Navigation in the Network of Individual Acquaintances

Pierre Fraigniaud
CNRS
University Paris Diderot

Outline of the talk

1. Observations
2. Model and analysis
3. Validation
4. Emergence

Statistical Properties of Interaction Networks

Interaction Networks

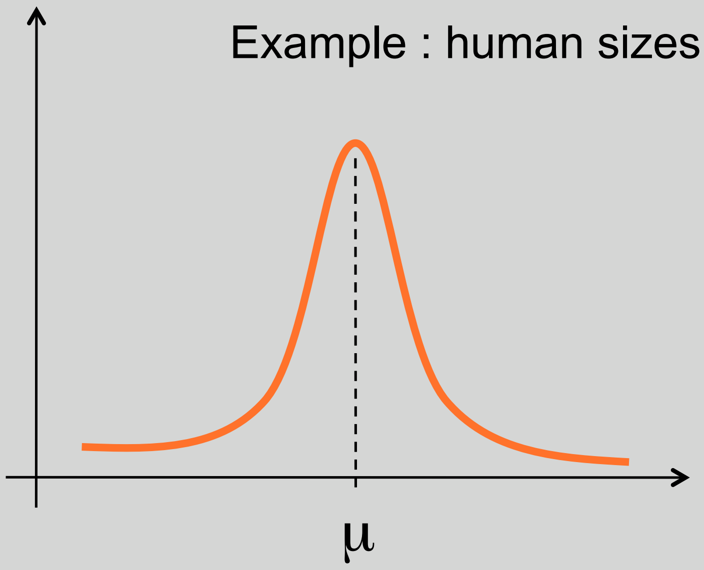
- Communication networks
  - Internet
  - Ad hoc and sensor networks
- Distributed networks
  - The Web
  - P2P networks (the unstructured ones)
- Social network
  - Acquaintance
  - Mail exchanges
- Biology, linguistics, etc.

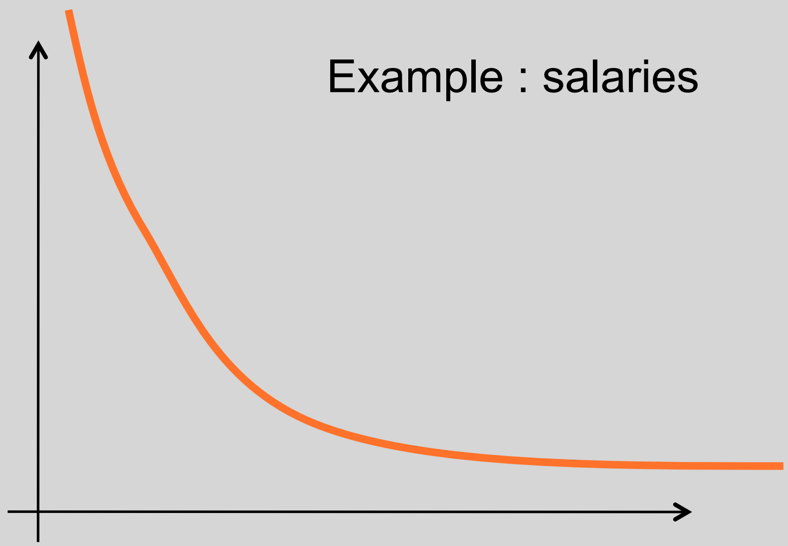
Common statistical properties

- 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.

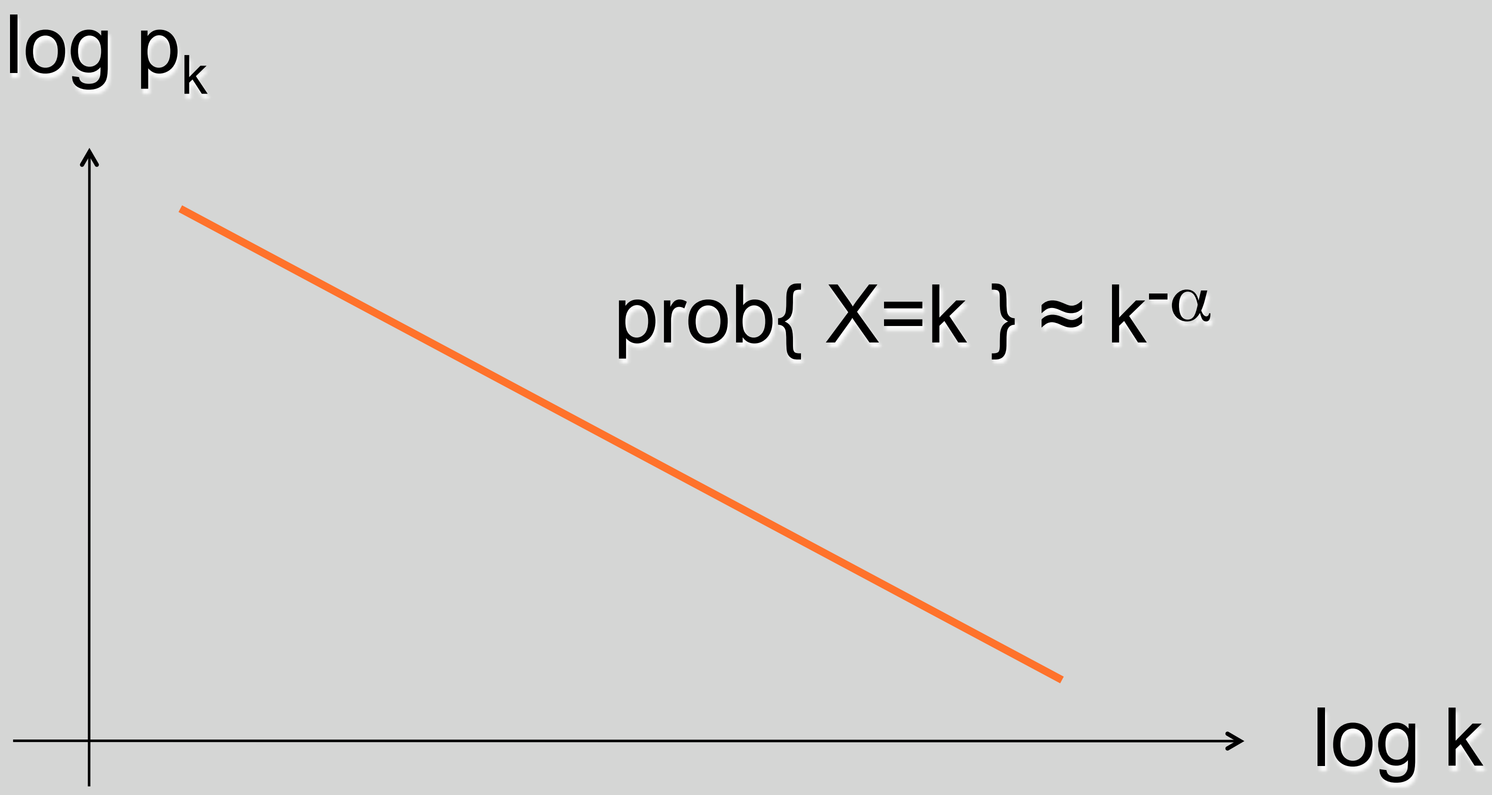
    p = clustering coefficient
- "Scale free" properties:
  - Heavy tailed probability distributions (e.g., of the degrees)

Gaussian vs. Heavy tail





Power law



$$\text{prob}\{ X=k \} \approx k^{-\alpha}$$

Random graphs vs. Interaction networks

- Random graphs ($p = c \ln(n)/n$ with $c>1$):
  - 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?

Navigability: Experiments and models

Milgram Experiment (60's)

Name: Paul Johnson
Professional occupation: Broker
Place of living: Boston, Massachusetts

Six degrees of separation

- Reproducible experiment

  [Dodds, Muhamad, and Watts, 2003]
  - Uses e-mail, and search for several targets
  - #successful chains depend on the target, but the average length of the chains seems independent from the target
- Structure of social networks

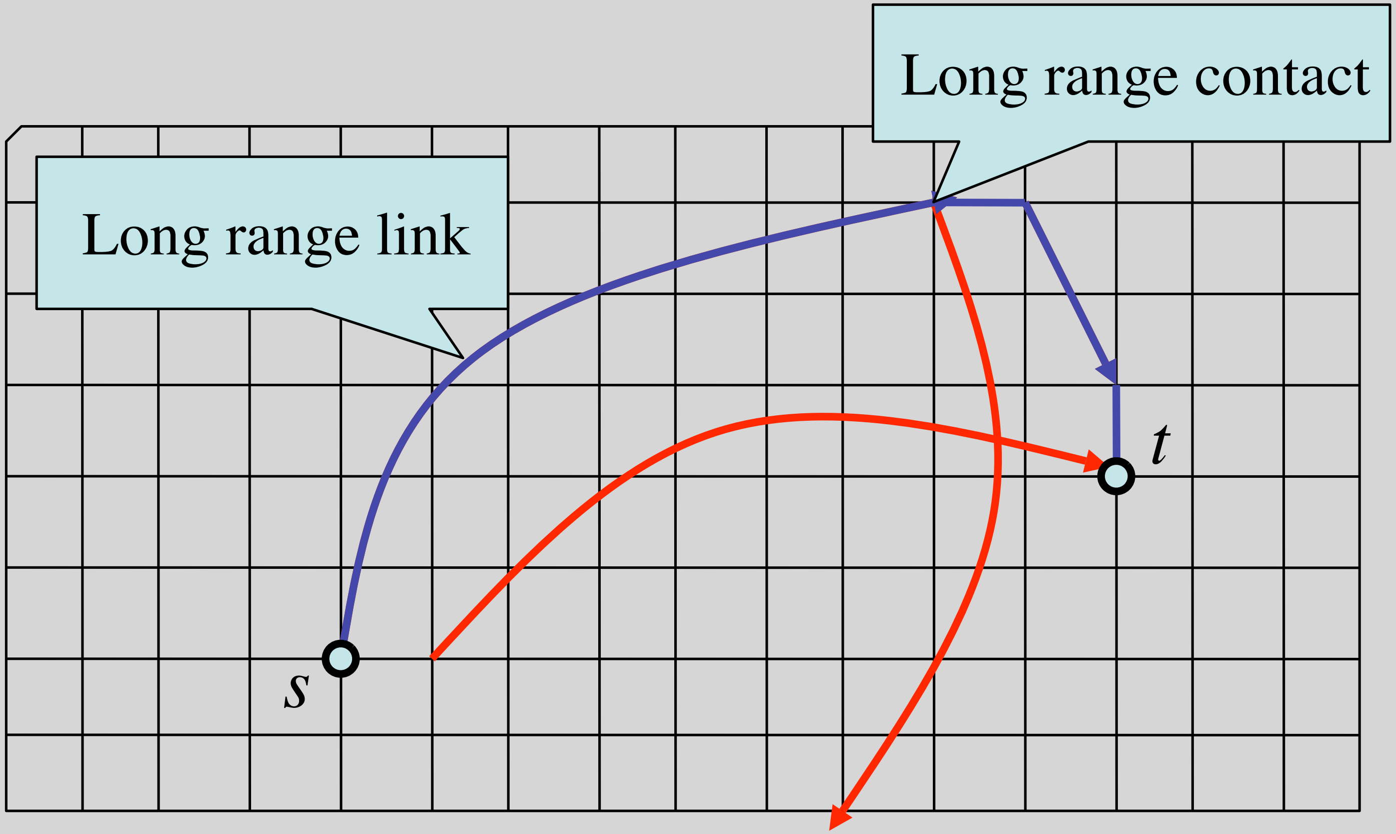
  [Liben-Nowell, Novak, Kumar, Raghavan, Tomkins, 2005]
  - for 2/3 of the friendships, the probability of befriending a particular person is inversely proportional to the number of geographically closer people

Quoting Jon Kleinberg (2000)

1. Why should there exist short chains of acquaintances linking together arbitrary pairs of strangers?
2. Why should arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?

Question 2 raises issues that lie truly beyond the scope of Question 1

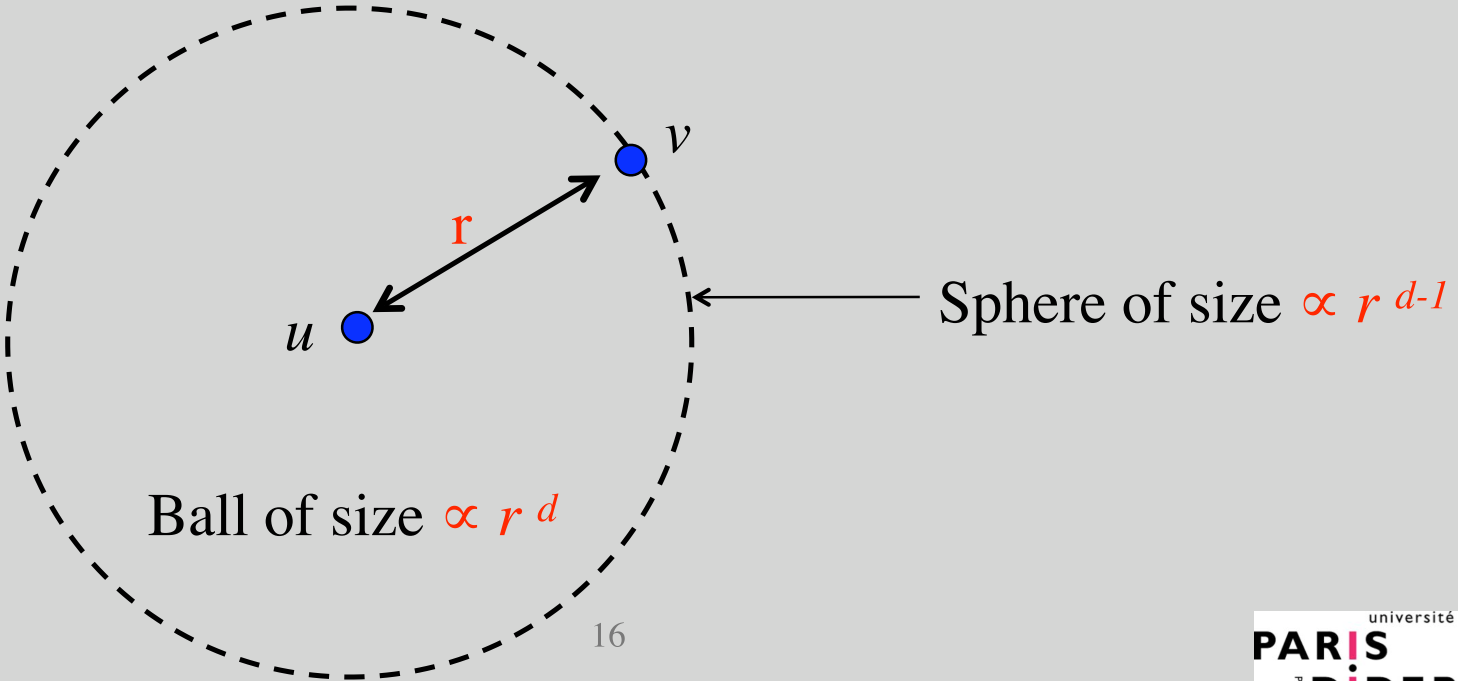
Greedy routing in augmented *d*-dimensional lattices



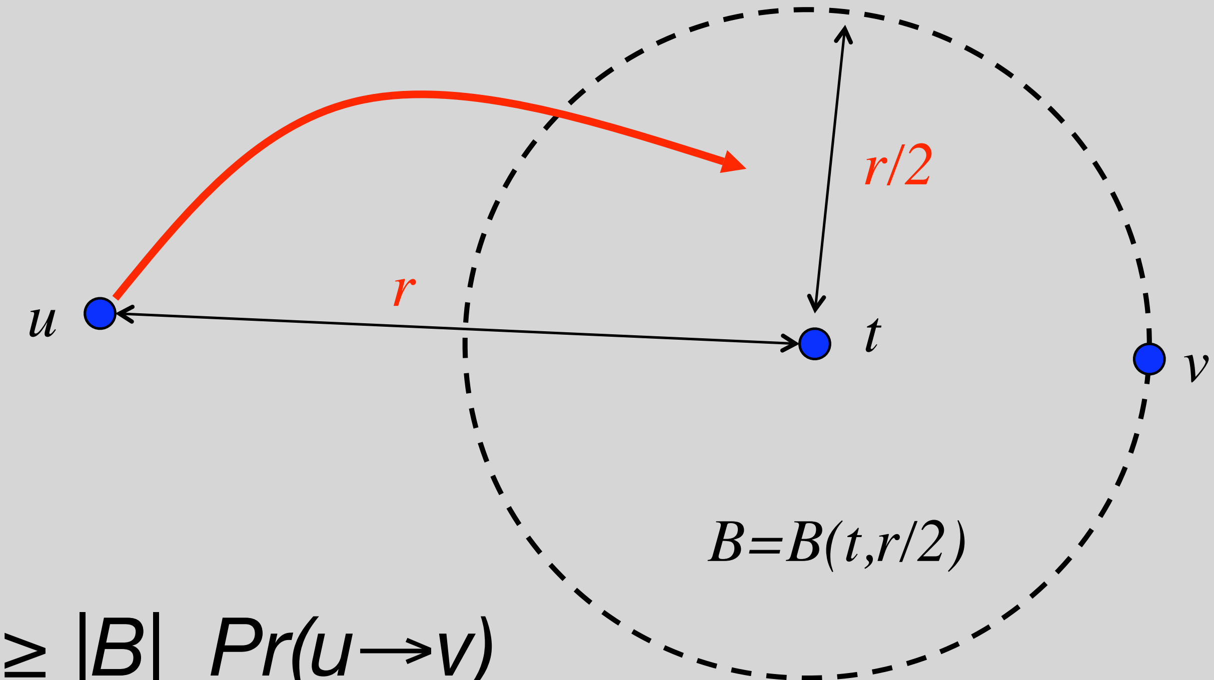
Trials of the long range links

d-Harmonic distribution:

$$Pr(u \rightarrow v) \propto 1/(\log n * dist(u,v)^{d})$$



Halving the distance



$Pr(u \rightarrow B) \geq |B| \;\; Pr(u \rightarrow v)$

$\geq (r/2)^d \;\; 1/(\log n * (3r/2)^d)$

$\geq 1/\log n$

Kleinberg Theorem

In the *d*-dimensional lattice augmented using the *d'*-harmonic distribution:

– If $d'=d$, then the expected #steps of greedy routing from any source *s* to any target *t* is $O(\log^2 n)$ for any *s* and *t*.

– If $d' \neq d$, then the average expected #steps of greedy routing is $\Omega(n^{\varepsilon})$, $\varepsilon > 0$.

