Enhancing Random Walks Efficiency

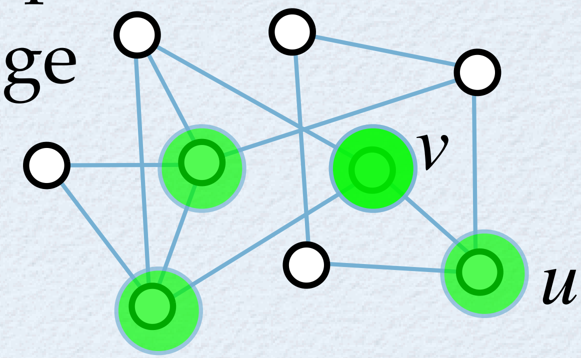
Chen Avin
Department of Communication Systems Engineering
Ben Gurion University

Talk Overview

* Motivation
* **Part I**: Many Random Walks are Faster Than One. [FOCS]
  - ✓ Noga Alon, Michal Koucky, Gady Kozma, Zvi Lotker and Mark Tuttle
* **Part II**: The Power of Choice in Random Walks. [MSWiM06]
  - ✓ Bhaskar Krishnamachari

Motivation

* Random Walk: visiting the nodes of a graph G in some sequential random order; Message moves randomly in the network.



* Random walks in networking setting: [SB02, BE02, DSW02, AB04, GMS04, SKH05, ASS06]
  - ✓ Searching, routing, query mechanism, self-stabilization.
  - ✓ Sensor networks, P2P, distributed systems, the Web.
* Why to use random-walk-based applications?
* How efficient is this process?

Cover Time

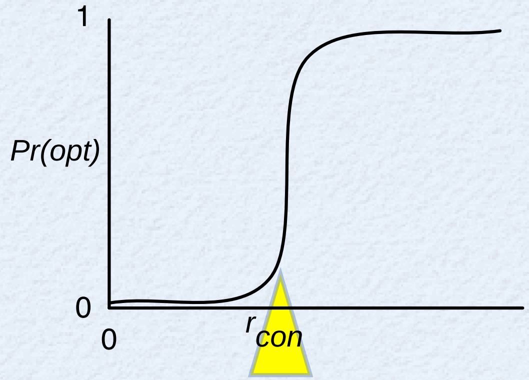
* **Cover time** $C(G)$: the expected time (messages) to visit all nodes in a graph G in a simple random walk. (starting at the worst node).
* **Hitting time**: $h(u,v)$. $h_{\max} = \max_{u,v} h(u, v)$
* **Mixing time**.
* Cover time, Known results:
  - ✓ Best cases: highly connected graphs. $\Theta(n \cdot \log n)$.
  - ✓ Worst cases: bottlenecks exist. $\Theta(n^3)$.
  - ✓ Random geometric graphs (random WSN). $\Theta(n \cdot \log n)$.

Cover Time of RGG

* What is the minimum radius r_{opt} that will guarantee w.h.p. that the random geometric graph has an optimal cover time of $\Theta(n \cdot \log n)$ and optimal partial cover time of $\Theta(n)$?

* Main Result: Theorem [AE05]:

$$r_{opt} = \Theta(r_{con})$$



*Diameter is long, $1/r = \Theta((n/\log n)^{½})$, the 2-dimensional grid is not optimal

The Speed Up Question

* Random walk is a sequential process, imply long delay in some applications.

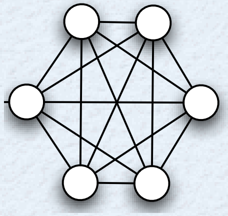
* **Can multiple walks reduce the latency**?
  * ✓ A similar question was first asked in [BKRU, FOCS 89].
  * ✓ What is the cover time, $C^k(G)$, of k parallel random walks, all starting at the same location?

* What is the **speed-up** of k-random walks on a graph G?

$$S^k(G) = \frac{C(G)}{C^k(G)}$$

The Answer is not Obvious

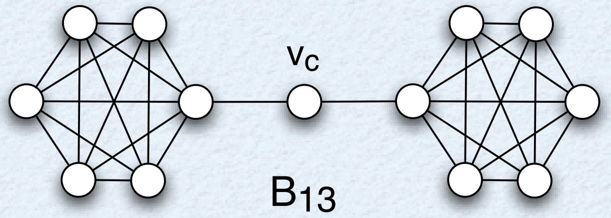
* First, easy case, Clique K_n of size n.



$$S^k(K_n) = k$$

The Answer is not Obvious

- What happens on the barbell B_n of size n?



