Mixing time on random walks

陈 敏
Min Chen

mc190@york.ac.uk

Univ.of York

research.btexact.com

2004 Nov 19 13:30

Outline

- What is a random walk?
- How to measure a random walk?
- Fastest mixing problem
- Computational results
- Applications
- Future work

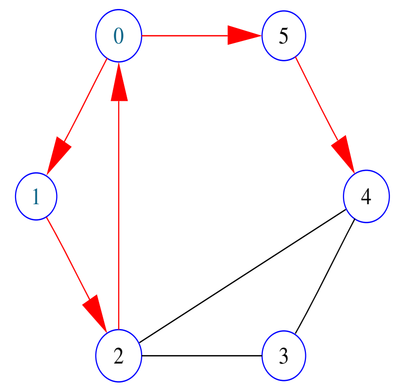
mixing time $\xrightarrow{describe}$ convergence speed

fast mixing $\longrightarrow$ fast convergence
of dynamical process on network

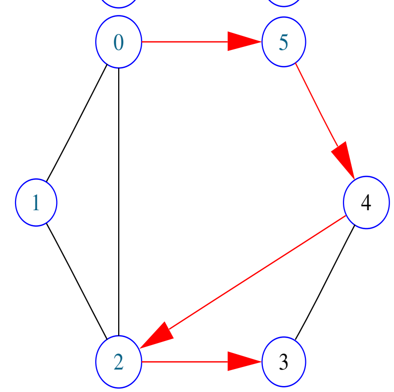
Random walk

- $\Gamma(v)$: neighbors of v in G
- $x_0 \xrightarrow{\text{random}} x_1 \in \Gamma(x_0) \xrightarrow{\text{random}} x_2 \in \Gamma(x_1) \xrightarrow{\text{random}} \cdots \xrightarrow{\text{random}} x_k \in \Gamma(x_{k-1})$
- Example:

  $\{v_0 v_1 v_2 v_0 v_5 v_4\}$ is one trial of a random walk



  $\{v_0 v_5 v_4 v_2 v_3\}$ is another trial of a random walk

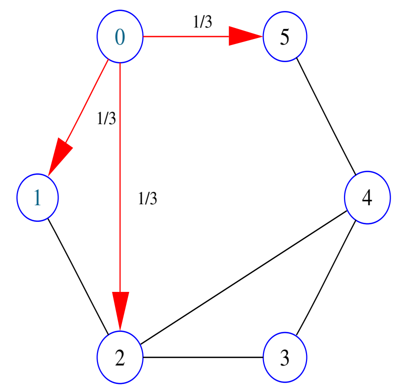


Random walk and Markov chain

- Transition probability matrix P

$$P_{ij} = \begin{cases} Pr[x_{t+1} = j | x_t = i], & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}$$