Exercise 1

Prove that, in the 2-dimensional $n \times n$ lattice augmented using the r-harmonic distribution with $r \neq 2$, the average expected #steps of greedy routing is $\Omega(n^{\varepsilon})$, $\varepsilon > 0$. (ε depends on r)

Correction 1 (sketch)

- **Case $0 \leq r < 2$**
  - $K = \Sigma_{v \neq u}\ 1/\mathrm{dist}(u,v)^r \geq \Omega(n^{2-r})$
  - Pick s,t u.a.r.; Δ ="dist(s,t) ≥ n/4"; Prob[Δ] ≥ ½
  - Set U = $B(t, n^{(2-r)/3}) \Rightarrow |U| \leq O(n^{2(2-r)/3})$
  - E = "within $\varepsilon n^{(2-r)/3}$ steps routing reaches a node ≠ t, with long-range contact in U" for some $\varepsilon > 0$ (to be fixed).
  - $E = \cup_i E_i$ where E_i is "E at step i"
  - $\mathrm{Prob}[E_i] \leq O(|U|/K) \leq O(1/n^{(2-r)/3})$
  - $\mathrm{Prob}[E] \leq \varepsilon$

Correction 1 (Continued)

- Pr[(not Δ) or E] ≤ ε + ½
- Pr[Δ and (not E)] ≥ ½ - ε
- X = #steps to reach t
- F = "routing reaches t in less than $\varepsilon n^{(2-r)/3}$ steps"
- We have: (Δ and (not E)) ⇒ not F
- Hence: Pr[F | Δ and (not E)] = 0
- Hence: **E**(X | Δ and (not E)) ≥ $\varepsilon n^{(2-r)/3}$
- **E**X ≥ **E**(X | Δ and (not E)) Pr[Δ and (not E)]
- Hence: **E**X ≥ (½ - ε)$\varepsilon n^{(2-r)/3}$ ∎

Correction 1 (continued)

- **Case r > 2**
  - based on the fact that long-range links are “too short”.

Local Routing Strategies in Lattices

Decentralized routing

Theorem [Manku, Naor, Wieder, 2004]

Greedy is optimal among all decentralized algorithms

Theorem [Lebhar, Schabanel, 2004]

There exists an ***exploration*** algorithm that finds routes of expected length $O(\log n * \log^2 \log n)$ between any pair of nodes, after exploring $O(\log^2 n)$ links.

Two-step greedy routing

Every node "knows" its neighbors' neighborhood (NoN)

Theorem

[Coppersmith, Gamarnik, Sviridenko, 2002]

[Manku, Naor, Wieder, 2004]

With *log n* long range contacts per node, the expected #steps of ***NoN*** from a source *s* to a target *t* is *O(log n/loglog n)* for any *s* and *t*.

Routing with partial knowledge

Every node "knows" the long range links of its log n closest neighbors in the mesh

Theorem

[Martel, Nguyen, 2004] Non-oblivious routing

[F., Gavoille, Paul, 2004] Oblivious routing

With *1* long range contact per node, the expected #steps of ***indirect greedy*** routing from a source *s* to a target *t* is $O(\log^{1+1/d} n)$ for any *s* and *t*.

Eclecticism Shrinks the World!

Augmented graph model

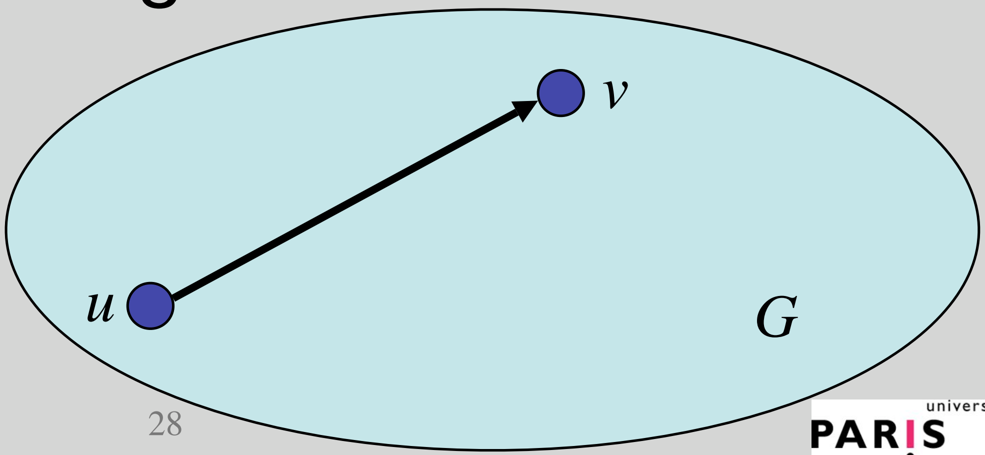
Model (G,φ)

- A graph G, called "base graph"
- A collection of probability distributions:

$$\varphi = \{\varphi_u : u \in V(G)\}$$

called "augmenting distribution"

$\varphi_u(v) = \Pr(u \rightarrow v)$



Greedy diameter of (G,φ)

- Greedy routing:
  - Every intermediate node selects its neighbor (possibly its long range contact) that is closer to the target in G, and forwards to it.
- Greedy diameter:
  - $\boldsymbol{E}_\varphi(s,t)$ = expected #steps of greedy routing from s to t in G augmented by φ
  - $gd(G,\varphi) = max_{(s,t) \in V(G)xV(G)} \boldsymbol{E}_\varphi(s,t)$
- Remark: $gd(G,\varphi) \geq diameter(G,\varphi)$

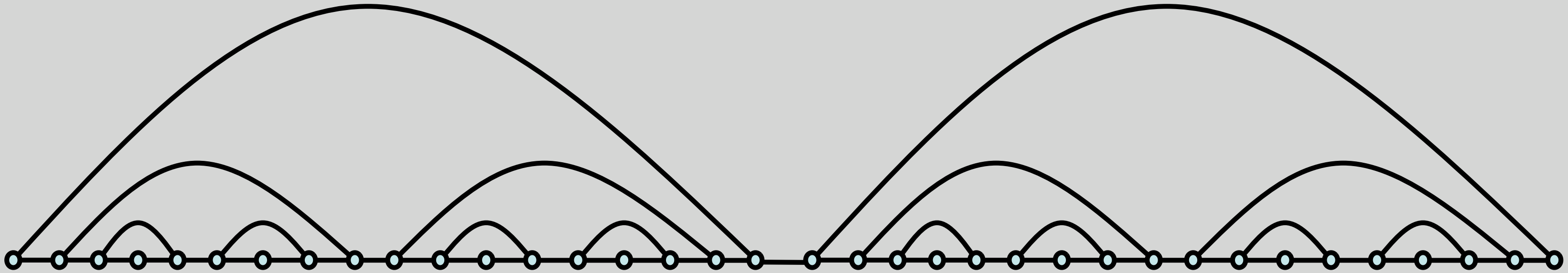
Examples

- $gd(M_{n,d}, H_d) = O(\log^2 n)$

  [Kleiberg 2000]

- $gd(P_n, \varphi) = O(\log n)$

  [Flammini et al. 2005, Abraham et al. 2005]



Complexity

Theorem

[Flammini, Moscardelli, Navarra, Perennes, 2005]

The following problem is NP-complete:

- Input: A graph G
- Question: Is there a φ such that $gd(G,\varphi) \leq 2$?

Polylog-navigable graphs

Ball growth

Definition

A graph G has ball growth $\leq \rho$ if for any node x and any radius r, $|B(x,2r)| \leq \rho\, |B(x,r)|$

Theorem

[Duchon, Hanusse, Lebhar, Schabanel, 2005, 2006]

For any graph G of ball growth $\leq \rho$ there exists φ such that $gd(G,\varphi) \leq polylog\ n$

Proof.

Density-based distribution:

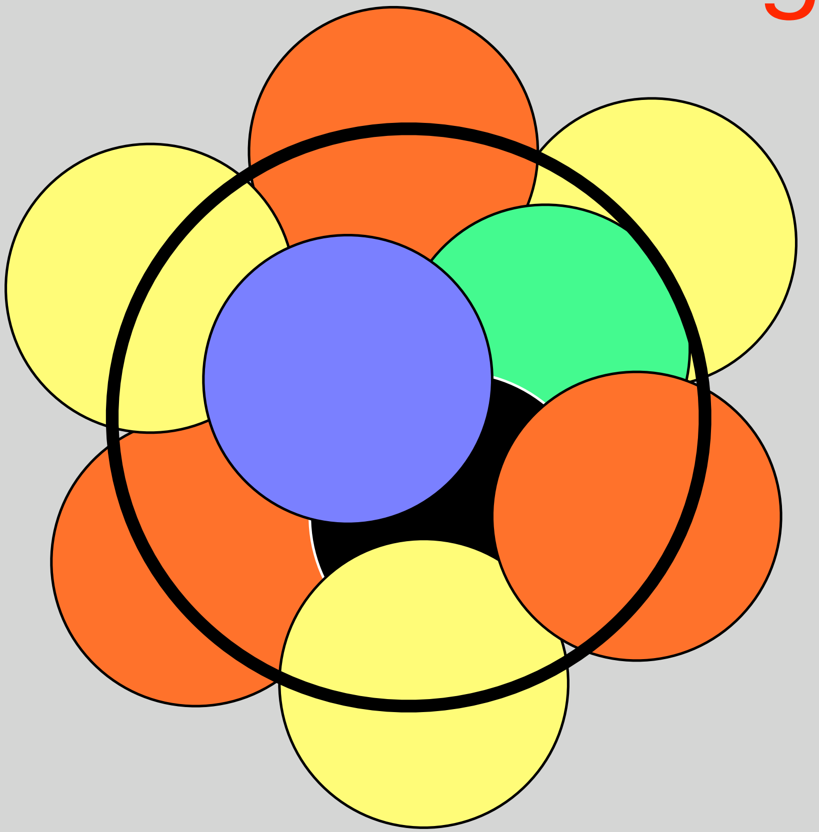
$$\text{If } dist_G(u,v) = r \text{ then } \varphi_u(v) \approx 1/|B(u,r)|$$

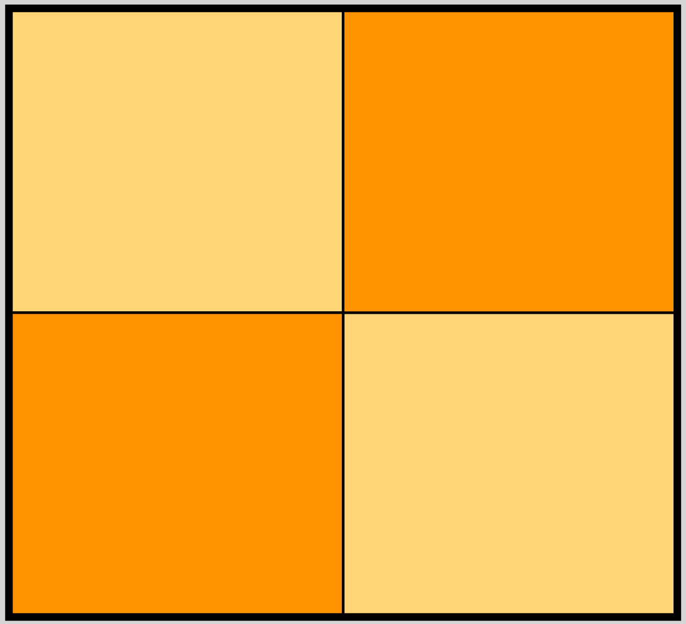
Exercise 2

Using the density-based distribution, prove that for any graph G of ball growth $\leq \rho$ there exists φ such that

$$gd(G,\varphi) \leq polylog\ n$$

Doubling dimension





Doubling dimension

Definition

A graph G has doubling dimension $\leq d$ if every ball $B(x,2r)$ can be covered by $\leq 2^d$ balls of radius r

Theorem [Slivkins, 2005]

For any graph G of doubling dimension $O(\log\log n)$ there exists φ such that $gd(G,\varphi) \leq \mathrm{polylog}\ n$

Proof.

For any graph G of doubling dimension d there exists a measure μ such that the node-weighted graph (G,μ) has a ball growth at most $O(d)$.

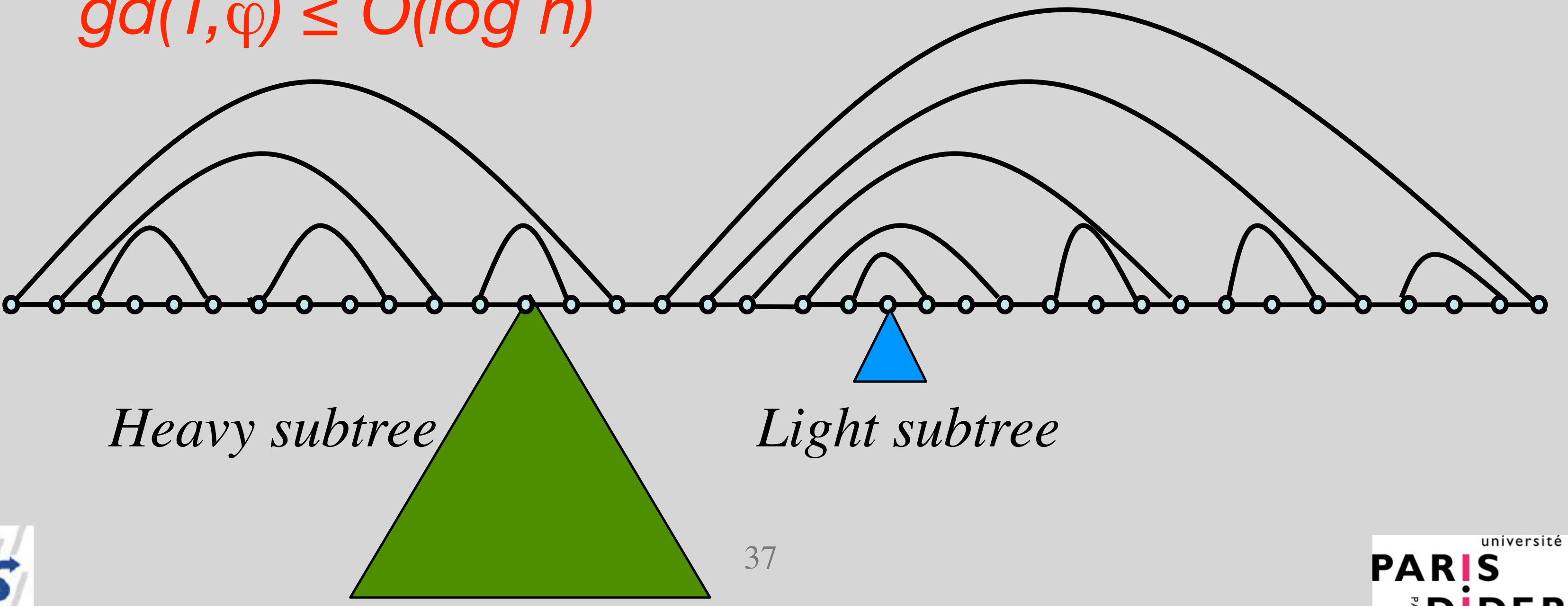
Trees

Theorem

[Flammini, Moscardelli, Navarra, Perennes, 2005]

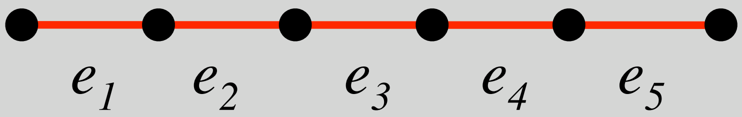
For any tree T there exists φ such that

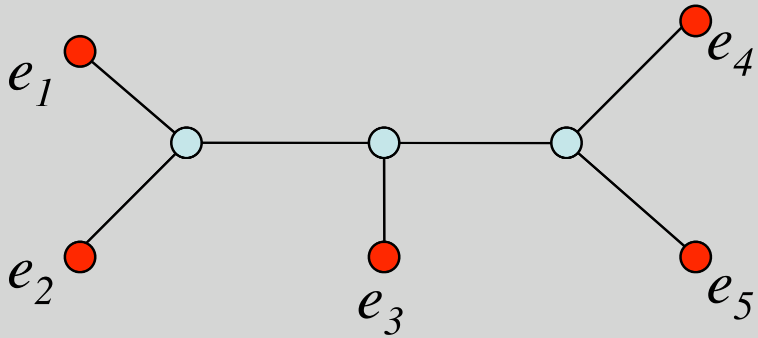
$gd(T,\varphi) \leq O(\log n)$



Graphs of bounded width
(treewidth, branchwidth, etc.)

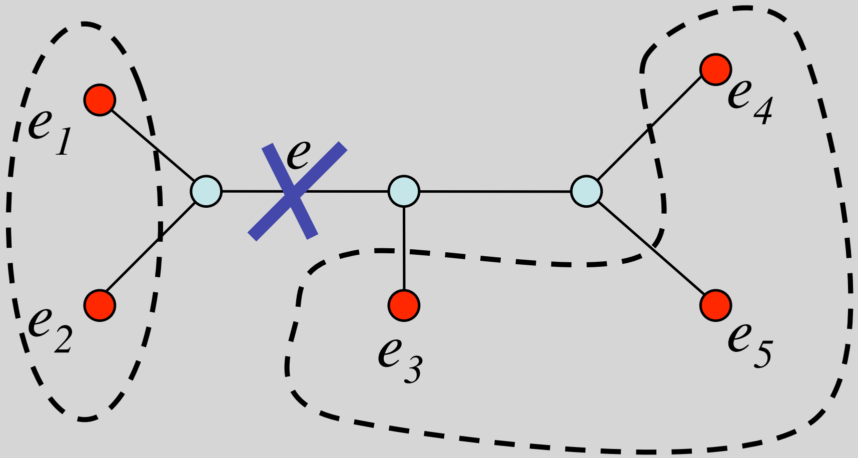
- Branch decomposition (T,π) of G:
  - A tree T with all internal nodes of degree 3
  - A one-to-one correspondence π between the leaves of T and $E(G)$.