- The cover time is $\Theta(n^2)$.

- $k=O(\log n)$ walks starting at v_c: w.h.p each clique will have $O(\log n)$ walks after the first step.

- w.h.p each clique is covered in $O(n)$ steps so, $C_{v_c}^{20 \log n}(B_n) = O(n)$

- This leads to an exponential speed up: $S_{v_c}^{k}(B_n) = \Omega(2^k)$

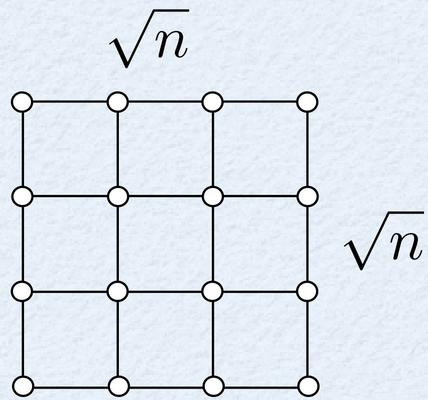
The Answer is not Obvious

- What happened on the path L_n of length n?

- The cover time is known to be $O(n^2)$.
- The probability for a single walk to cover the line in n steps is 2^{-n} .
- So, $2^{\Omega(n)}$ walks are needed for a linear speed up.
- Generalize for k, the speed up $S^k(L_n)=\Theta(\log k)$.

The Answer is not Obvious

- What happens on the 2-dimensional grid G_n?



Logarithmic ?

Exponential?

Linear?

- The cover time is $O(n \log^2 n)$.

- We show that for $\epsilon > 0, \; k \leq (\log n)^{1-\epsilon}$

$$S^k(G_n) = k - o(1)$$

Main Results

- **Theorem**: For a large collection of graphs, as long as *k* is not too big there is a speed up of *k*-o(1).

- The collection includes all graphs for which there is a **gap** between the cover time *C* and h_{max} and *k* such:

$$\frac{C}{h_{\max}} \to \infty \qquad k \leq (C/h_{\max})^{1-\epsilon}$$

  - d-dimensional grids for d ≥ 2
  - d-regular balanced trees for d≥2
  - Erdős-Rényi random graphs
  - random geometric graphs

$$S^k = k - o(1)$$

$$k \leq (\log n)^{1-\epsilon}$$

Proof for linear speed-up

* **Matthews' bound** [MAT88]: for any graph *G*

$$C(G) \leq h_{max} \cdot H_n$$

* **Theorem**: For any graph *G* and $k \leq log\ n$

$$C^k(G) \leq \frac{e + o(1)}{k} \cdot h_{max} \cdot H_n$$

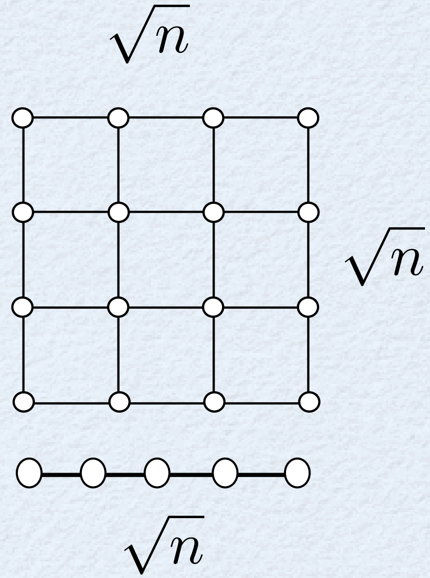
* **Corollary**: when Matthews' bound is tight there is a **linear** speed up

Proof

- $\Pr[u \rightsquigarrow v \geq e \cdot h_{\max}] \leq \frac{1}{e}$ (Markov inequality) - a trial
- $\Pr[u \rightsquigarrow v \geq er \cdot h_{\max}] \leq \frac{1}{e^r}$ (independent r trials)
- $\Pr[\text{k walks: } u \rightsquigarrow v \geq er \cdot h_{\max}] \leq \frac{1}{e^{kr}}$ $\leq \frac{1}{n \ln^2 n}$
- $r = \lceil (\ln n + 2 \ln \ln n)/k \rceil$
- $\Pr[\text{k walks of } erh_{\max} \text{ from u covers the graph}] \geq 1 - \frac{1}{\ln^2 n}$
- $C^k(G) \leq erh_{\max} + C(G)/\ln^2 n$
- $C^k(G) \leq (e + o(1))h_{\max} H_n / k$ $\square$