- We can choose each node uniformly



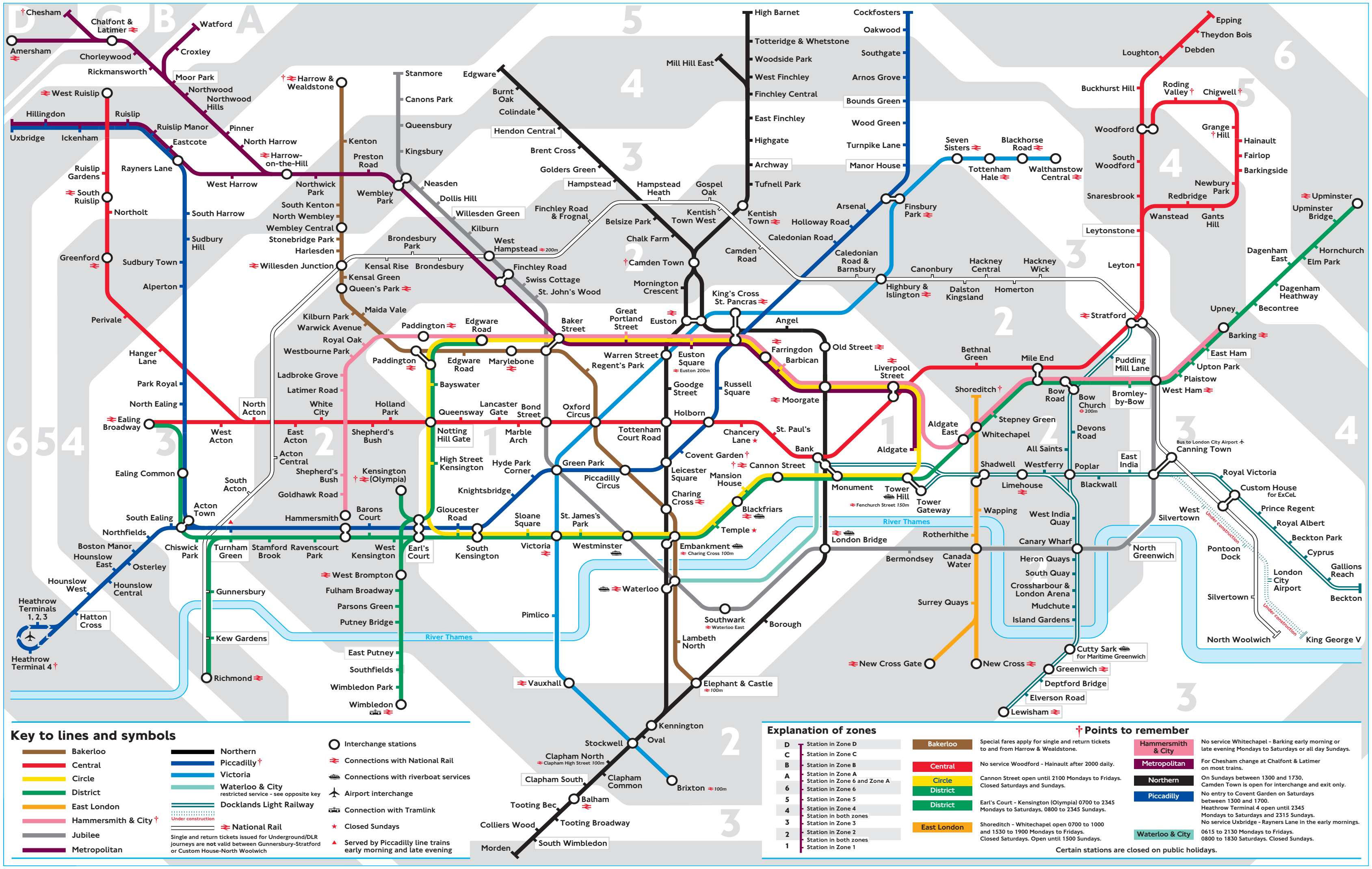
$$P_{ij} = \begin{cases} 1/d(i), & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}$$

where $d(i)$ is the degree of i

- distribution at t: $\pi(t) = \pi(0)P^t$
- equilibrium distribution: $\pi = \pi P$
- reversible Markov chain: $\pi_i P_{ij} = \pi_j P_{ji} \quad i, j \in V$
- symmetric chain: $P^T = P$