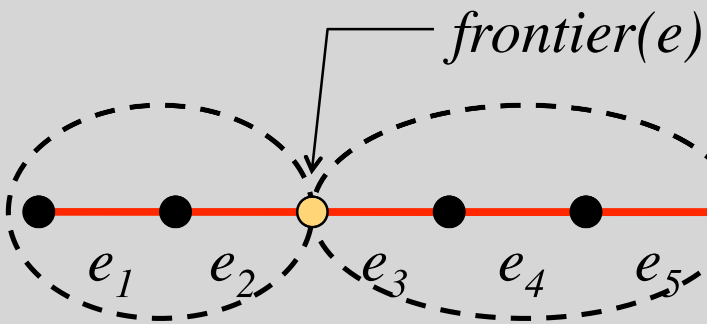




Branchwidth

- width(T,π) = $\max_{e \in E(T)}$ |*frontier(e)*|

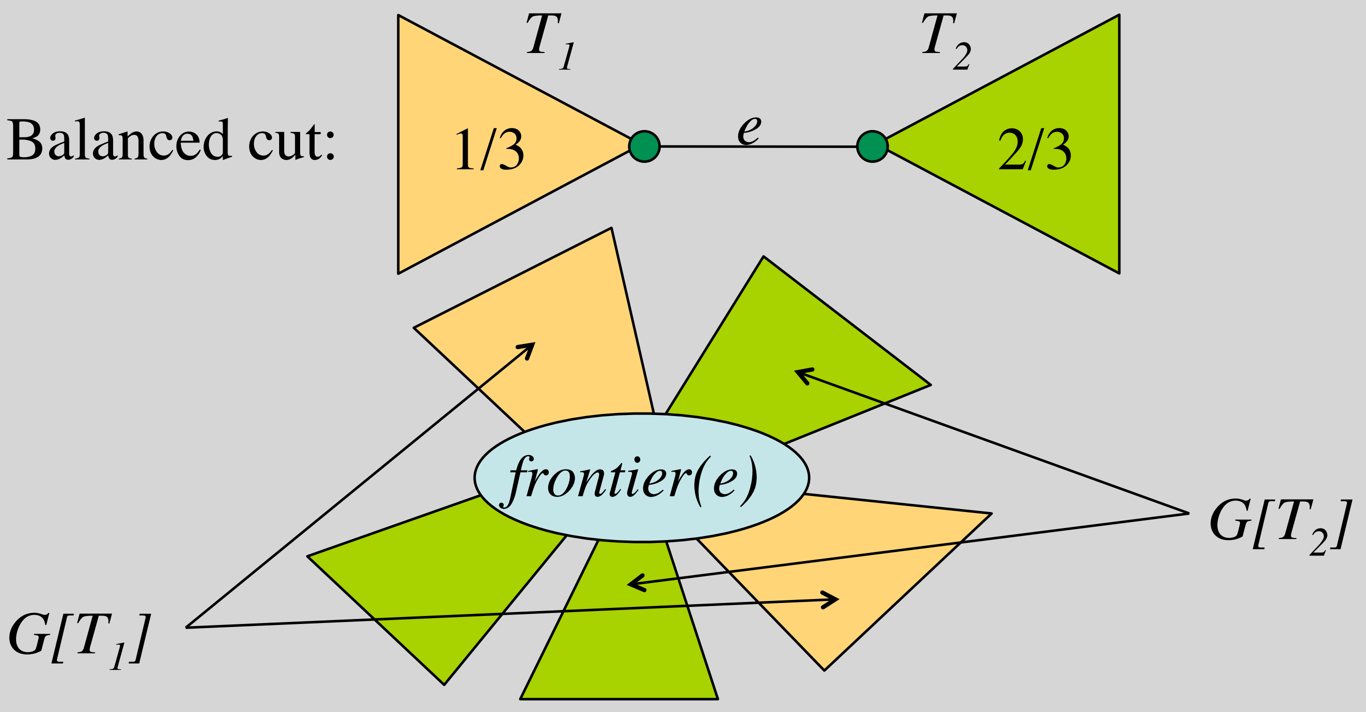




- bw(G) = $\min_{(T,\pi)}$ width(T,π)

Recursive augmentation

Balanced cut:



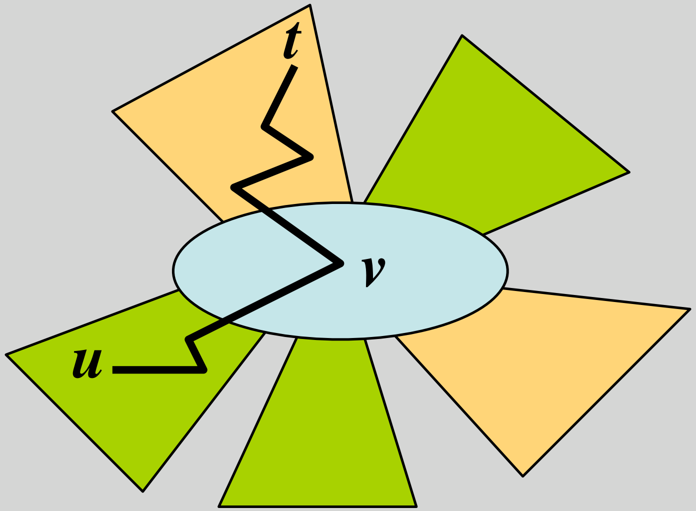
Apply $\Theta(\log n)$ times recursively in T_1 and T_2

Augmenting distribution φ :

(1) selects a "level" in $\{1,2,\dots,\log n\}$

(2) selects its long range contact in the frontier of this level

Greedy routing



$\varphi_u(v) \approx 1/\log n * 1/bw(G)$

$\Rightarrow E(s,t) \leq O(bw(G) \log^2 n)$

Theorem [F., 2005]

If G is of bounded branchwidth then

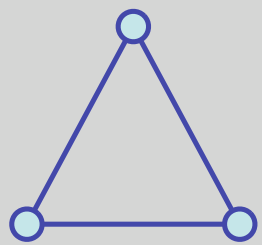
$gd(G,\varphi) \leq O(\log^2 n)$

Excluding a minor

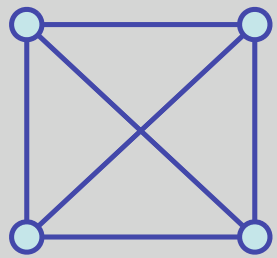
- A graph H is a minor of G if it can be obtained from G by a sequence of:
  - edge deletion
  - node deletion
  - edge contraction
- A graph family $\mathcal{G}$ is H-minor free if for any $G \in \mathcal{G}$ we have H **not** a minor of G

Examples

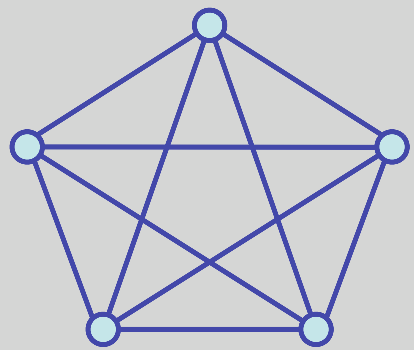
- Trees = K_3-free



- Series-Parallel = K_4-free



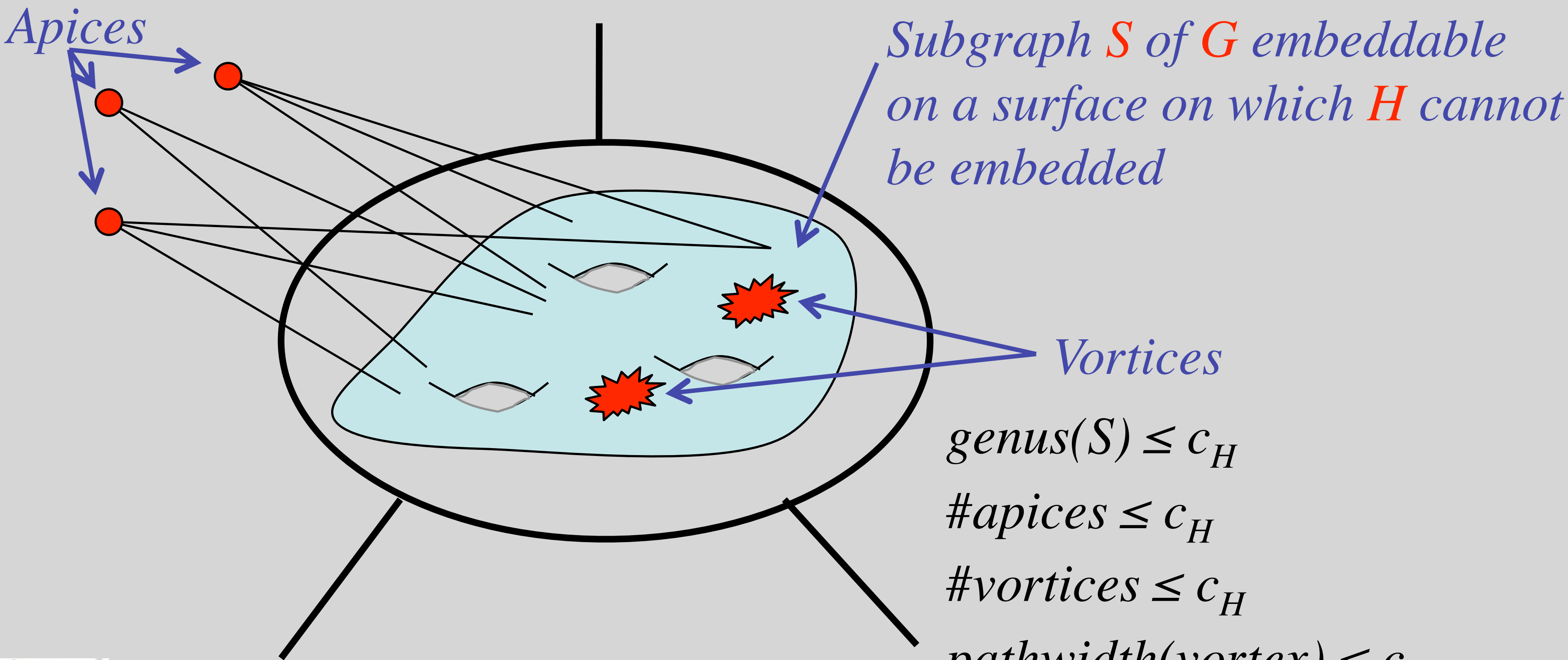
- Planar graphs ⊂ K_5-free



Characterization

[Robertson and Seymour, 1994]

Every *H*-minor free graph *G* as a tree-decomposition:



$genus(S) \leq c_H$

$\#apices \leq c_H$

$\#vortices \leq c_H$

$pathwidth(vortex) \leq c_H$

Path separability

Theorem [Abraham and Gavoille, 2006]

- Every H-minor free graph is k-path separable, where k is a constant depending only on H.
- For every k-path separable graph G, there exists φ such that

$$gd(G,\varphi) \leq k^2 \, polylog(n).$$

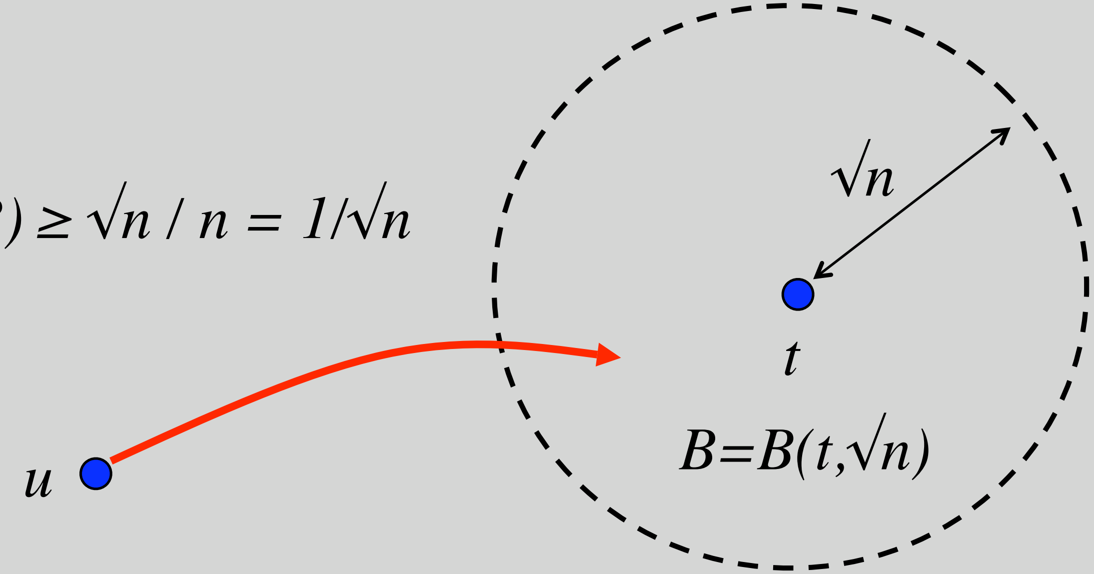
Universal augmentation schemes

O(√n) upper bound

Theorem [Peleg 2005]

The uniform augmenting scheme *U* satisfies *gd(G,U) ≤ O(√n)* for any *G*

$$Pr(u \rightarrow B) \geq \sqrt{n} / n = 1/\sqrt{n}$$



Universal augmentation problem: f-navigability

- Find the smallest function f: $\mathbf{N} \rightarrow \mathbf{R}$ satisfying that for any n-node graph G, there exists an augmenting distribution φ such that $gd(G,\varphi) \leq f(n)$
- We have $f(n) \leq O(\sqrt{n})$
- Is there an $\alpha \geq 0$ such that $f(n) \leq \log^{\alpha} n$?

Graphs with limited navigability

Theorem [F., Lebhar, Lotker, 2006]

Let d such that

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

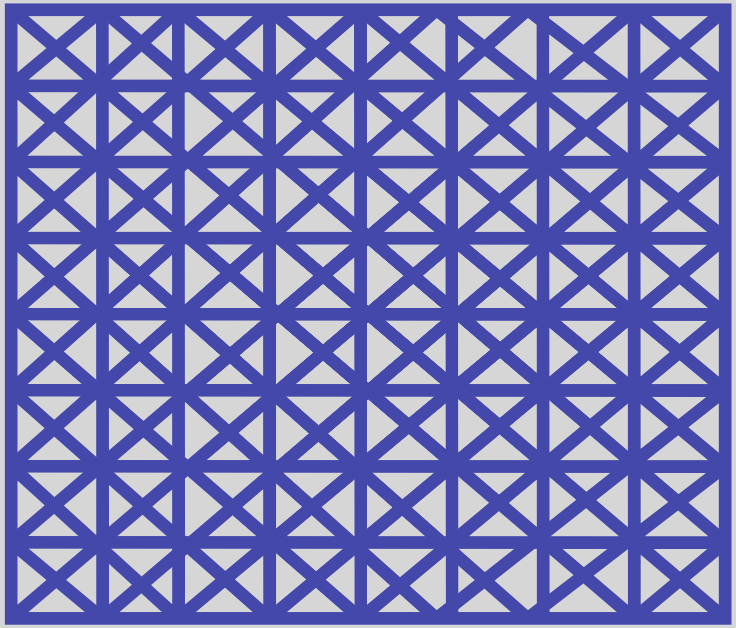
There exists an infinite family of n-node graphs with doubling dimension at most $d(n)$ that are not $polylog(n)$-navigable.

Consequences:

