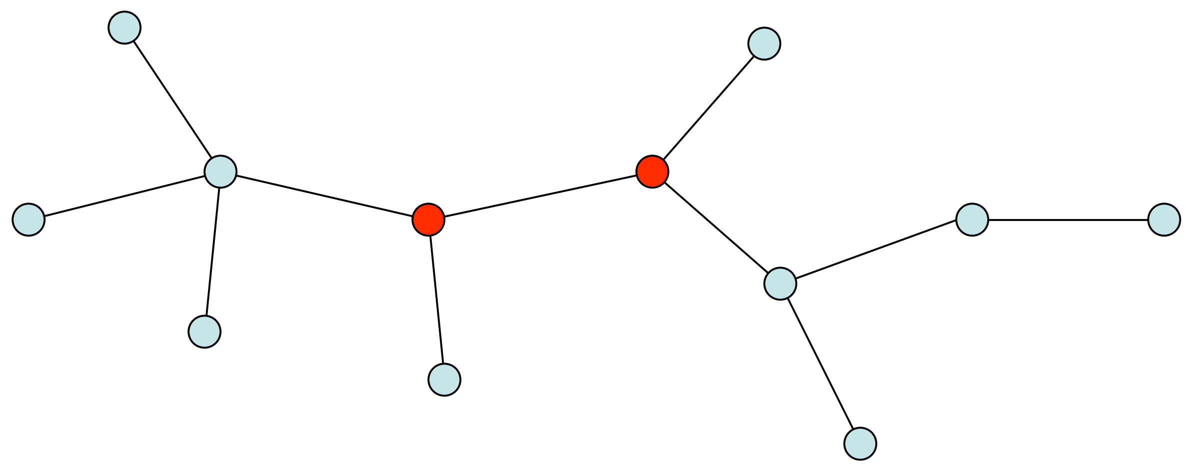
Treewidth

- The width of a tree-decomposition (T,X) is: $\text{width}(T,X) = \max_{i \in I} |X_i| - 1$
- The treewith of a graph G is the minimum width of any tree-decomposition (T,X) of G:

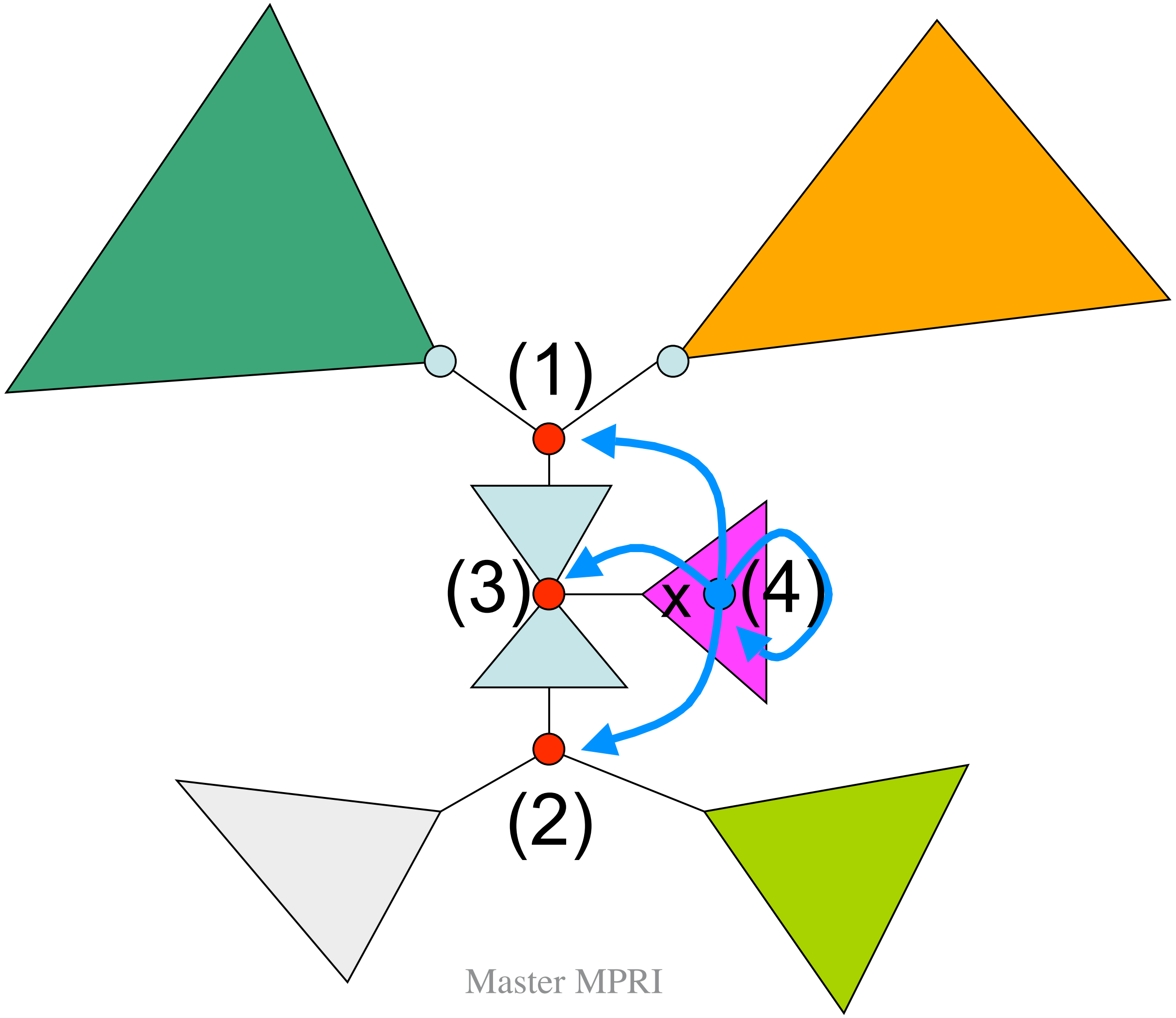
$$tw(G) = \min_{(T,X)} \text{width}(T,X)$$

Centroid

A centroid of an n-node tree T is a vertex v such that T-{v} is a forest of trees, each of at most n/2 vertices.



Tree-Decomposition Based Distribution



THEOREM

- For any n-node graph G of treewidth k, there exists a tree-decomposition based distribution D such that greedy routing in G+D performs in

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

  expected number of steps.
- Application: graphs of bounded treewidth.

Proof Sketch

- Let c be the centroid separating the current node x and t.
- It takes O(log n) expect. #steps to reach a node in c.
- The centroid c cannot be visited more than tw(G)+1 times
- There are ≤ log n levels of centroids

References

- Cf. Jon Kleinberg's Homepage

  http://www.cs.cornell.edu/home/kleinber/

- His talk at the International Congress of Mathematicians, 2006

  http://icm2006.org/AbsDef/Invited/15_kleinberg.pdf