Measurements of Random walk

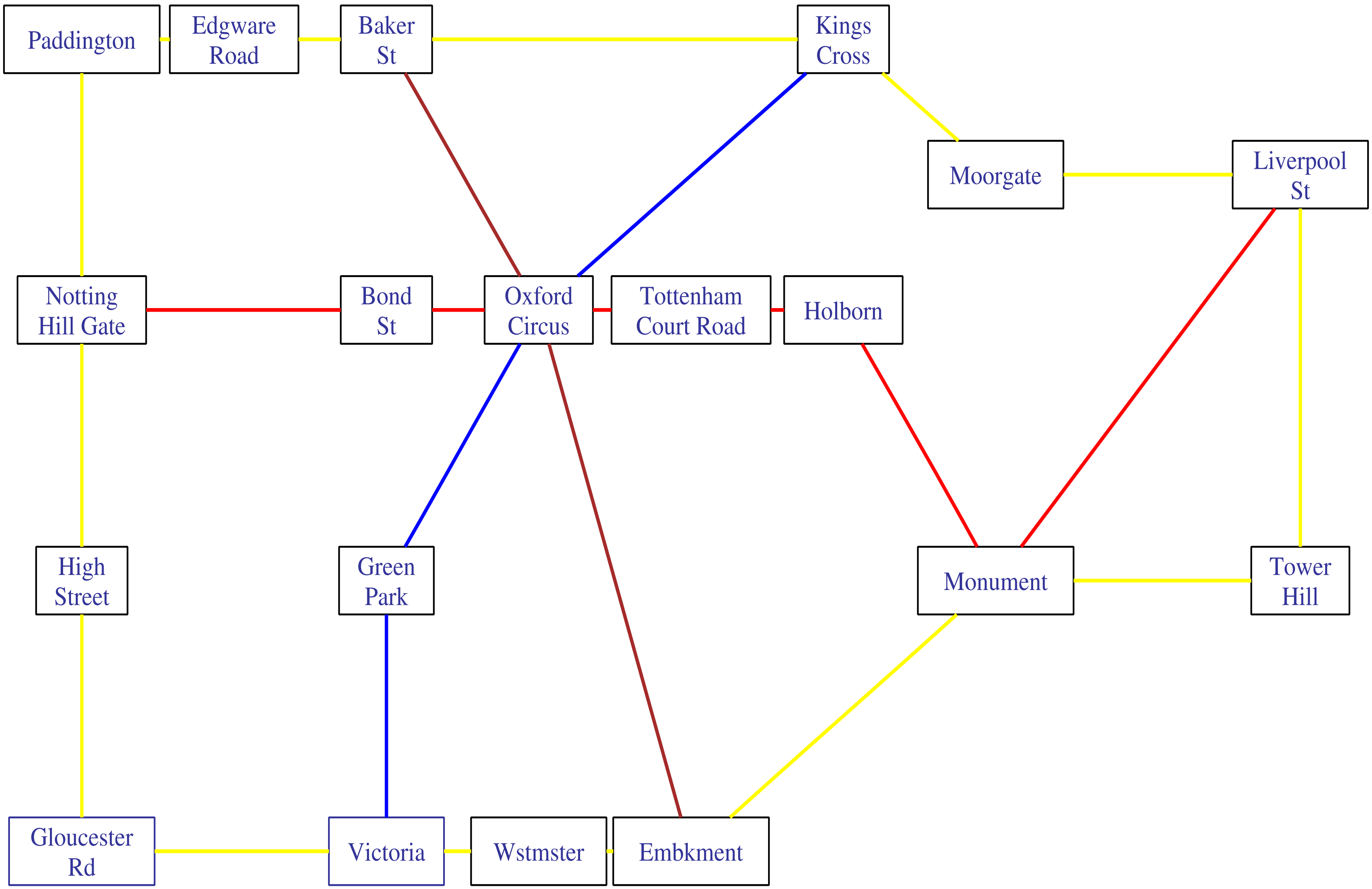
- Example: London tube graph



● Transport for London

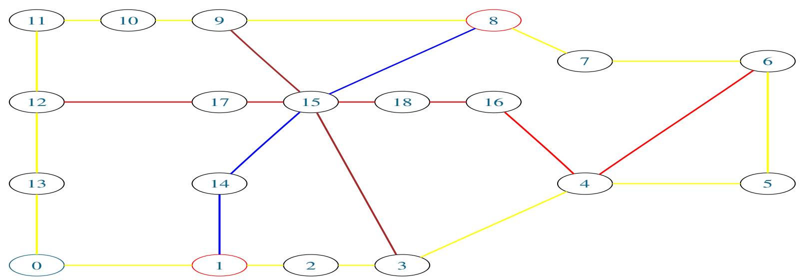
Example: London tube graph (continued)