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

Proof of limited-navigability

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



$x = x_1 x_2 \ldots x_d$

is connected to all nodes

$y = y_1 y_2 \ldots y_d$

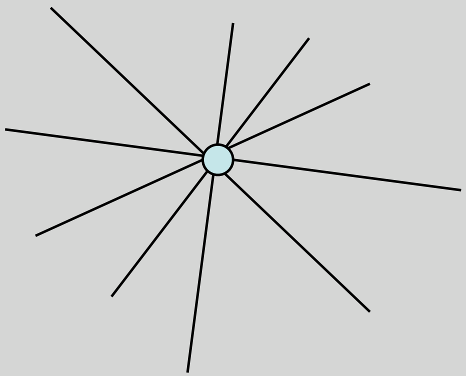
such that $y_i = x_i + a_i$ where

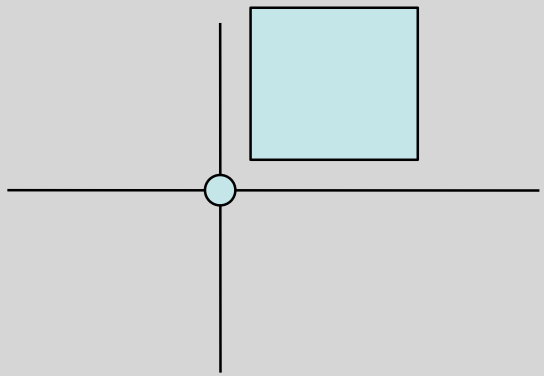
$a_i \in \{-1,0,+1\}$

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_∞

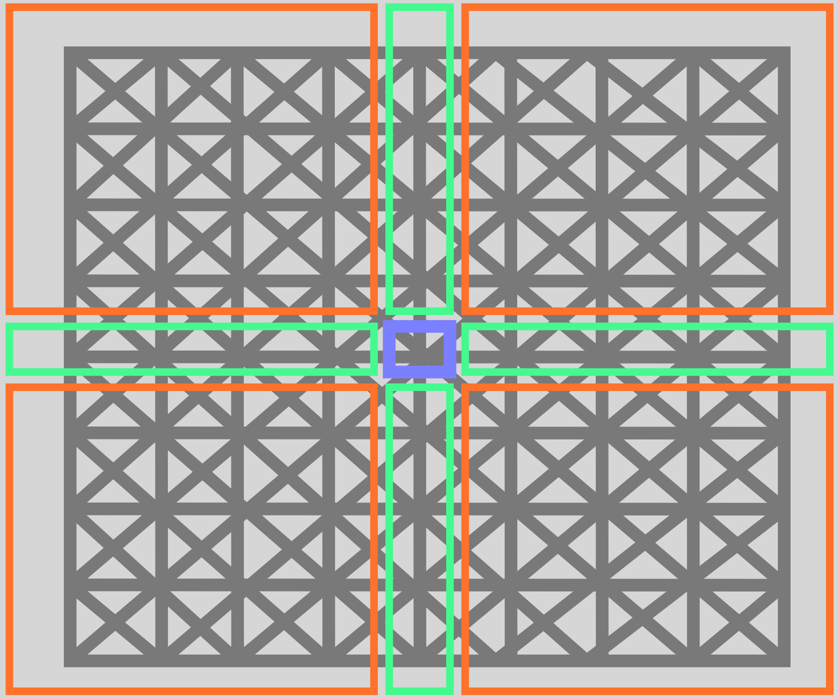




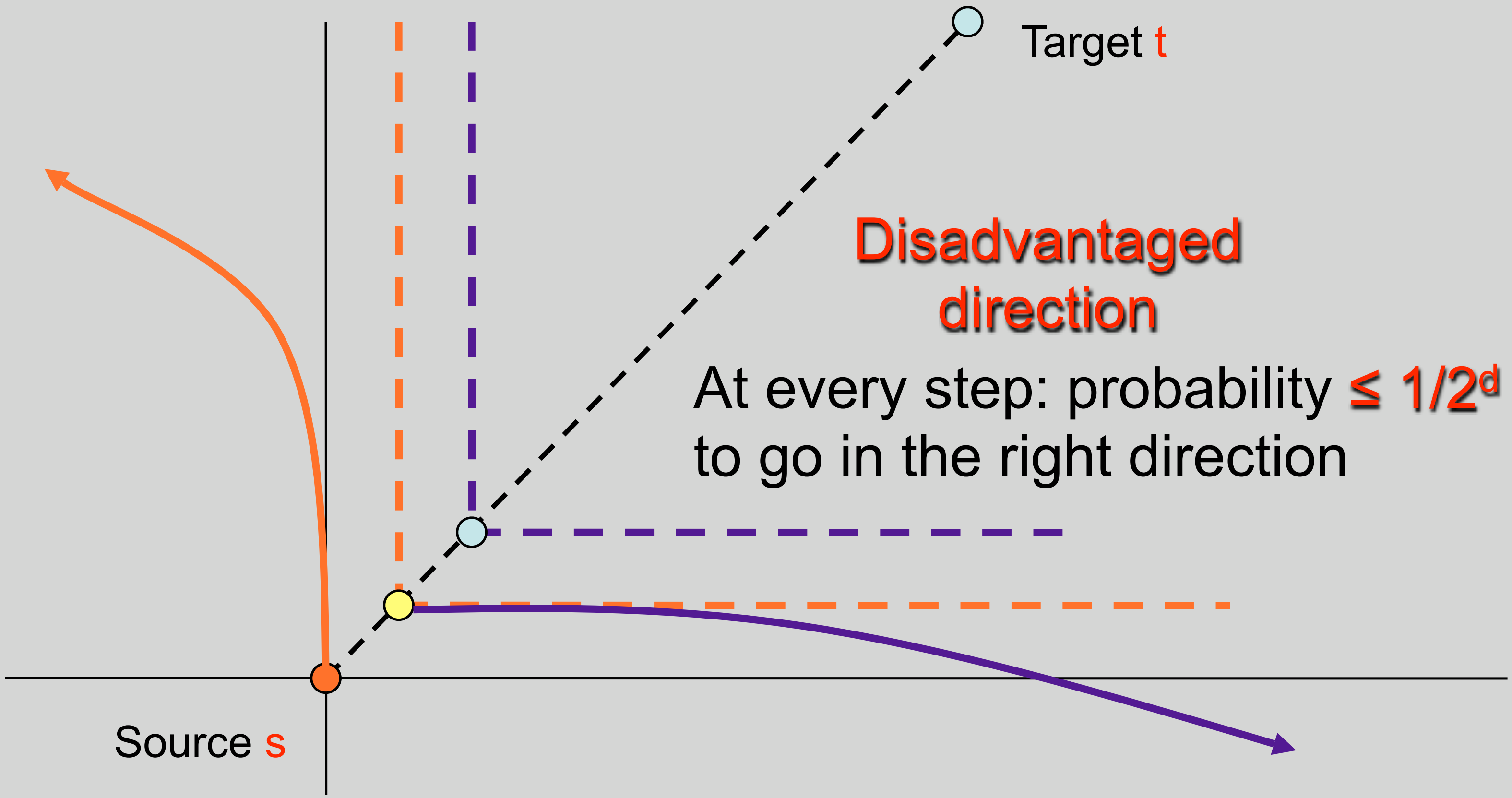
Directions

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

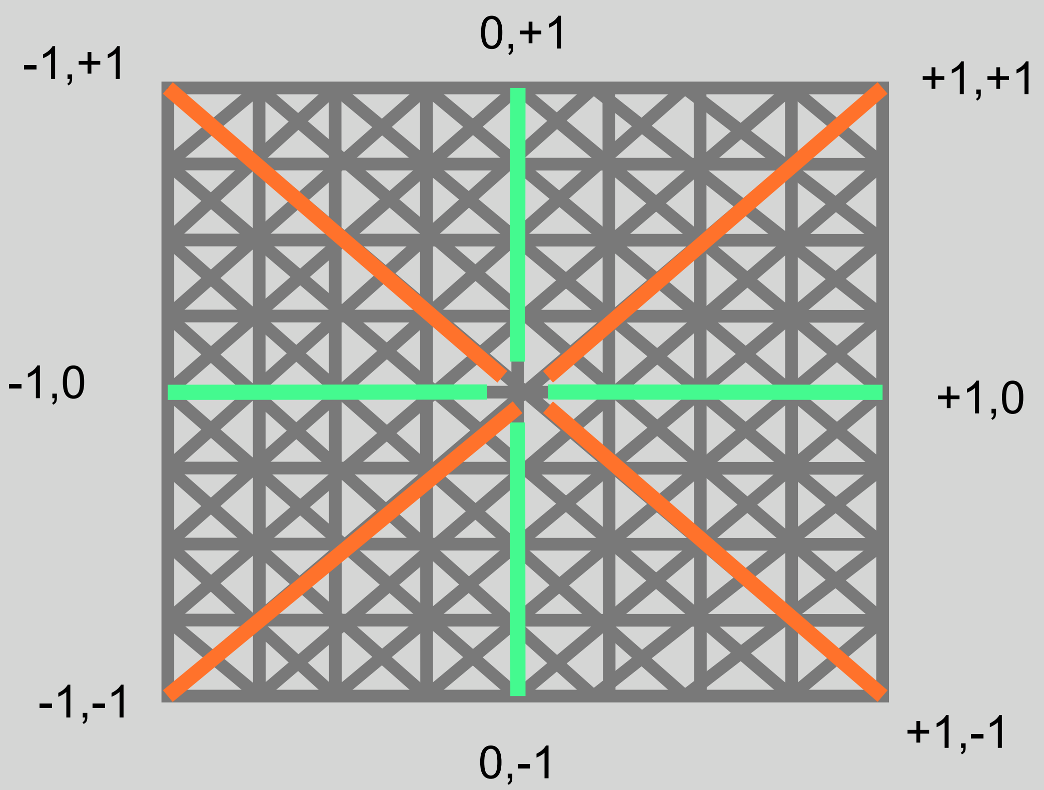
$Dir_\delta(u)=\{v \,/\, v_i = u_i + x_i\, \delta_i$ where $x_i = 1 \ldots p/2\}$



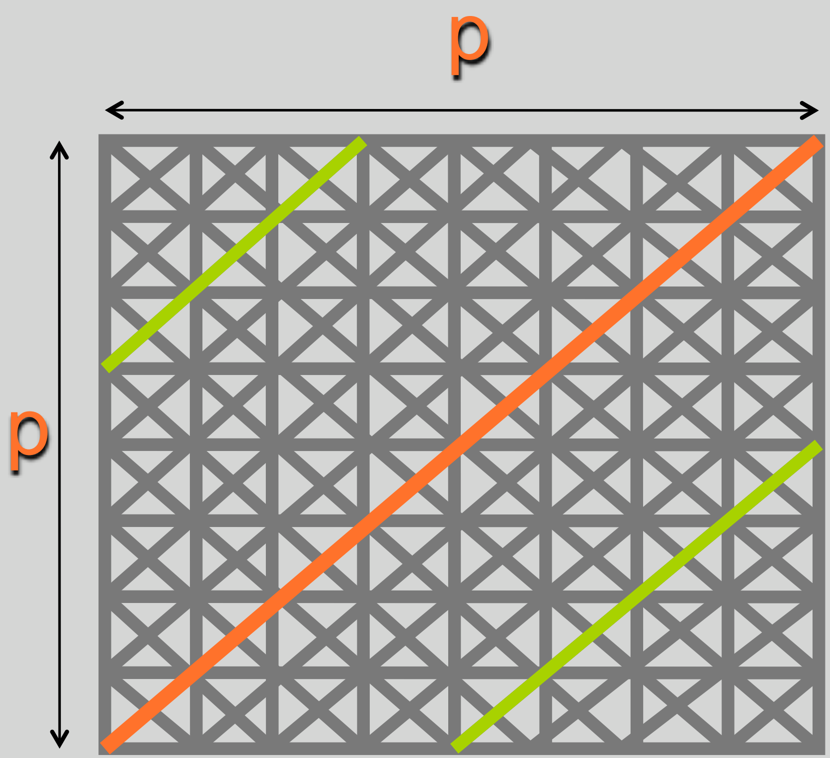
Case of symmetric distribution



-- General case --
Diagonals

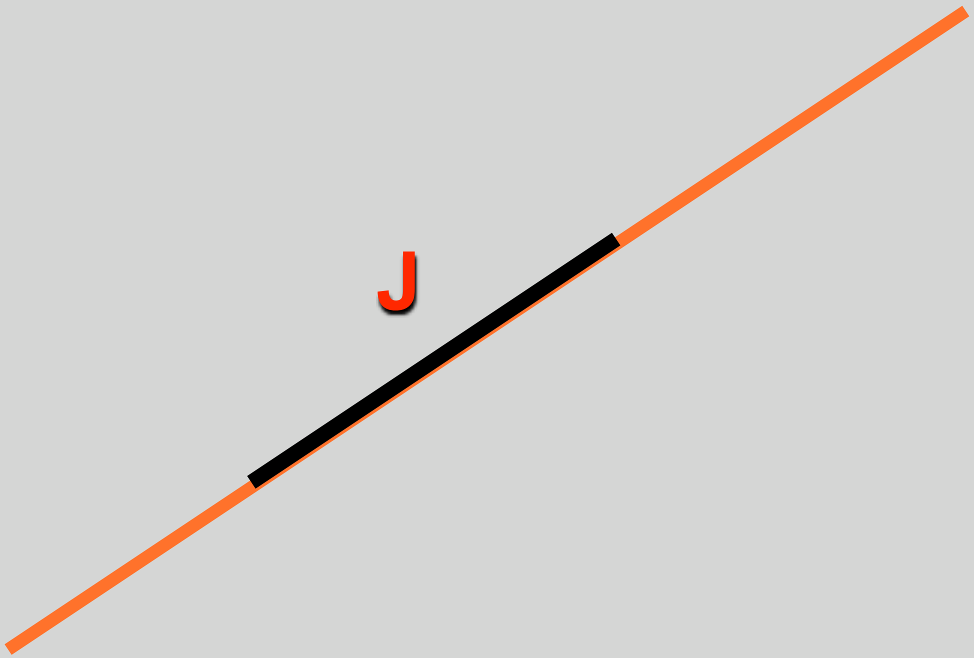


Lines

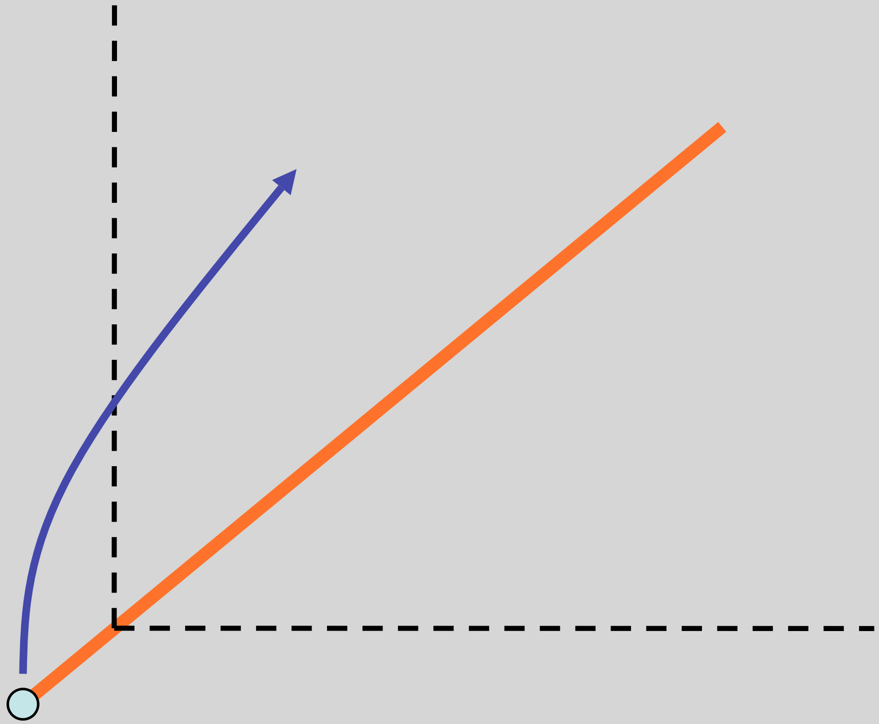


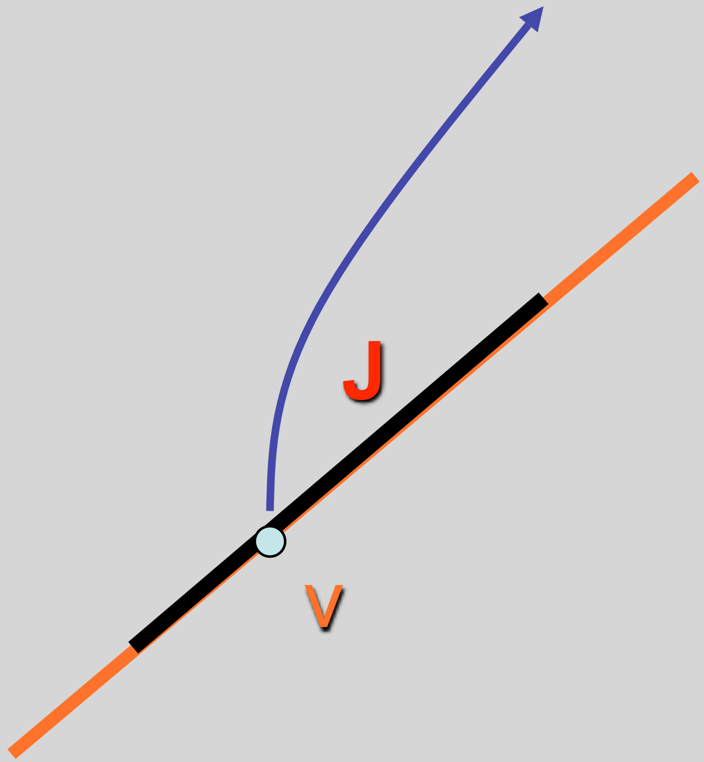
p lines in each direction

Intervals



Certificates



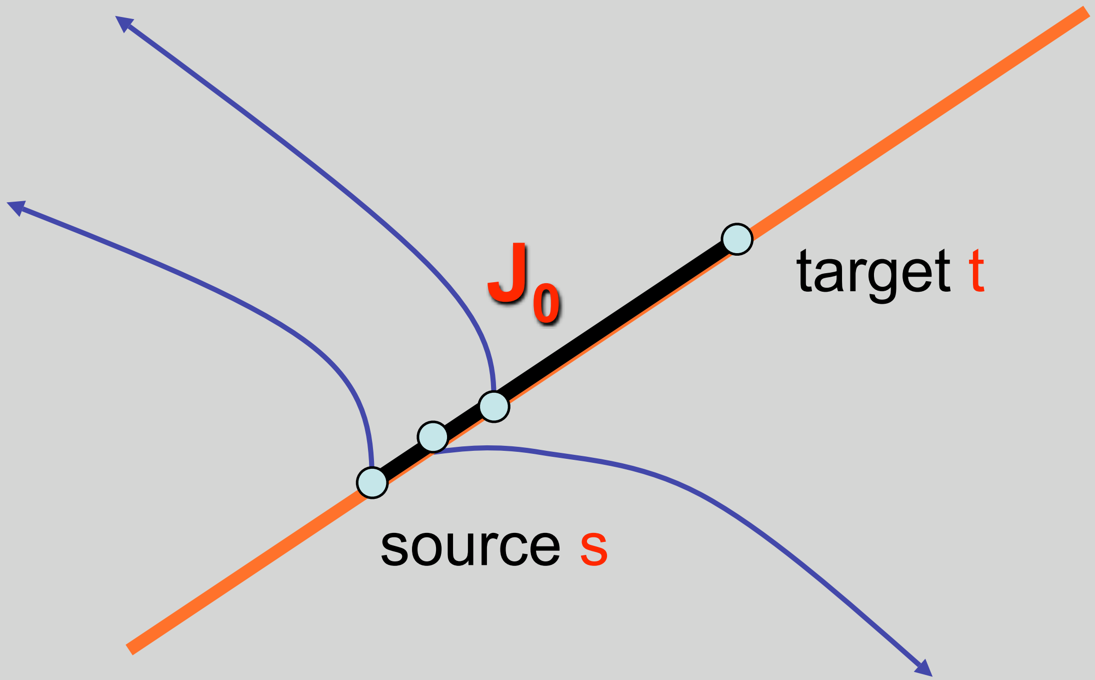


v is a certificate for J

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 $\mathrm{E} = \mathbf{E}_D$(#interval without certificate) $\geq n$
- On the other hand:

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

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

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

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

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

Lower bound

- For any augmenting distribution φ, $gd(H_d,\varphi) \geq 2^d$ where $2^d \leq n^{1/d}$.

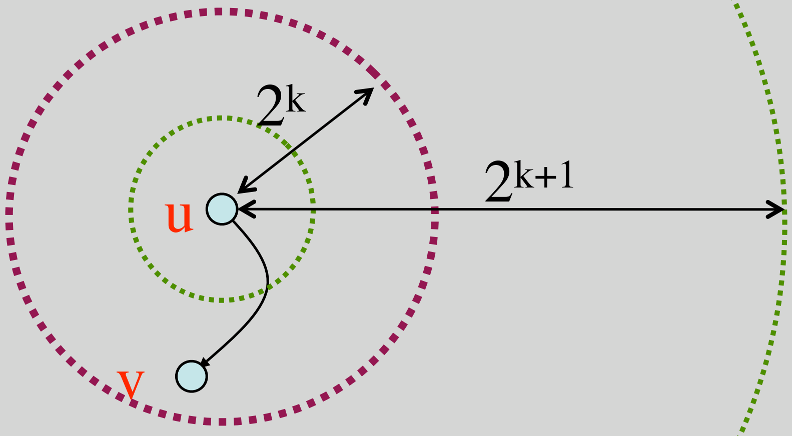
Corollary [F., Lebhar, Lotker, 2006]

$$f(n) \geq \Omega(n^{1/\sqrt{\log(n)}})$$

$\tilde{O}(n^{1/3})$-navigability

log n levels of balls

1) node u selects a level k UAR
2) node u chooses v UAR in $B(u,2^k)$



Theorem [F., Gavoille, Kosowski, Lebhar, Lotker, 2007] $f(n) \leq \tilde{O}(n^{1/3})$

Exercise 3

Prove the upper bound $O(n^{2/5})$ by using the following 2-level augmentation scheme for every node u:

– With probability 1/2, the long-range contact of u is chosen u.a.r. among all the nodes

– With probability 1/2, the long-range contact of u is chosen u.a.r. among the nodes at distance $\leq n^{2/5}$ from u.