The Result is Tight

* For the 2-dimensional grid (torus)
  - ✓ $k \leq (\log n)^{1-\epsilon}$ gives $S^k = k - o(1)$
  - ✓ $k \geq \Omega(\log^3 n)$ gives $S^k = o(k)$
* On the grid could, $C^{\log^3 n} = n/\log n$?
* On the cycle could, $C^{\log^3 n} = n/\log n$?



* Fore a more restricted families of graphs we can have linear speed up for a larger *k*.

Expanders

- **Theorem:** For expanders there is a **linear** speed up for $k \leq n$.

- Proof idea: having a much stronger bound on the hitting probability (instead of markov inequality we used earlier).

- The cover time is $O(n \log n)$, diameter $O(\log n)$.

- The result is tight: for $k = \omega(n)$ the k cover time cannot be $C^k = o(\log n)$.

Part I: Open Problems

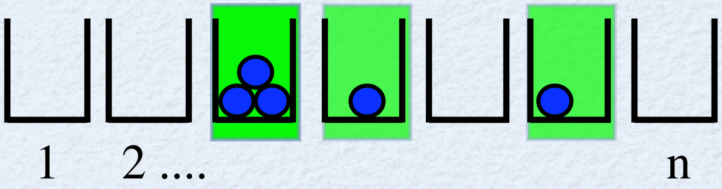
* What is the graph property that captures the speed up of k random walks?
* Is speed-up always $\Omega(\log k)$?
* Is speed-up always $O(k)$?
* ....

Talk Overview

* Motivation
* Part I: Many Random Walks are Faster Than One.
* **Part II: The Power of Choice in Random Walks**

The Power of Choice

- “Balls in Bins” (n balls, n bins)

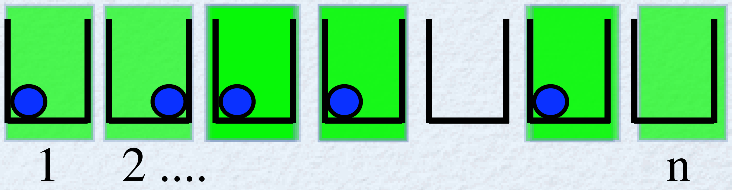


- The most loaded bin is

$$\frac{\log n}{\log \log n}$$

- “Balanced Allocation” [ABKU94]: Adding Choice of d.
- The most loaded bin with choice of d >1:

$$\frac{\log \log n}{\log d}$$



- n → ∞. unbounded improvement! diminishing returns.

RWC(d): Random Walk with Choice

* Observation: "Balls in Bins" is a random walk on the complete graph!

* Idea: add choice for RW on arbitrary graph

* Algorithm:

RWC(d)

1. update the number of visits
2. select d neighbors independently and uniformly with replacement.
3. step to the node that minimizes (# visits)/(node degree)

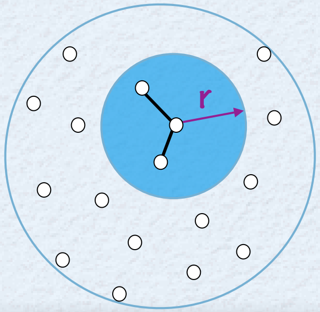
Theoretical results

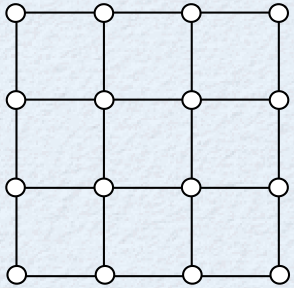
- Lemma: On a complete graph RWC(d) gives improvement of order d:

$$\frac{\text{cover time of SRW}}{\text{cover time of RWC(d)}} \approx d$$

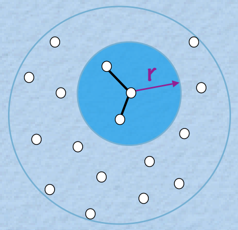
- Question: Can we get unbounded improvement for other graphs ?