Pick out a simple graph from it:



Example: London tube graph - Hitting time

- King's Cross $\underbrace{\xrightarrow{\text{random walk}}}_{\text{? steps}}$ Victoria



- Hitting time:[LL00]

$$H_{ij} = 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left[\frac{v_{ki}^2}{d(i)} - \frac{v_{ki} v_{kj}}{\sqrt{d(i)d(j)}} \right]$$

$\{\lambda, v\}$ is the eigensystem of $N = D^{1/2} A D^{1/2}$, A is the adjacency matrix, $D = \text{diag}(1/d(i))$, m is the number of edges in the graph

- the mean number of steps between King's Cross and Victoria is

$$H_{kc-v} = H_{81} = 38.5$$

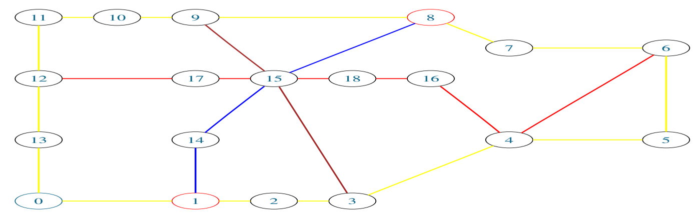
Example: London tube graph - Hitting time matrix

- Hitting time matrix

0.0	12.7	30.1	24.9	36.1	58.8	47.5	49.7	31.2	30.0	46.5	45.0	...
28.1	0.0	21.3	20.1	32.6	55.4	44.4	47.4	29.6	29.8	49.6	51.4	...
40.6	16.4	0.0	11.0	27.0	50.5	40.0	44.7	28.7	29.9	51.3	54.5	...
51.1	30.9	26.7	0.0	19.5	43.5	33.7	40.1	25.8	28.1	50.9	55.7	...
58.4	39.4	38.7	15.5	0.0	26.4	18.9	32.5	25.4	30.3	54.0	59.6	...
60.3	41.5	41.4	18.8	5.7	0.0	10.5	27.8	24.4	30.8	54.8	60.7	...
60.2	41.7	42.2	20.1	9.4	21.6	0.0	21.1	21.5	29.2	53.5	59.7	...
58.8	41.1	43.3	23.0	19.4	35.4	17.6	0.0	11.8	23.5	48.8	55.9	...
55.5	38.5	42.5	23.9	27.4	47.2	33.2	26.9	0.0	15.9	42.0	50.1	...
53.3	37.7	42.7	25.2	31.4	52.6	39.9	37.7	14.9	0.0	29.8	41.6	...
50.1	37.8	44.4	28.3	35.4	56.9	44.5	43.3	21.4	10.1	0.0	21.8	...
45.0	36.0	44.0	29.4	37.4	59.1	47.1	46.8	25.8	18.2	18.2	0.0	...
37.8	32.1	41.6	28.6	37.3	59.4	47.7	48.3	28.3	24.4	34.3	26.2	...
19.9	23.4	36.8	27.7	37.7	60.1	48.6	50.0	30.7	28.2	41.4	36.6	...
40.7	17.9	30.8	21.3	31.5	54.0	42.7	44.7	26.0	26.4	48.0	51.6	...
51.3	33.7	38.4	20.4	28.4	50.7	39.0	40.0	20.3	21.0	44.5	49.8	...
58.0	39.5	40.6	19.2	11.5	36.5	27.6	37.0	25.7	29.2	52.8	58.3	...
45.6	33.9	41.0	25.5	33.9	56.0	44.3	45.2	25.3	23.7	40.4	39.0	...
55.7	37.6	40.5	20.8	21.0	44.6	34.3	39.5	24.0	26.1	49.7	55.1	...