Designated metrics

RELAXATION

- Greedy routing in **(G,φ)** according to **$\mu: V\times V \to R^+$**
- Greedy routing does not work for arbitrary function **μ** (dead-end or ping-pong):

  Example: discrete metric: **$\mu(x,y)=1 \Leftrightarrow x \neq y$**

- Greedy routing converges, e.g., if:

  μ is the distance metric in a spanner of **G**

NAVIGABILITY DIAMETER

- Definition The *navigability diameter* of **(G,φ,μ)** is

$$\mathbf{nav(G,\varphi,\mu) = max_{s,t}\ E(X_{s,t})}$$

  where $\mathbf{X_{s,t}}$ is #steps of greedy routing in **(G,φ)** according to **μ**.

- Remark By [Flammini et al], for any **G**, there exist **φ** and **μ** such that **nav(G,φ,μ) ≤ O(log n)**.

THE NAVIGABILITY PROBLEM

Given any (connected) graph **G**, find

- an augmentation **φ**
- a metric **μ**

such that:

- **nav(G,φ,μ)** is small (**polylog(n)**)
- **μ** has small stretch: **$\max_{x,y}$ μ(x,y)/**

POLYLOG NAVIGABILITY

Our result: For any **n**-node connected graph **G** with a positive edge cost function, there exist

- an augmenting distribution **φ,** and
- a semimetric **μ** with stretch **O(log n)**,

such that **nav(G,φ,μ) ≤ O(polylog(n))**.

The semimetric **μ** can be encoded at every node using **O(polylog(n) log(Δ))** bits where **Δ** denotes the normalized diameter of **G**.

MAIN RESULT

Theorem For any **n**-node connected graph **G** with a positive edge cost function, and any integer **k≥1**, there exist

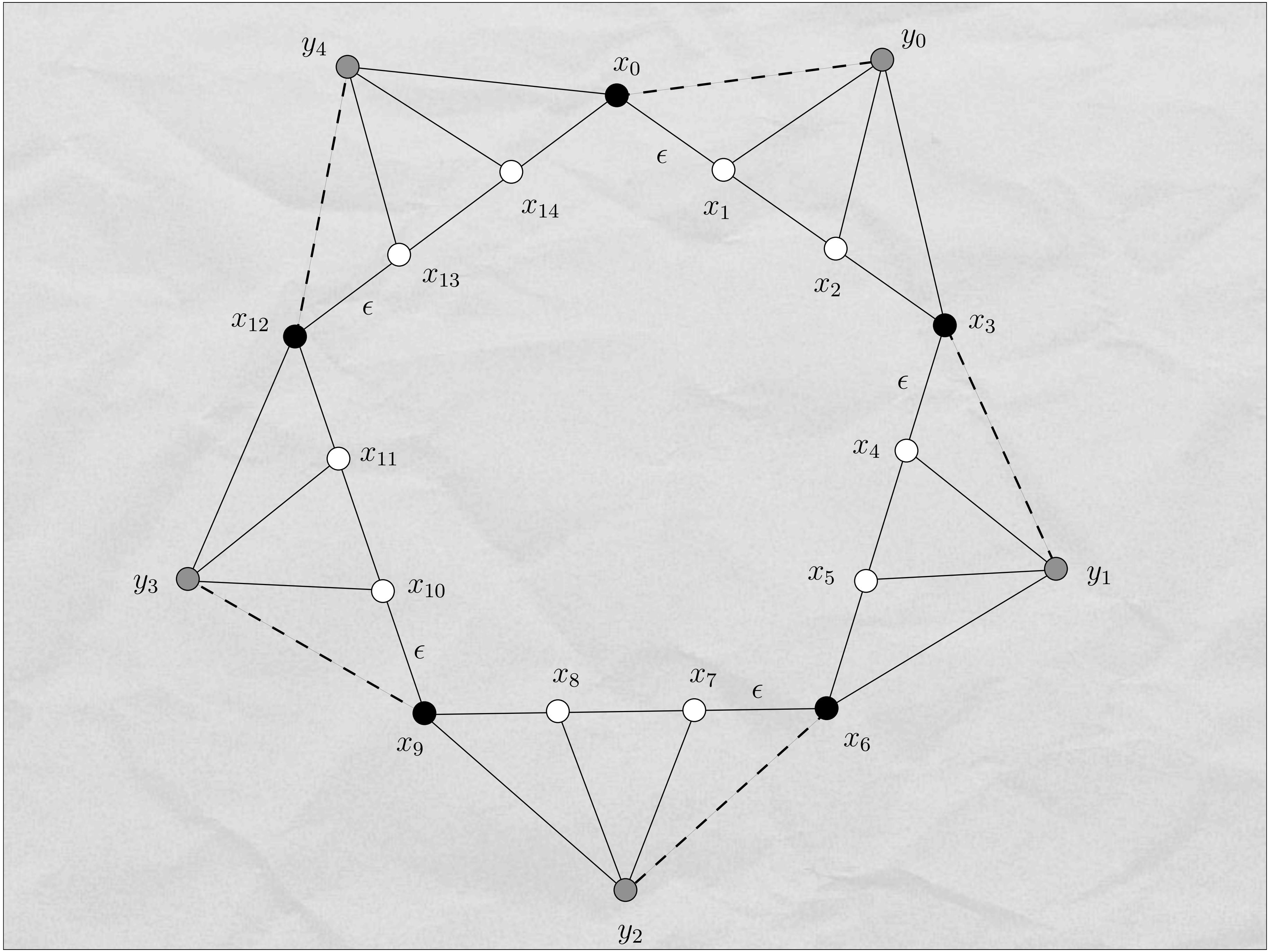
- an augmenting distribution **φ,** and
- a semimetric **μ** with stretch **2k-1**,

such that $\mathbf{nav(G,\varphi,\mu) \leq O(k^2\, n^{2/k} \log^2 n)}$.

The semimetric **μ** can be encoded at every node using $\mathbf{O(k\, n^{1/k} \log n \log(k\Delta))}$ bits.

SPANNER-BASED METRICS

- Proposition There exist an edge-weighted graph **G**, a **(1+ε)**-spanner **S** of **G**, and an augmentation **φ** for **S**, such that:
  - **nav(S,φ,dist$_S$) = O(polylog n)**
  - **nav(G,φ,dist$_S$) = Ω(n)**



y_4
x_0
y_0
ϵ
x_{14}
x_1
x_{13}
x_2
x_{12}
ϵ
x_3
ϵ
x_4
x_{11}
y_3
x_{10}
x_5
y_1
ϵ
x_8
x_7
ϵ
x_9
x_6
y_2

PROOF OF THEOREM: TREE-COVER

- Definition A **(σ,δ)**-tree-cover of **G** is a collection **C** of trees such that:
  - $\forall T \in \mathbf{C}$, T is a subgraph of **G**
  - $\forall x \in V(\mathbf{G})$, $\exists T \in \mathbf{C}$ / $x \in V(T)$
  - $\forall x,y \in V(\mathbf{G})$, $\exists T \in \mathbf{C}$ / $dist_T(x,y) \leq \boldsymbol{\sigma}\, dist_G(x,y)$
  - $\forall x \in V(\mathbf{G})$, $|\{T \in \mathbf{C} \,/\, x \in V(T)\}| \leq \boldsymbol{\delta}$

SPARSE TREE-COVER

- Theorem [Thorup, Zwick, J. ACM 2005] Any graph has a **$(2k-1, O(k\, n^{1/k}))$**-tree-cover, for all **$k \geq 1$**.

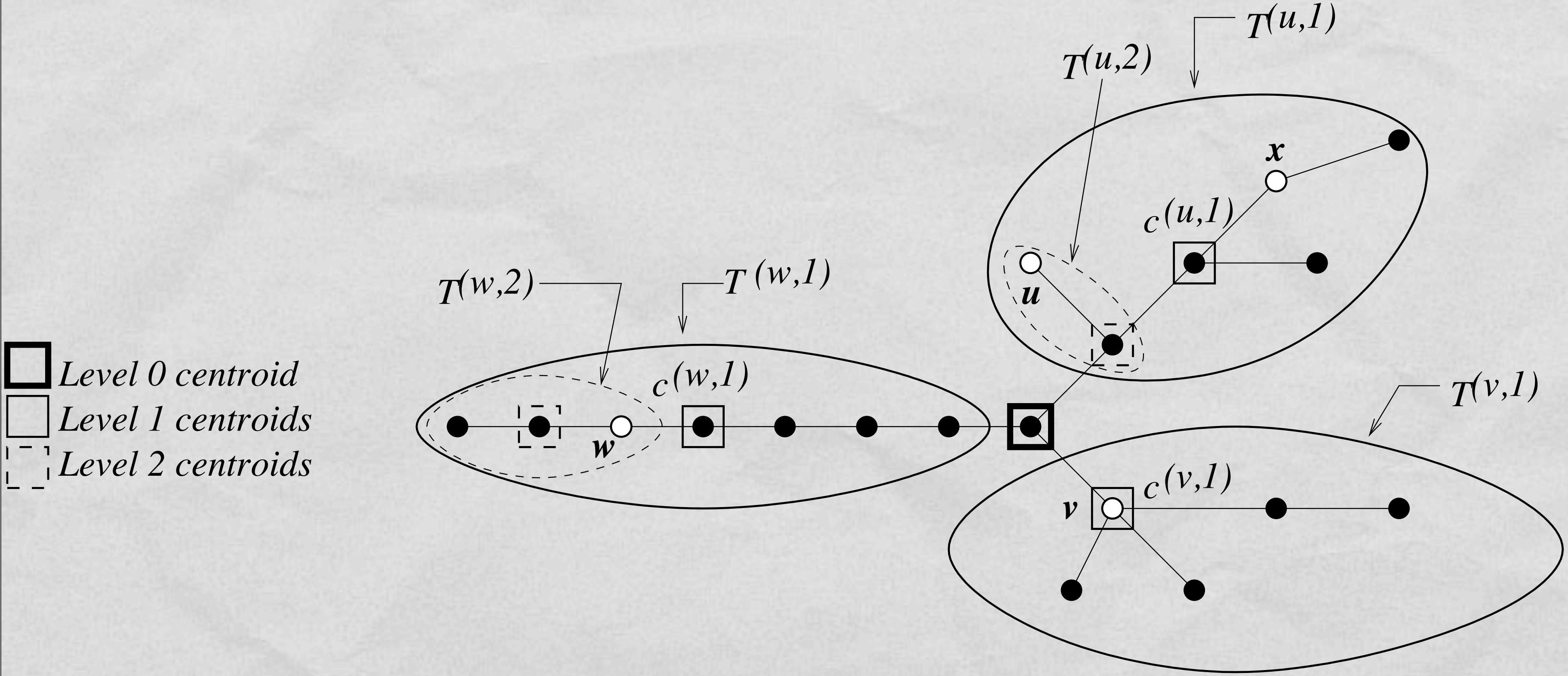
- Lemma If **G** has a **(σ,δ)**-tree-cover with **$\delta \leq n$**, then there exist an augmenting distribution **φ,** and a semimetric **μ** with stretch **σ**, such that **$nav(G,\varphi,\mu) \leq O(\delta^2 \log^2 n)$**.

  The semimetric **μ** can be encoded at every node using **$O(\delta \log n \log(\sigma\Delta))$** bits.

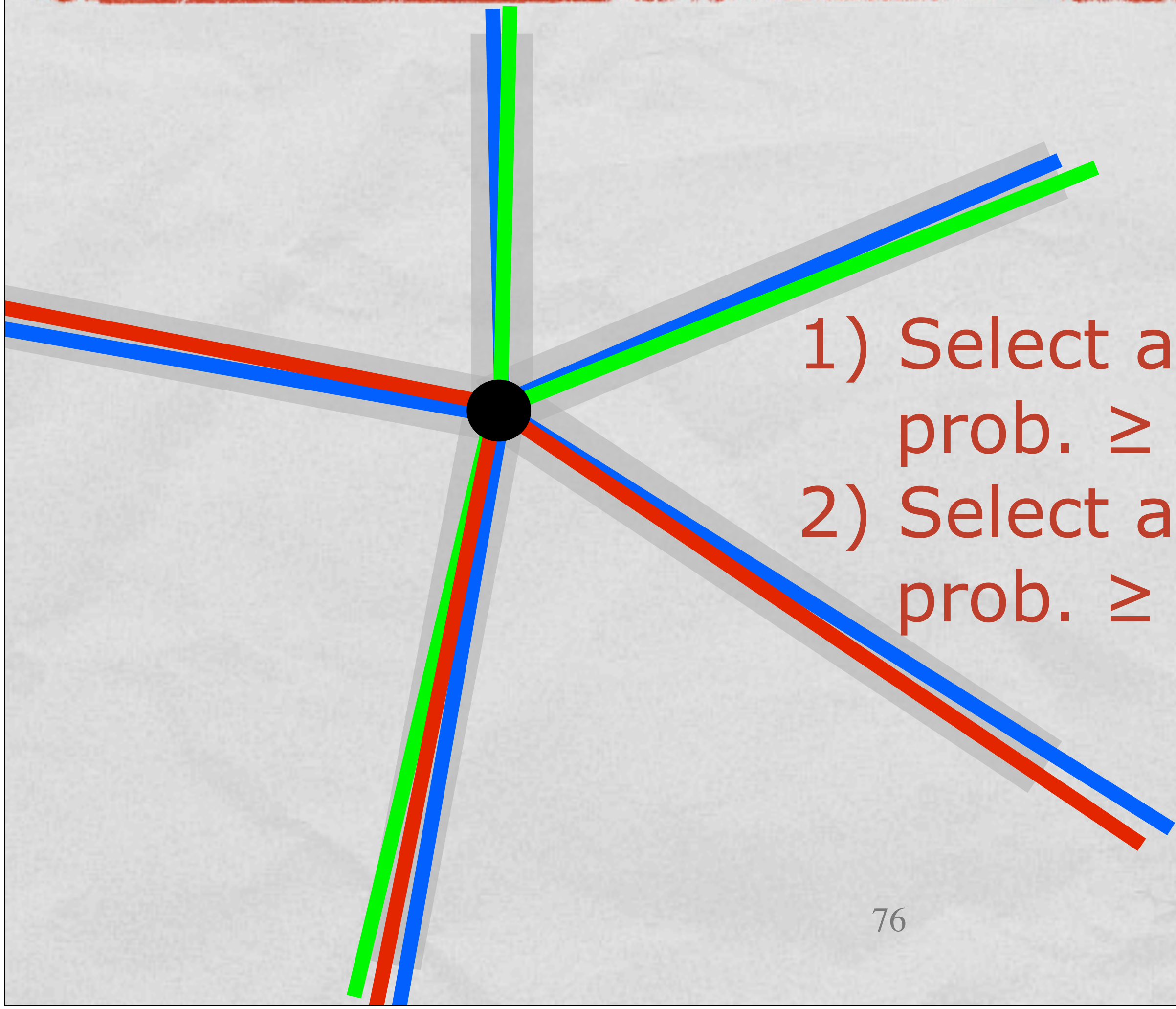
THE SEMIMETRIC

- $\mathbf{C_{u,v}} \subset \mathbf{C}$ is the set of trees containing **u** and **v**.
- Setting: $\boldsymbol{\mu(u,v) = \min_{T \in C_{u,v}} dist_T(u,v)}$
- Remark:
  - **μ** has stretch **σ**
  - **μ** does not satisfy the triangle inequality (it is a semimetric)

CENTROIDAL DECOMPOSITION



AUGMENTATION



1) Select a tree u.a.r.
   prob. ≥ **1/δ**
2) Select a centroid u.a.r.
   prob. ≥ **1/log n**

GREEDY ROUTING ANALYSIS

- Route from source **s** to target **t**: $u_0, u_1, \ldots, u_r$
- $c_{j,k}$ denotes the centroid for **t**, of level **k** in tree **j** containing **t**.
- $\Phi(u)$ = #centroids $c_{j,k}$ closer to target than **u**
- $\Phi_{j,k}(u) = 1$ if $\mu(u,t) > \mu(c_{j,k},t)$, and **0** otherwise

$$\Phi(u) = \Sigma_{j,k} \Phi_{j,k}(u)$$

ANALYSIS (2)

$\mathbf{\Phi(s) \leq \delta \log(n)}$ and $\mathbf{\Phi(t) = 0}$

$\mathbf{Z_i}$ = #steps to reduce $\mathbf{\Phi}$ by at least 1
from $\mathbf{\Phi=i}$ to $\mathbf{\Phi \leq i-1}$

$$\mathbf{nav(G,\varphi,\mu) \leq \textstyle\sum_i E(Z_i)}$$

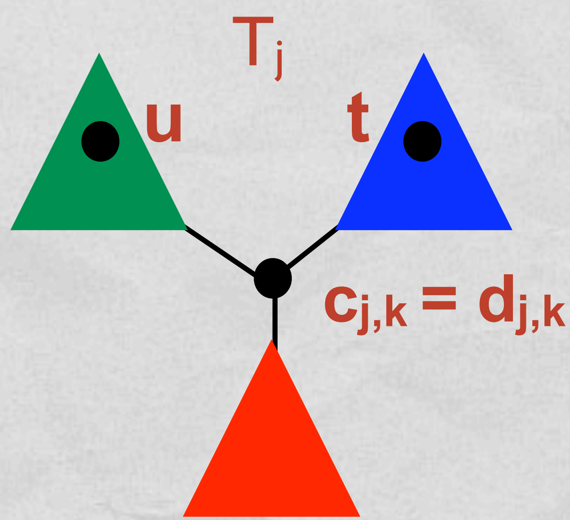
<u>Note:</u> there are at most $\mathbf{\delta \log(n)}$ terms in the sum

<u>Claim</u> $\mathbf{E(Z_i) \leq \delta \log(n)}$

ANALYSIS (3)

- $\mathbf{d_{j,k}}$ denotes the centroid for current node **u**, of level **k** in tree **j** containing **u**.
- Let **j** such that $\boldsymbol{\mu(u,t) = dist_{T_j}(u,t)}$

Let **k** be largest index such that $\mathbf{c_{j,k} = d_{j,k}}$



ANALYSIS (4)

- If $\mathbf{u} = \mathbf{c_{j,k}} = \mathbf{d_{j,k}}$ then $\mathbf{\Phi}$ decreased by 1
- Assume $\mathbf{u} \neq \mathbf{c_{j,k}} = \mathbf{d_{j,k}}$
    - $\boldsymbol{\mu(u,t) = dist_{Tj}(u,t)}$
    - $\boldsymbol{dist_{Tj}(u,t) = dist_{Tj}(u,d_{j,k}) + dist_{Tj}(d_{j,k},t)}$
    - $\boldsymbol{\mu(d_{j,k},t) \leq dist_{Tj}(d_{j,k},t) < dist_{Tj}(u,t)}$
    - Thus if the long-range contact of $\mathbf{u}$ is $\mathbf{c_{j,k}} = \mathbf{d_{j,k}}$ then $\mathbf{\Phi}$ decreases by 1
    - This event occurs with probability $\boldsymbol{\geq 1/(\delta \log(n))}$