- We present simulation results on Random Geometric Graphs (RGG) and Grids.





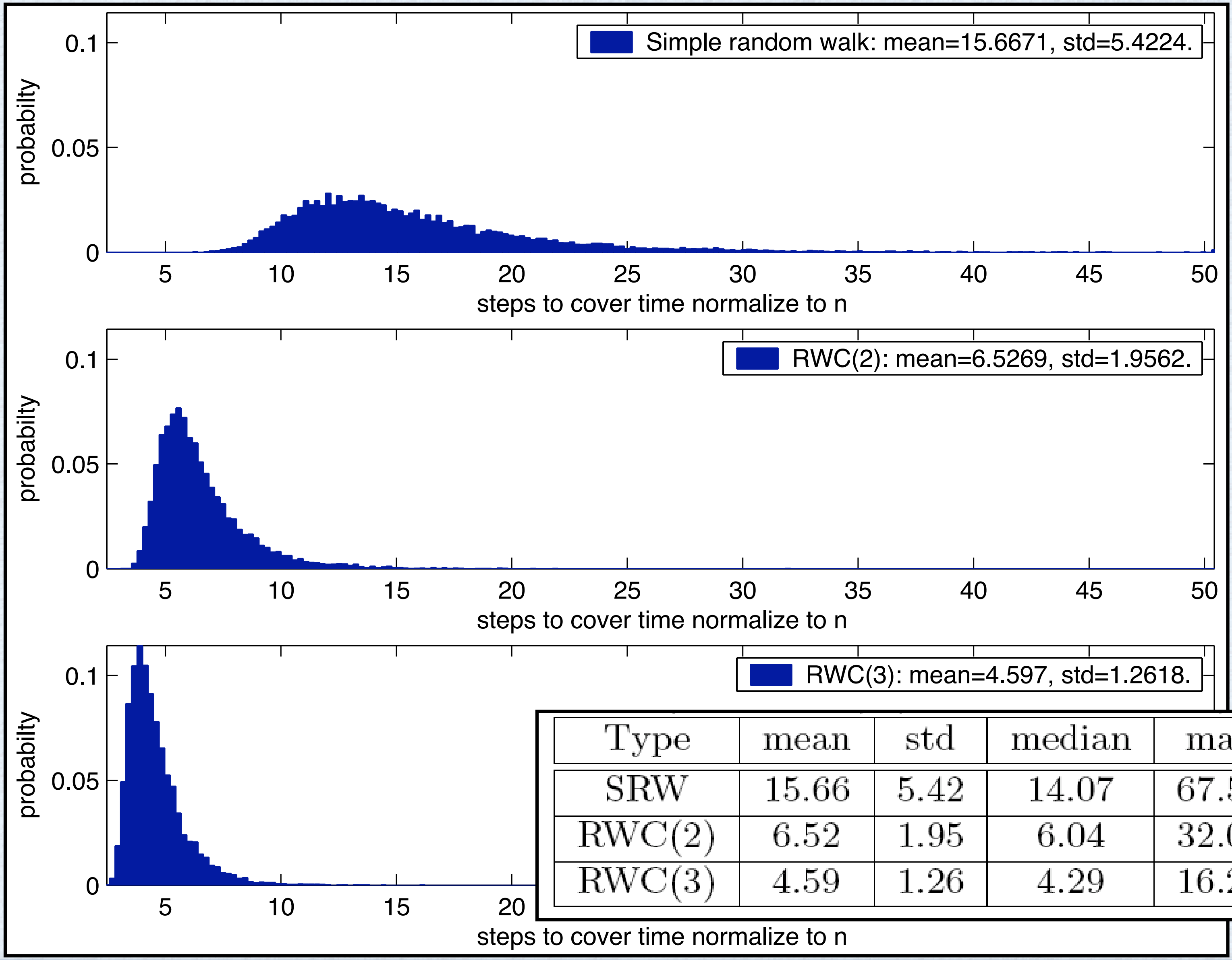
Random Wireless Network: Cover time distribution



SRW

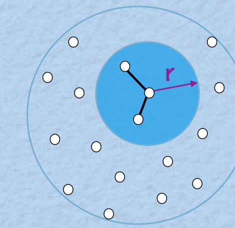
RWC(2)

RWC(3)