Example: London tube graph - Commute time

- King's Cross $\underbrace{\xrightarrow{\text{rw}} \text{Victoria} \xrightarrow{\text{rw}}}_{\text{? steps}}$ King's Cross



- commute time [LL00]

$$k_{ij} = H_{ij} + H_{ji}$$

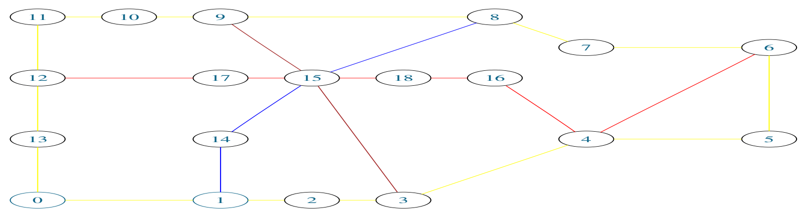
- So we know that the commute time from King's Cross to Victoria then back to King's Cross is

$$k_{k-v} = k_{81} = H_{81} + H_{18} = 38.5 + 29.6 = 68.1$$

- Symmetry
  - $H_{ij} \neq H_{ji}$ unless i and j are vertex-transitive.
  - $k_{ij} = k_{ji}$ for all i, j

Example: London tube graph - Mixing time

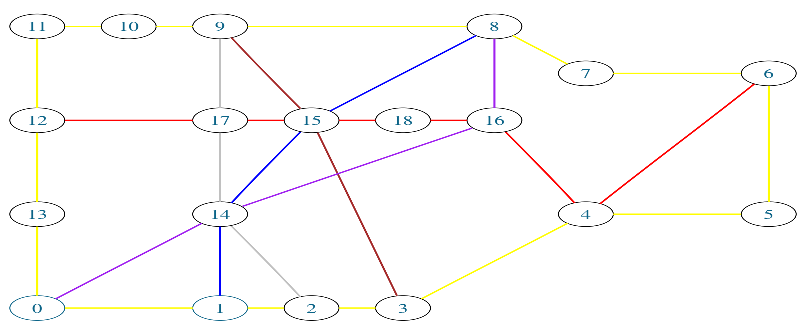
- $\underbrace{\pi(0) \rightarrow \pi(1) \rightarrow \cdots \rightarrow \pi}_{?\ \text{steps}}$



- $mixing\ rate = \log(1/\mu(P))$[SB04]

  where $\mu(P) = \max_{i=2,\dots,n} |\lambda_i(P)| = \max\{\lambda_2(P), -\lambda_n(P)\}$
- mixing time: $\tau = 1/(\text{mixing rate}) = 1/\log(1/\mu)$[SB04]
- $mixing\ rate = 0.101568$

  $mixing\ time = 9.845655$
- build new lines



$mixing\ rate = 0.168313$

$mixing\ time = 5.941294$

Definition of random walk properties

hitting time (access time)	H_{ij} is the expected number of steps in a random walk starting from node i and before node j.	$2m\sum_{k=2}^{n}\frac{1}{1-\lambda_k}\left[\frac{v_{ki}^2}{d(i)}-\frac{v_{ki}v_{kj}}{\sqrt{d(i)d(j)}}\right]$
commute time	k_{ij} is the expected number of steps in a random walk starting at i, the first time return to i via j.	$H_{ij}+H_{ji}$
mixing rate	measure of how fast the random walk converges to its stationary distribution.	$\rho=-\log(\mu(P))$
mixing time	the time scale (in steps) for reaching the stationary distribution.	$\tau=-\frac{1}{\log(\mu)}$