COMPACT ENCODING

- Theorem [Gavoille, Peleg, Perenes, Raz, 2004] There exists a distance labeling scheme **(λ,α)** for trees with
  - Labels **$\lambda(T,u)$** of size **$O(\log(n)\log(\Delta))$** bits
  - Decoding **α** of size **$O(1)$** and time **$O(1)$**
- Labeling:

$$L(G,u)=\Big((\beta_1,\lambda(T_{\beta_1},u)),(\beta_2,\lambda(T_{\beta_2},u)),\ldots,(\beta_\delta,\lambda(T_{\beta_\delta},u))\Big)$$

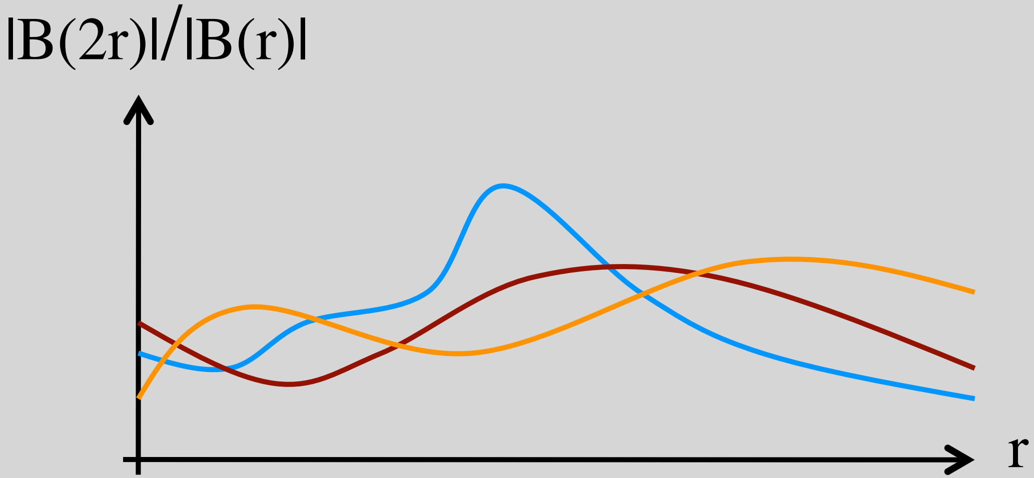
OPEN PROBLEMS

- Is it possible to design an augmenting distribution **φ,** and a semimetric **μ** with *constant* stretch, such that **nav(G,φ,μ) ≤ O(polylog n)**?
- Is it possible to design an augmenting distribution **φ,** and a *metric* **μ** with polylog stretch, such that
    - **nav(G,φ,μ) ≤ O(polylog n)** and
    - **μ** can be encoded at every node using **O(polylog n)** bits?

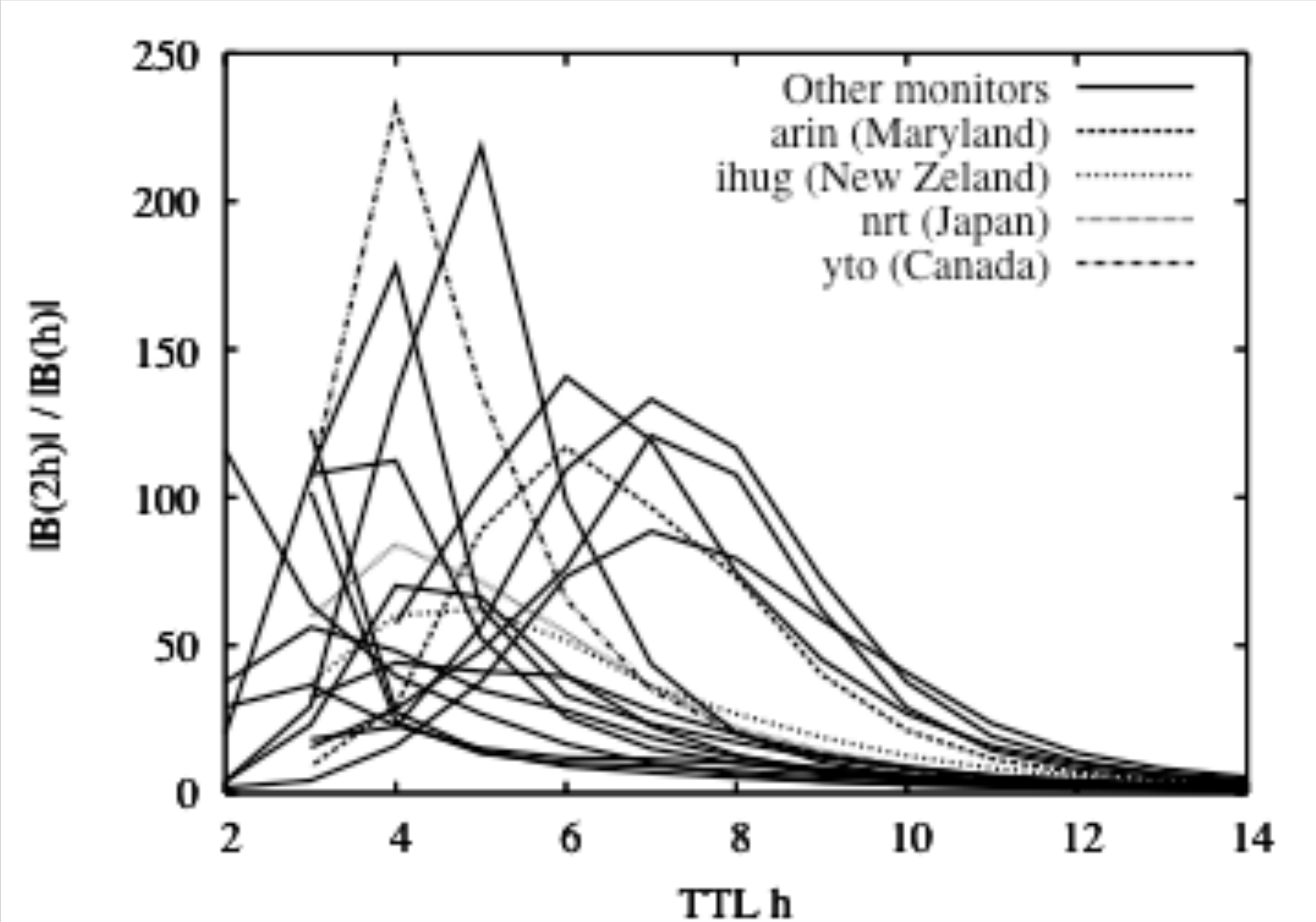
Validation of the model

Internet ball growth

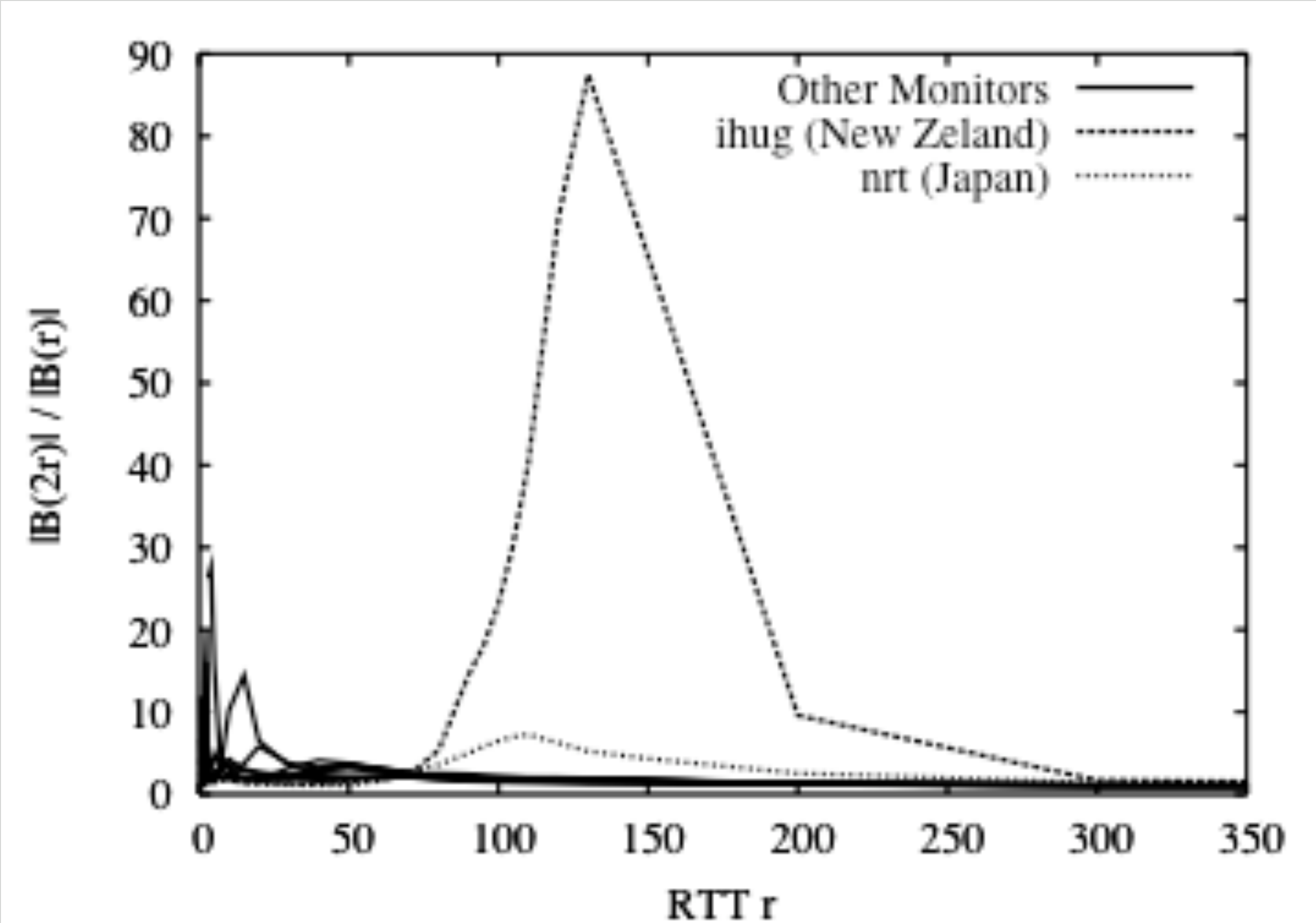
- Experiments [F., Lebhar, Viennot, 2007]
  - Caida's Skitter project
  - TRACEROUTE queries from a small set of monitors to a large set of IP addresses



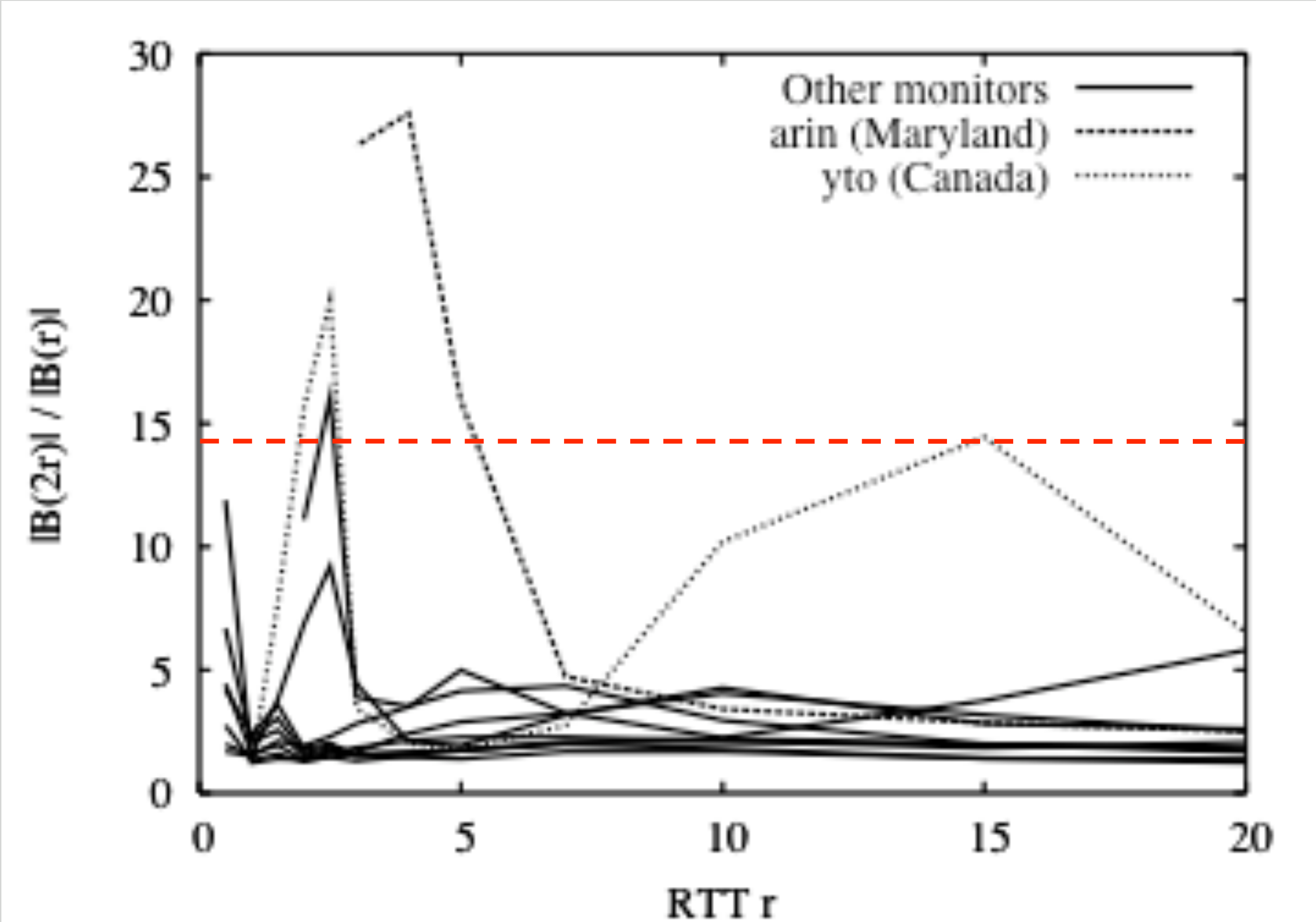
Time to Live (TTL), i.e., #hops



Round Trip Time (RTT), i.e., milliseconds



Round Trip Time (RTT), i.e., milliseconds



Internet doubling dimension

- Experiments [F., Lebhar, Viennot, 2007]
  - King method [Gummadi, Saroiu, Gribble, 2002]
  - Applied to:
    - 2500x2500 matrix of Meridian project

      [Wong, Slivkins, Sirer, 2005]
    - similar results obtained on the 1740x1740 matrix of P2PSim

      [Gil, Kaashoek, Li, Morris, Stribling]

Measurements

- Greedy heuristic to cover $B(s)$ by $B(r)$
- $N(s,r)$ = #balls to cover $B(s)$ by $B(r)$
- Scaling:

d-doubling metric:

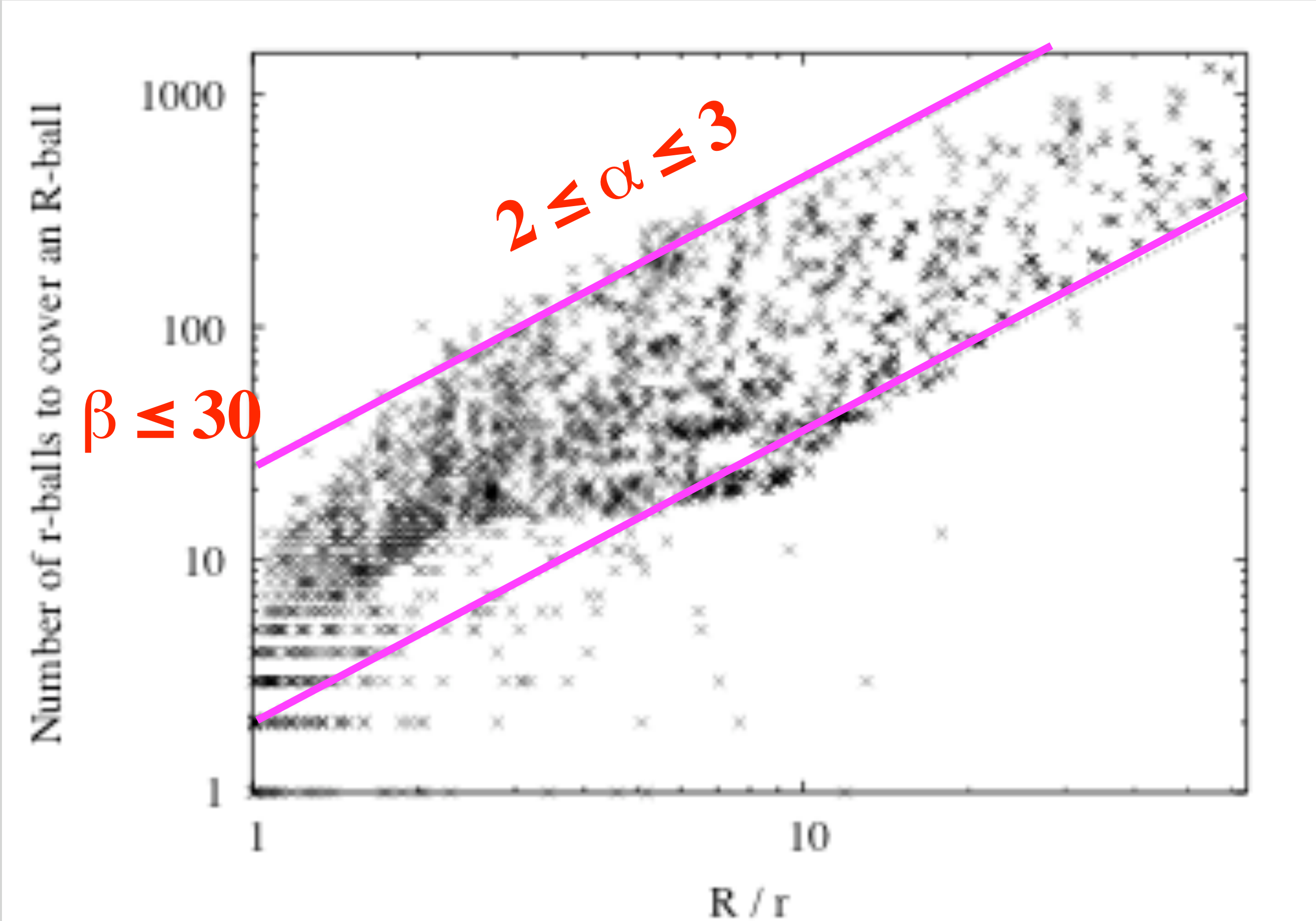
$$N(s,r) = 2^{d \log(s/r)}$$

$$\Rightarrow \log(N(s,r)) = d \log(s/r)$$

(α,β)-doubling:

$$N(s,r) = 2^{\beta}\, 2^{\alpha \log(s/r)}$$

$$\Rightarrow \log(N(s,r)) = \alpha \log(s/r) + \beta$$



1000
100
10
1
1
10
Number of r-balls to cover an R-ball
R / r
2 ≤ α ≤ 3
β ≤ 30

Extracting the long range links

Let $H \in (G,\varphi)$, i.e., $E(H) = E(G) \cup F$

Objective: extracting F from H.