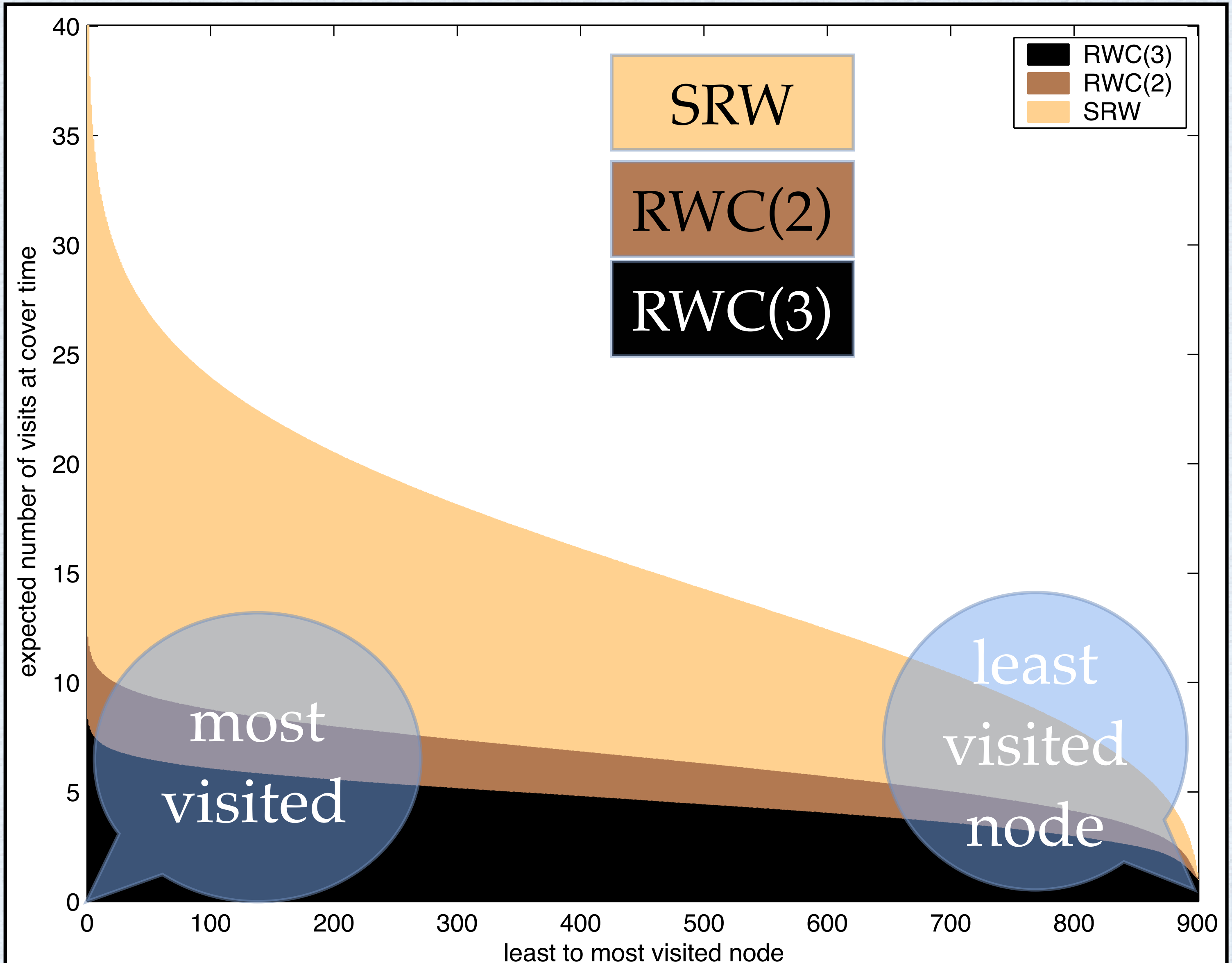


Type	mean	std	median	max	95%
SRW	15.66	5.42	14.07	67.59	25.99
RWC(2)	6.52	1.95	6.04	32.03	10.25
RWC(3)	4.59	1.26	4.29	16.21	7.03

Random Wireless Network: Number of visits at cover



number of visits



SRW

RWC(2)

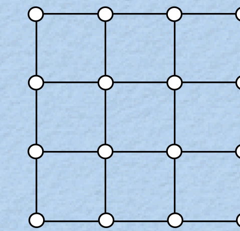
RWC(3)