Fastest mixing problem - topology

- Mixing time - convergence speed
- fastest mixing - smallest mixing time/different topology

change the topological structure of graphs

↓

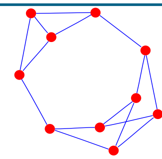
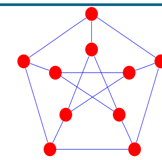
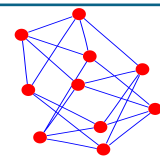
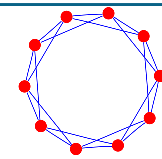
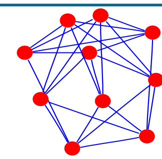
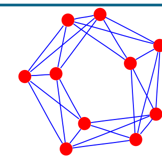
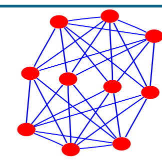
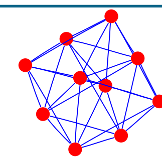
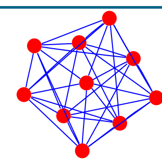
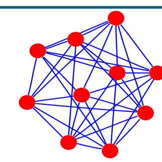
fastest mixing chain

- Optimization description

$$
\begin{array}{ll}
\min_G & \mu(P(G)) \\
\text{s.t.} & P_{\cdot} \geqslant 0 \\
 & P\mathbf{1} = \mathbf{1} \\
 & P_{ij} = \begin{cases} 1/d(i), & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}
\end{array}
$$

Computational results - small regular graphs

- the min/max mixing time for 10 nodes regular graphs

n	deg	num	maxtime	graph	mintime	graph	avertime
10	3	17	15.5896		2.4663		6.8216
10	4	58	7.7220		1.7195		3.4093
10	5	59	7.4542		1.2427		2.2145
10	6	21	2.4663		0.9102		1.5722
10	7	5	1.1802		1.0168		1.1475

Fastest mixing problem - weights

- Stephen Boyd's idea, Stanford Univ.
- fastest mixing - smallest mixing time/different weights

fix the topology, changing the weights

$\downarrow$ *(reversible chain)*

fastest mixing chain

- Optimization description

$$
\begin{array}{ll}
\min_P & \mu(P) \\
\text{s.t.} & P_{\cdot} \geqslant 0 \\
 & P\mathbf{1} = \mathbf{1} \\
 & \Pi P = P^T \Pi \\
 & P_{ij} = 0, \quad i,j \notin E
\end{array}
$$

Fastest mixing problem - combined problem

- Combined Boyd's idea together with ours
- Fastest mixing chain both on topology and weights

all different topologies reversible chain

↓ *change the weights*

fastest mixing weights on each different topology

↓

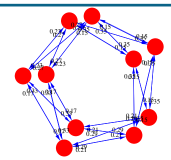
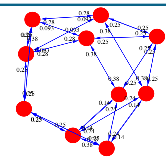
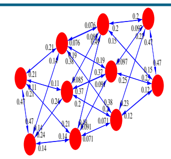
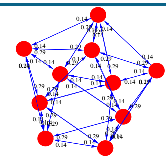
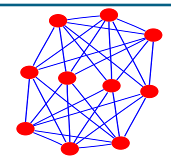
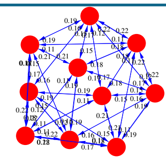
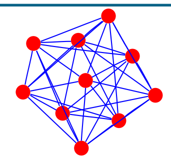
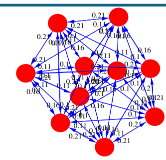
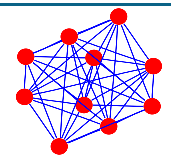
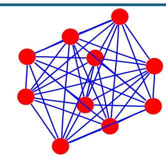
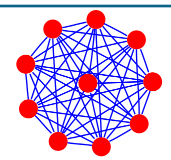
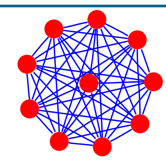
fastest mixing chain

- Optimization description

$$\begin{array}{ll} \min_G \min_{P(G)} & \mu(P) \\ \text{s.t.} & P_{\bullet} \geqslant 0 \\ & P\mathbf{1} = \mathbf{1} \\ & \Pi P = P^T \Pi \\ & P_{ij} = 0, \quad i,j \notin E \end{array}$$