Maximum likelihood method

Compute the edge-set $S \subseteq E(H)$ of cardinality n that maximizes

$$Pr(H \mid S \text{ is the set of long range links})$$

Difficulties

1. Exponential number of sets S
2. Requires the knowledge of φ

Local Maximum Likelihood (LML)

- For $H \in (G,\varphi)$, with $E(H) = E(G) \cup F$, LML decides whether $e \in F$ depending on $Pr(H \mid e \in F)$
- Augmenting distribution φ :
  - [Chung, Lu, 2004] [Andersen, Chung, Lu, 2006]

    Power law distribution
  - [F., Lebhar, Lotker, 2007]

    Density-based distribution

    If $dist_G(u,v) = r$ then $\varphi_u(v) \approx 1/|B(u,r)|$

Difficulty: locality!

Clustering hypothesis on G

- For every $e = (u,v) \in E(G)$, there exist at least k disjoint paths of length at most L between u and v [Chung, Lu, 2004]
- For every $e = (u,v) \in E(G)$, there at least k units of flow can be pushed from u to v along paths of length at most L [Andersen, Chung, Lu, 2006]
- Every $e \in E(G)$ belongs to at least $\Omega(\log(n)/\log\log(n))$ triangles [F., Lebhar, Lotker, 2007]

Extraction algorithm

Given $H \in (G, \varphi)$, with $E(H) = E(G) \cup F$

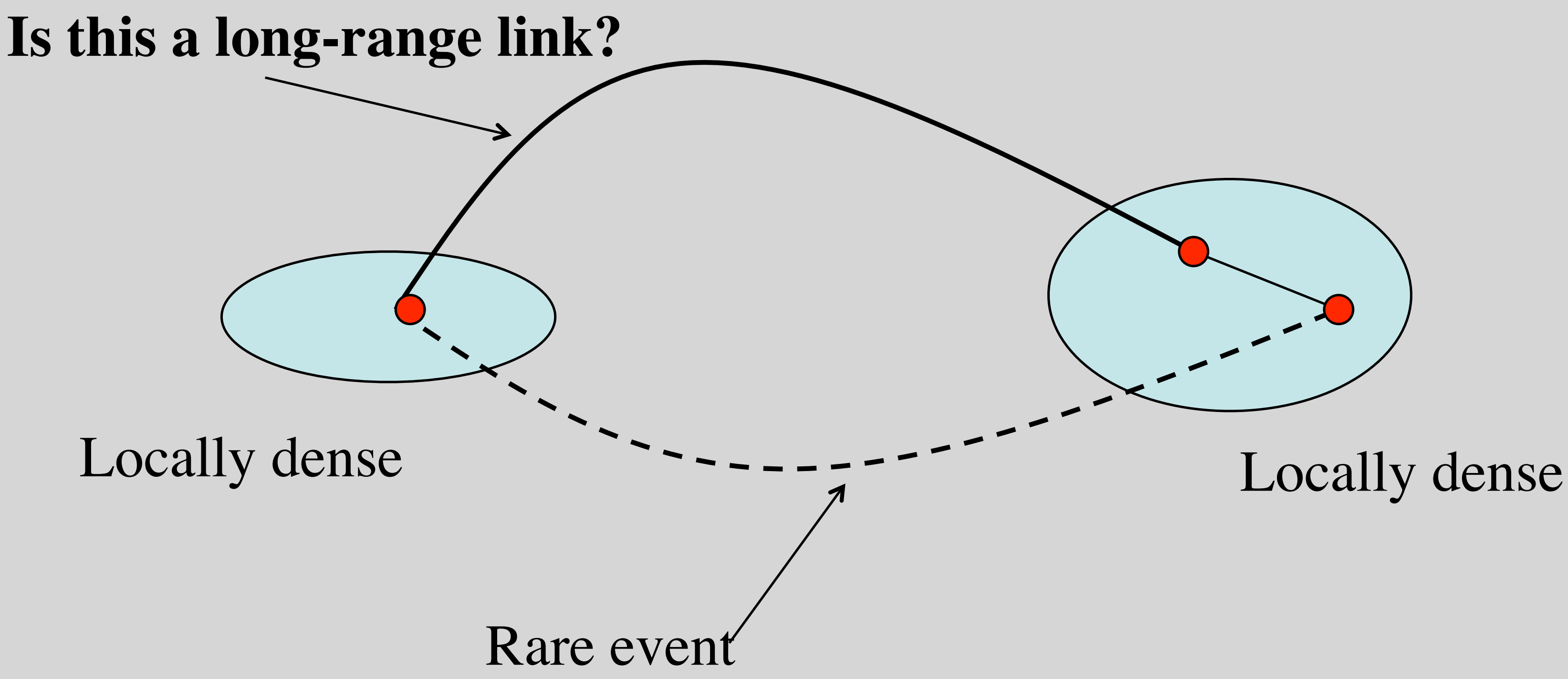
- Density based augmenting distribution φ
- Base graph G with
  - bounded ball growth or bounded doubling dimension
  - clustering $\Omega(log(n)/loglog(n))$

Theorem [F., Lebhar, Lotker, 2007]

There is a LML algorithm that returns $F' \subseteq F$ where:

- W.h.p., $H\text{-}F'$ contains at most *polylog n* edges of stretch larger than *polylog n*, and
- Greedy routing using the map $H\text{-}F'$ performs in *polylog n* expected number of steps

Idea of the proof

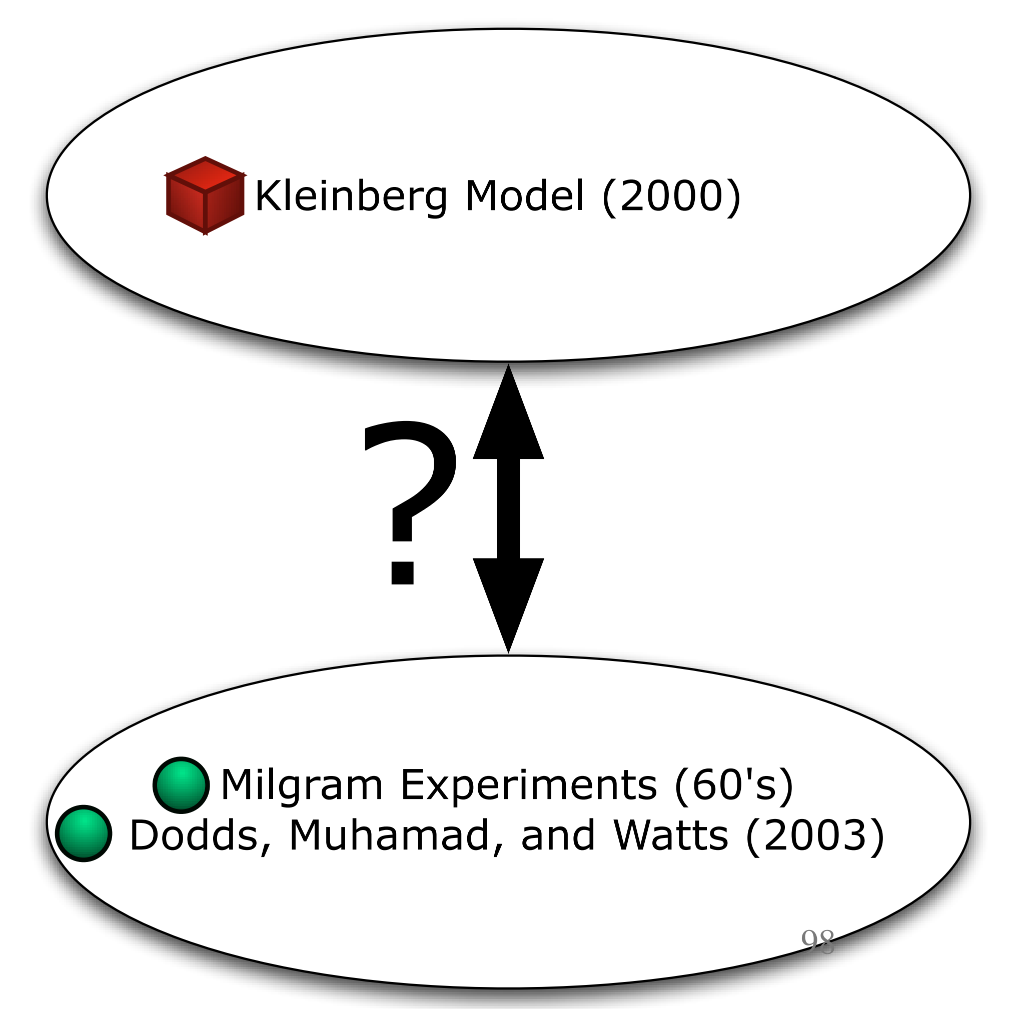


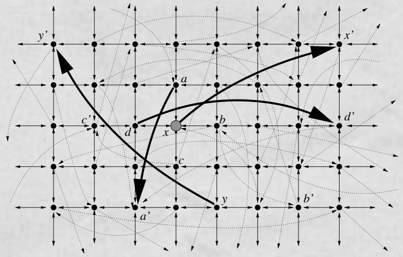
Clustering Coefficient Revisited

- Theorem [F., Lebhar, Lotker, 2007]
  For any $0<\varepsilon<1/5$, any LML algorithm for recovering the base graph C_{2n+1} in $H\in(C_{2n+1},h)$ fails in the detection of an expected number $\Omega(n^{5\varepsilon} / \log n)$ of long range links of stretch $\Omega(n^{1/5 - \varepsilon})$.