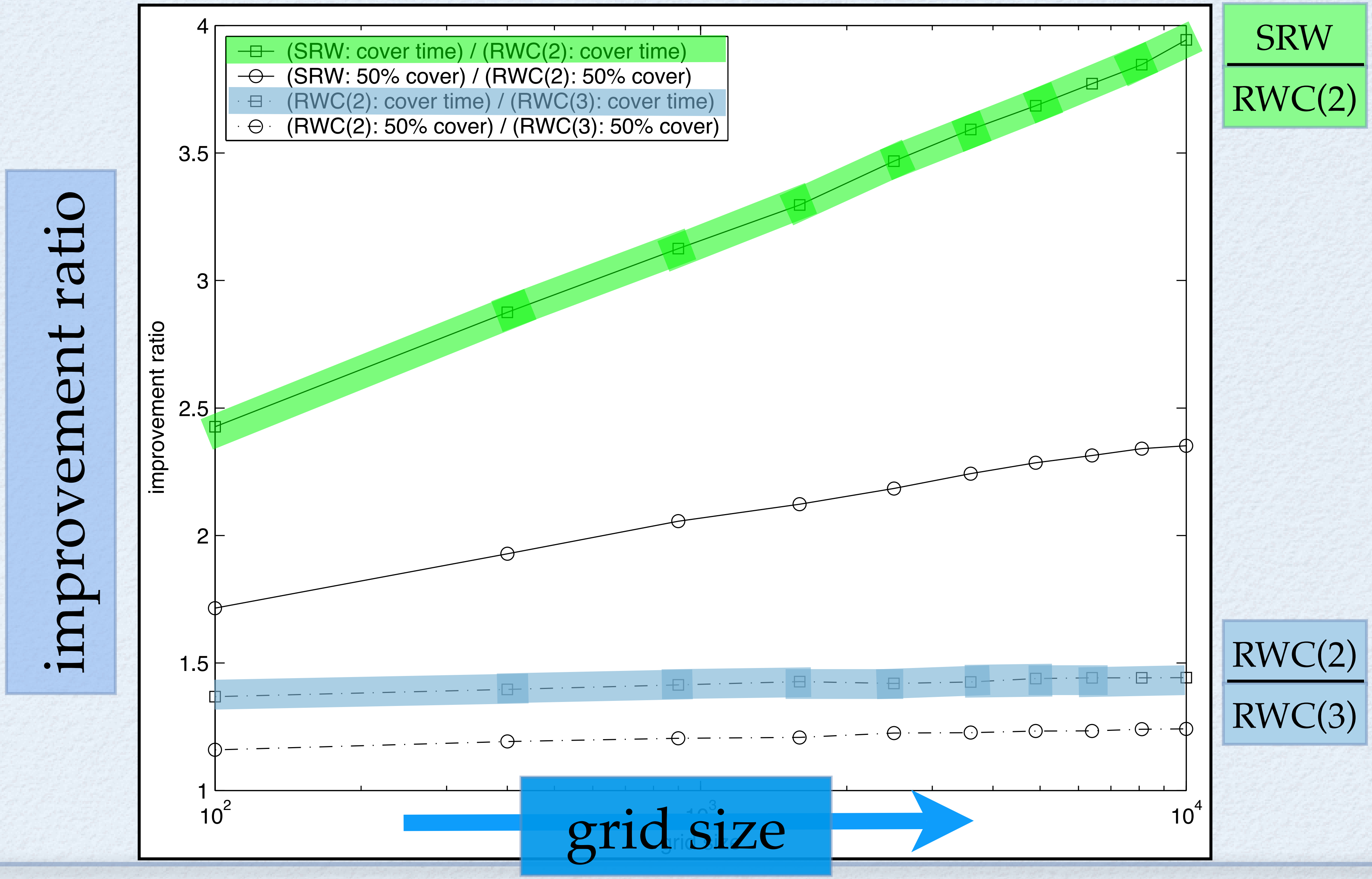
most visited

least visited node

Nodes

2D grids of different sizes: improvement ratio in cover time





Part II: Open Problems

* Prove the power of choice in random walks.
* What is the cover time speed-up in RWC(d)?
* What is the mixing time of RWC(d)?
* What is the stationary (empirical) distribution of RWC?
* What is the decrease in the most visited node in RWC(d)?
* ....

Thank You