Emergence

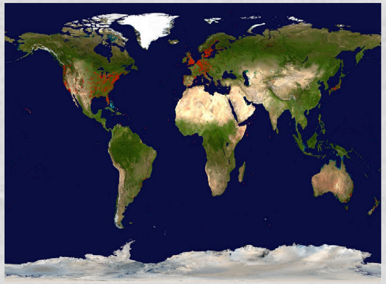
NAVIGABILITY VS. MODELS



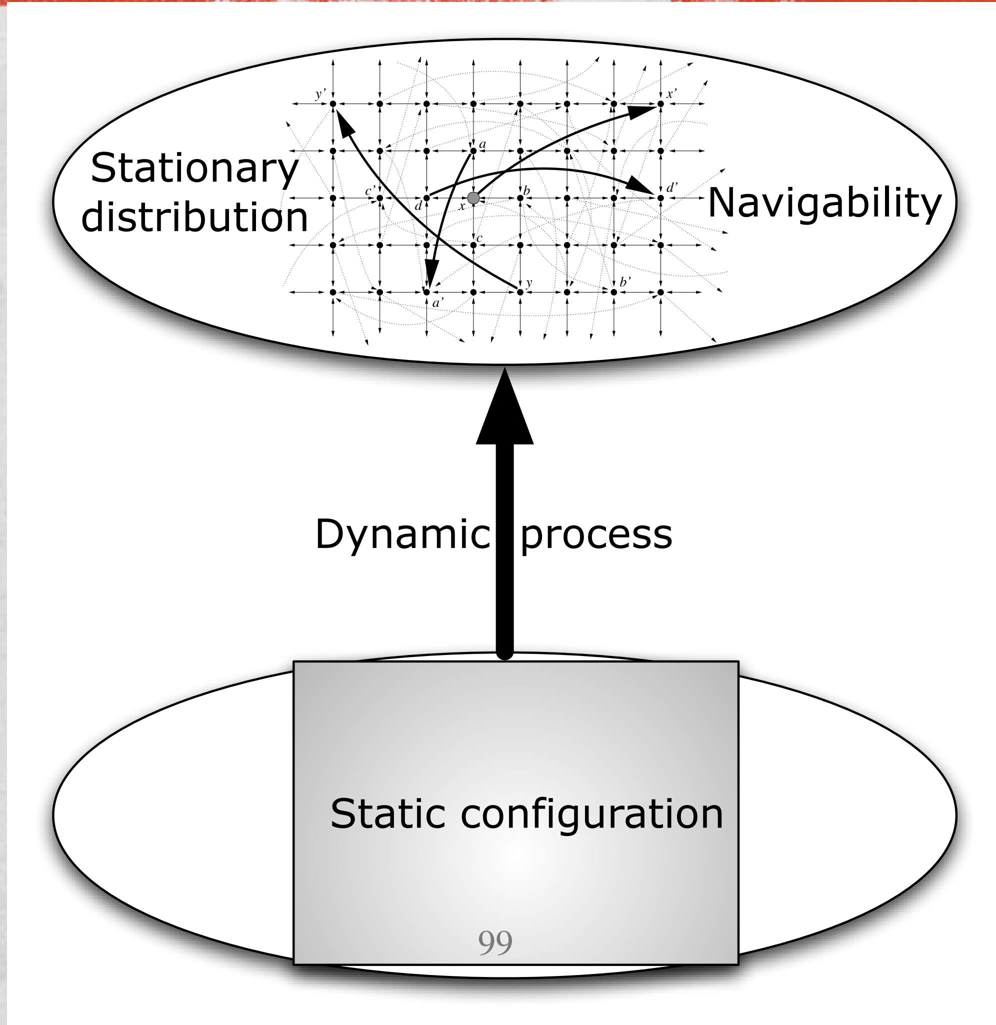


Graph models

Acquaintanceship networks

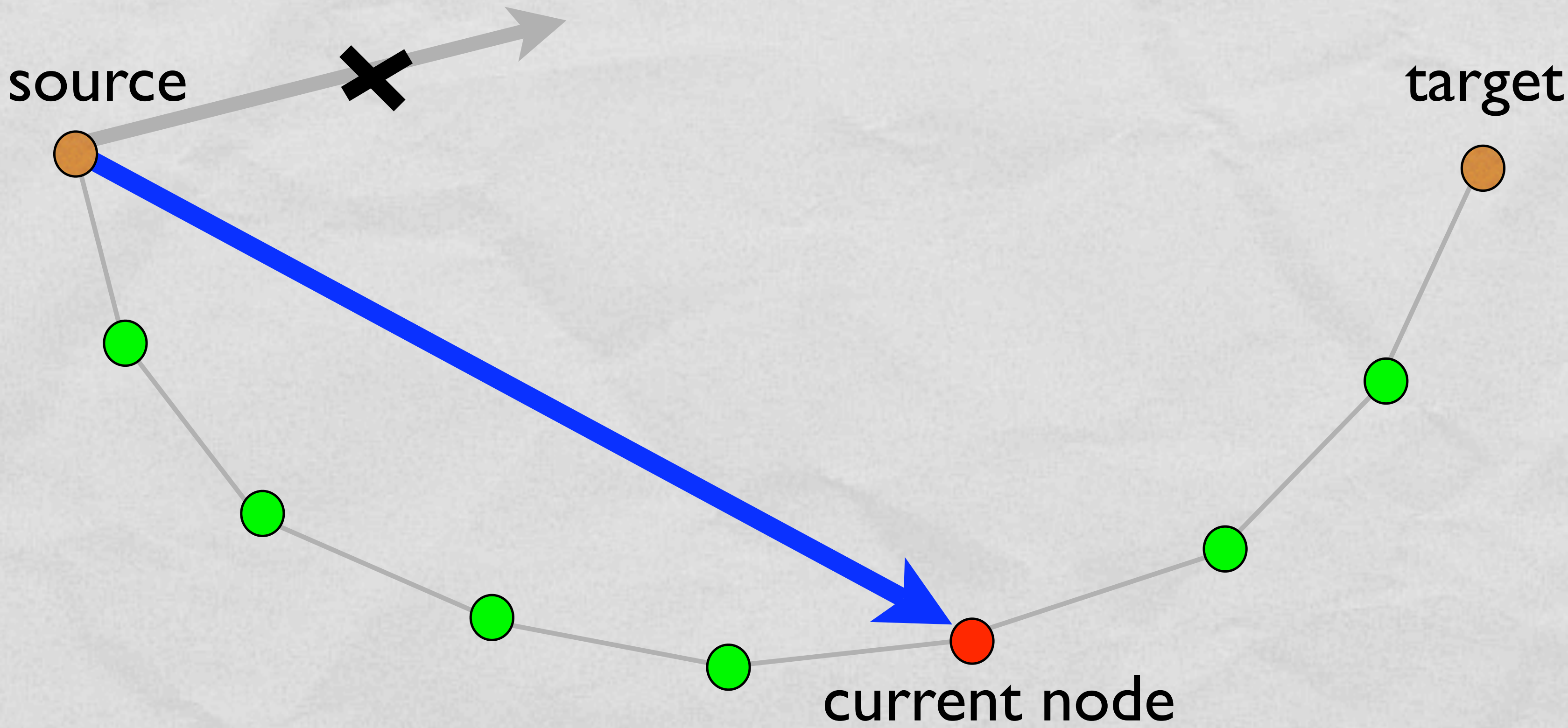


NAVIGABILITY AS AN EMERGING PROCESS



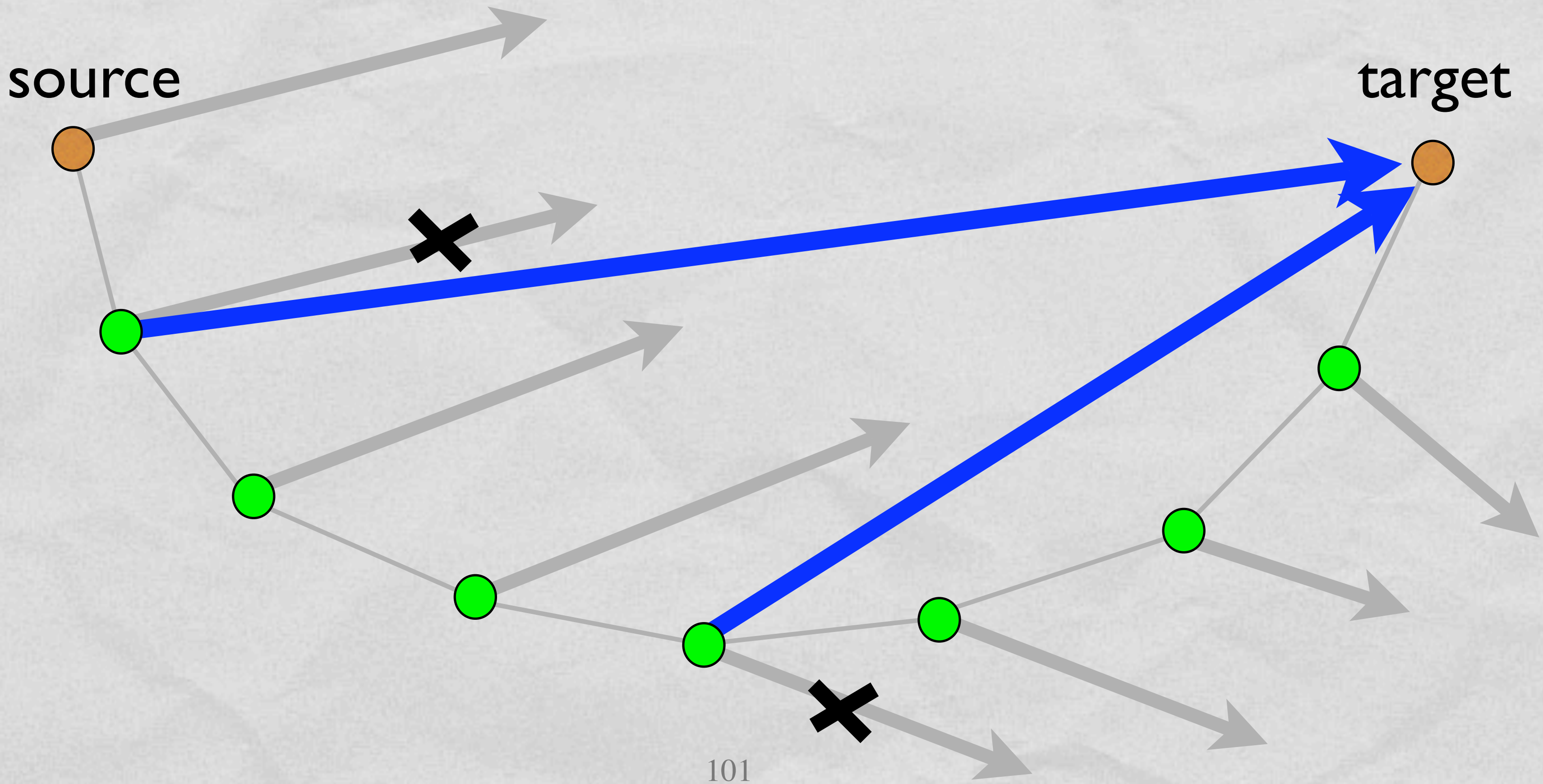
CLAUSET AND MOORE

Web surfer analogy

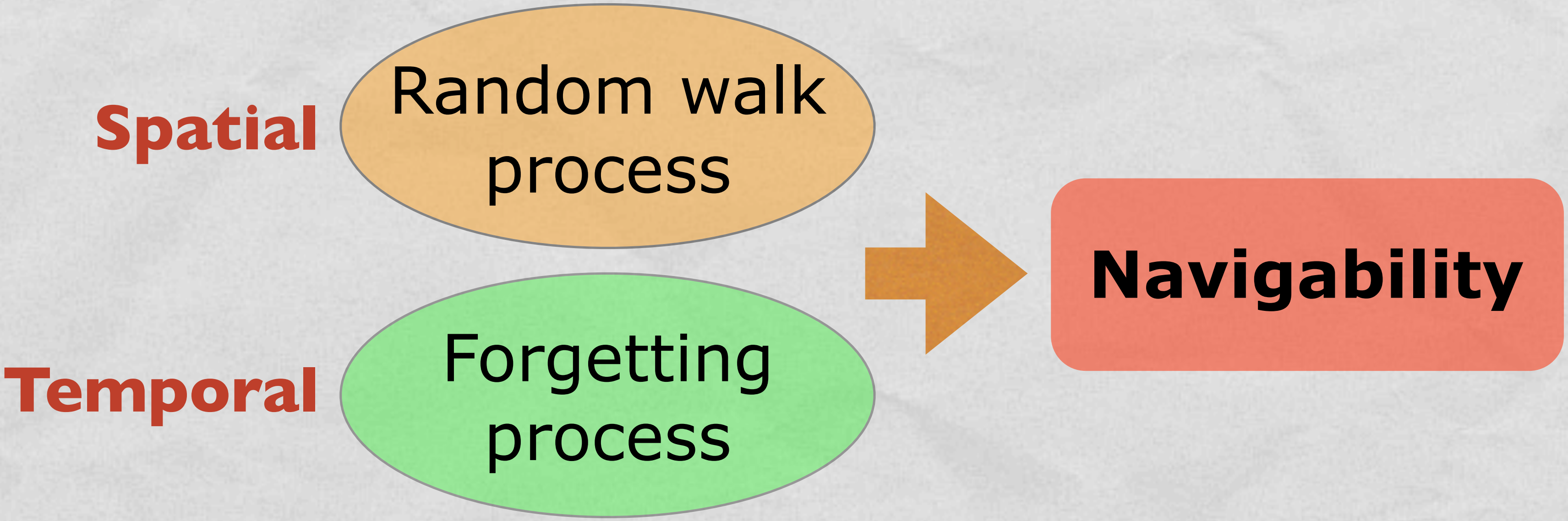


CLARKE AND SANDBERG

Freenet analogy



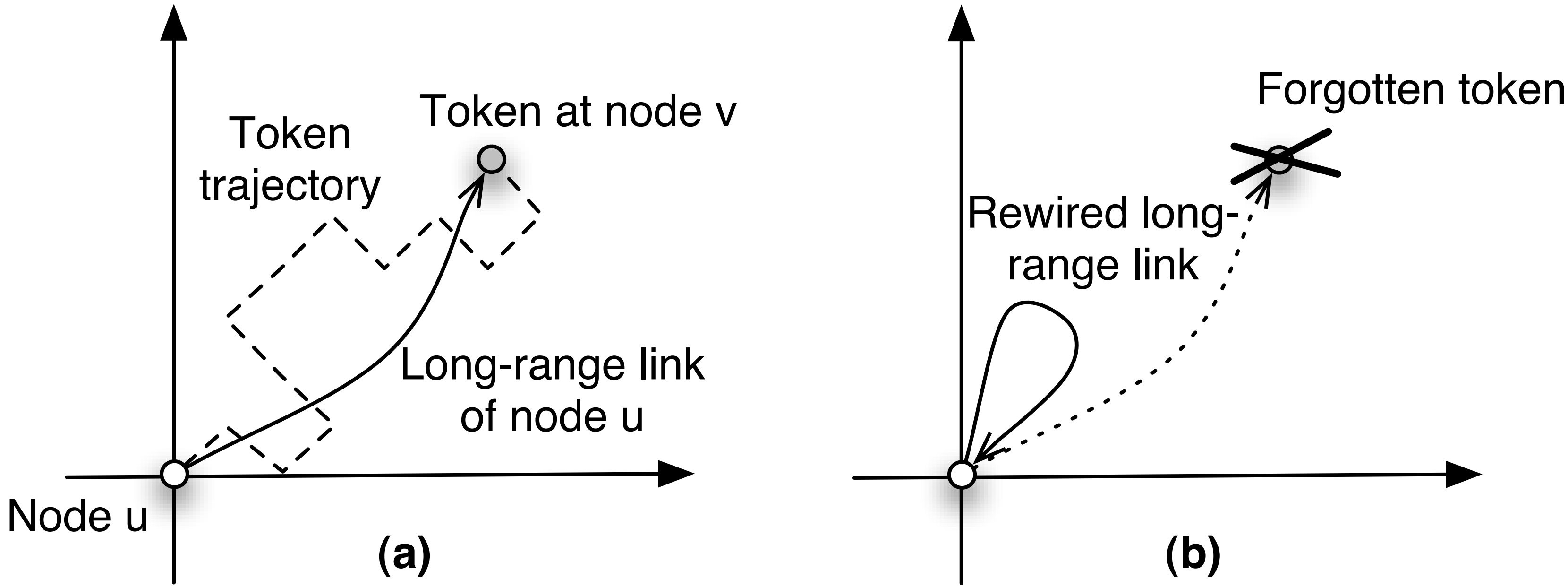
OUR RESULT


Spatial
Random walk process
Temporal
Forgetting process
Navigability

SUMARY OF RESULTS

	Harmonic link distribution	Navigability
Clauset and Moore	Simulations	Simulations
Sandberg and Clarke	Proof	Simulations
Move-and-Forget (M&F)	Proof	Proof

MOVE-AND-FORGET (M&F)



SETTINGS

- Random walk:

$$X(t) = \big(X_1(t), X_2(t), \dots, X_k(t)\big)$$

$X_i(t) = X_i(t-1) \pm 1$ with probability 1/2

- Forgetting process:

Pr(link is forgotten | link of age a) = $\Phi_\varepsilon(a)$

$$\Phi_\varepsilon(a) = 1 - \big((a-1)\log^\varepsilon(a-1)\big) / \big(a \log^\varepsilon(a)\big)$$

HARMONIC FORGETTING PROCESS

- $\Pr(\text{link is forgotten} \mid \text{link of age } a) \approx 1/a$

$$\Rightarrow \Pr(\text{forget } \alpha a) \approx \Pr(\text{forget } a) / \alpha$$

STATIONARITY

- $\pi(a)$ = Pr(link has age a)

  Lemma $\pi(a) \approx 1/a$

- $f(\mathbf{d}) = \Pr(\mathbf{u} \rightarrow \mathbf{u}+\mathbf{d})$

  Lemma $f(\mathbf{d}) = \Sigma_{a \geq 1}\ \pi(a) \Pr\{ X(a)=\mathbf{d} \}$

LINK DISTRIBUTION

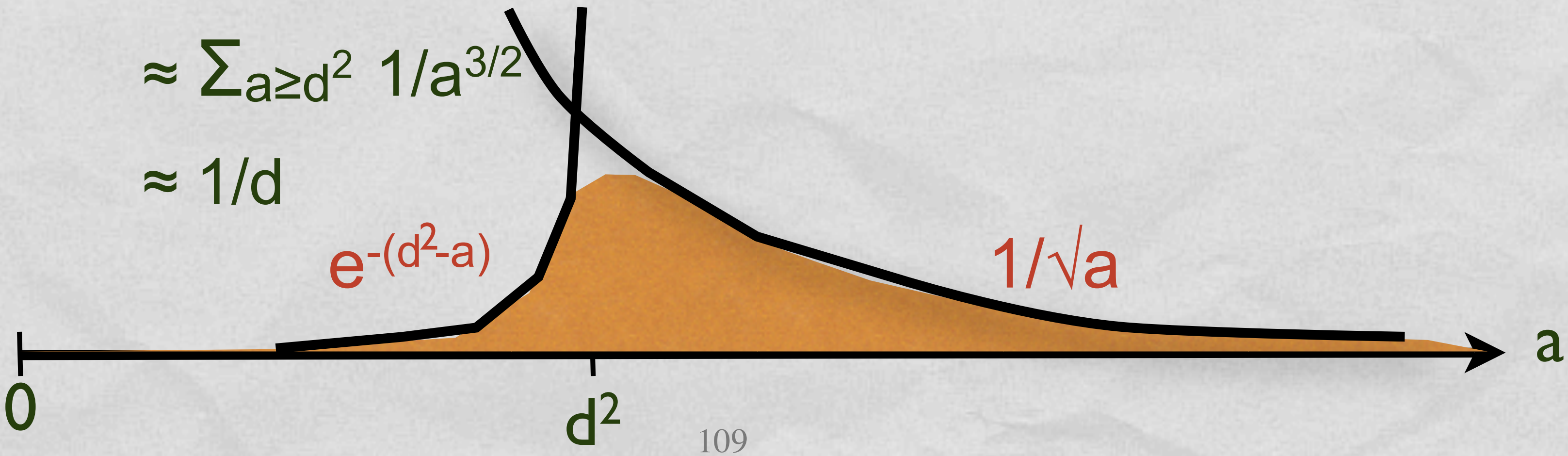
Theorem There exists $d_0 \geq 0$, and $c, c' \geq 0$, such that, for any $\mathbf{d}=(d_1,...,d_k)$, with $|d_i| \geq d_0$ for any $i=1,...,k$, we have

$$\frac{c}{||\mathbf{d}||^k \log^{1+\varepsilon}||\mathbf{d}||} \leq f(\mathbf{d}) \leq \frac{c' \log^{k/2}||\mathbf{d}||}{||\mathbf{d}||^k \log^{1+\varepsilon}||\mathbf{d}||}$$

IDEA OF THE PROOF

1-dimensional lattice

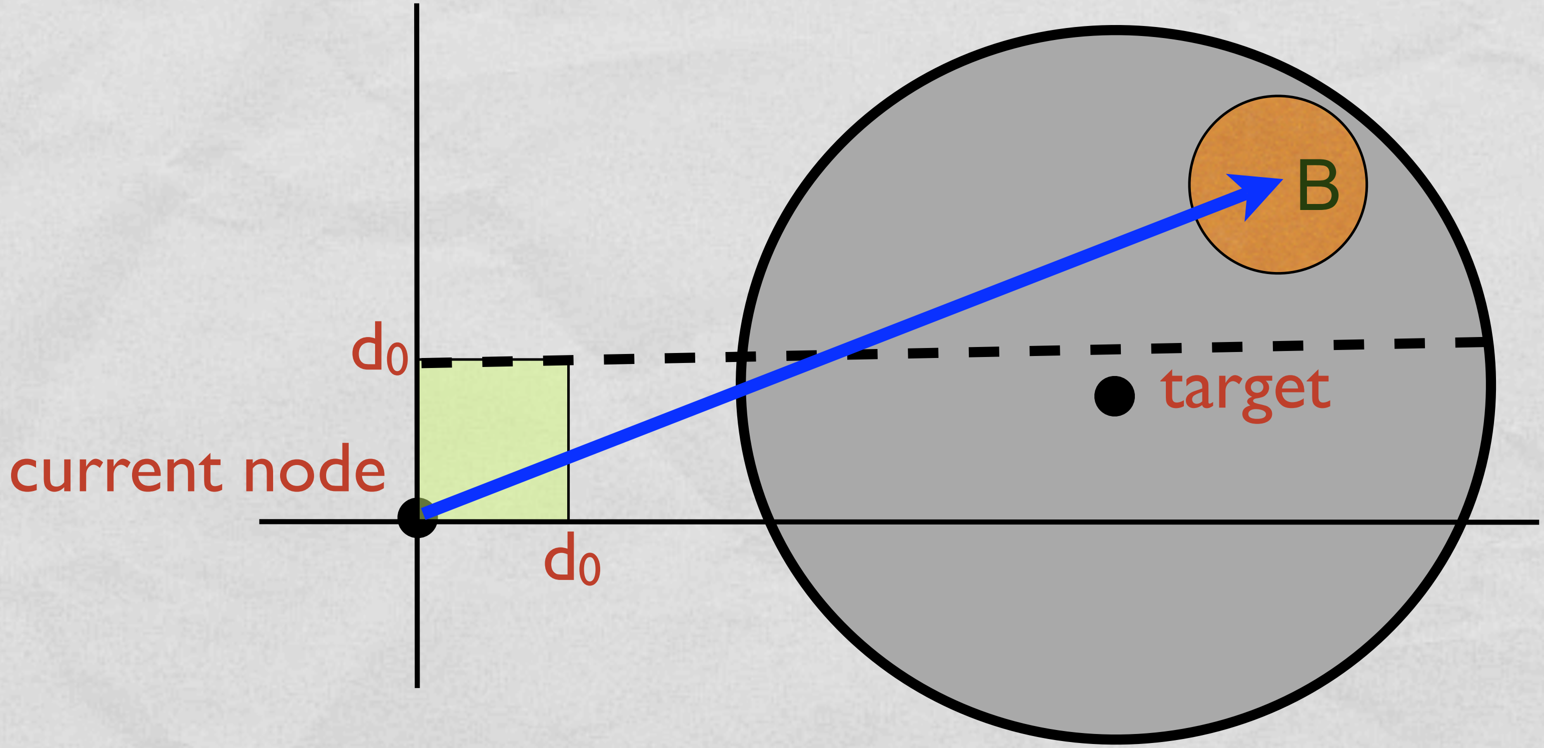
$$f(d) \;=\; \Sigma_{a\geq 1}\, \pi(a)\, \Pr\{\, X(a)=d \,\}$$
$$\approx \Sigma_{a\geq d^2}\, \pi(a)\, \Pr\{\, X(a)=d \,\}$$
$$\approx \Sigma_{a\geq d^2}\, \pi(a)\, 1/\sqrt{a}$$
$$\approx \Sigma_{a\geq d^2}\, 1/a^{3/2}$$
$$\approx 1/d$$



GREEDY ROUTING

- Assumption: the k-dimensional lattice augmented with long-range links, each at the stationary distribution of the dynamical process **M&F**.
- Theorem The expected number of steps of greedy routing from any source node s to any target node t at distance d is $O(\log^{2+\varepsilon} d)$.
- Proof just needs $f(\mathbf{d}) \geq \Omega(1/(||\mathbf{d}||^k \log^{1+\varepsilon}||\mathbf{d}||))$

PROOF SKETCH



$$|B| \,/\, ||\mathbf{d}||^k = \Omega(1/\log^{1+\varepsilon}||\mathbf{d}||)$$

OTHER APPLICATIONS

- Spatial gossip
- Opportunistic routing

  1st ACM Workshop on Online Social Networks (WOSN 2008)

- Perspective: All moving nodes

SUMMARY

Navigability in acquaintanceships networks emerges from the combination of two simple processes:

- a spatial process (simple random walk)
- a temporal process (harmonic forgetting)

Conclusion and further works

Combinatorial aspects

- Close the gap between
  - $\Omega(n^{1/\sqrt{\log n}})$ lower bound
  - $\tilde{O}(n^{1/3})$ upper bound
- Matrix-based augmentation schemes
  - Universal matrix M
  - Appropriate labeling L of the nodes
- Extension to matroïds

Algorithmic aspects

- Given G, compute φ such that $gd(G,\varphi)$ is minimum (in deterministic or probabilistic settings)
  - Approximation algorithms
  - Exact (exponential) algorithms

Social science aspects

- Are we truly living in a small world?
- Validation of the augmented graph model
  - Long range links extraction algorithms
  - Experiments on real data sets


Teşekür Ederim!