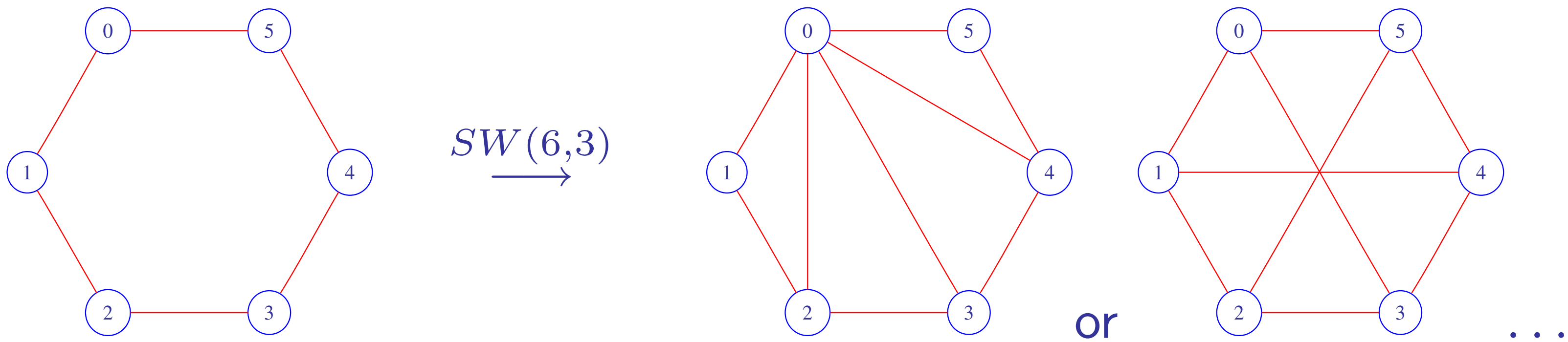
FMRMC for 10 nodes regular graphs

The mixing time for 10 nodes regular graphs

nodes	degree	max time	graph	min time	graph
10	4	4.7185		1.5767	
10	5	4.3522		1.1802	
10	6	2.4663		0.9064	
10	7	1.1802		0.7670	
10	8	0.7213		0.7213	
10	9	0.4551		0.4551	

Random graph results - small world $SW(n,m)$

- $SW(n,m)$ model: n nodes cycle + m more cross links



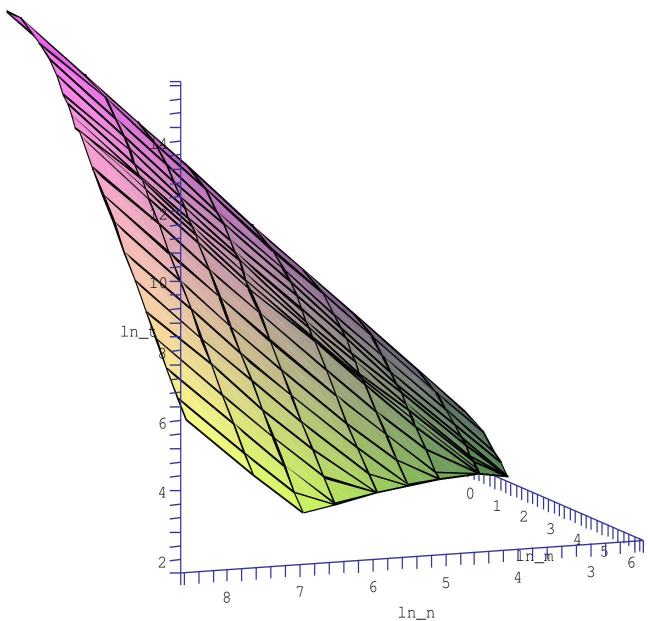
- (n,m) ⇒topology structure⇒mixing time
- How does n affect mixing time? $n \xrightarrow{?} \tau$
- How does m affect mixing time? $m \xrightarrow{?} \tau$

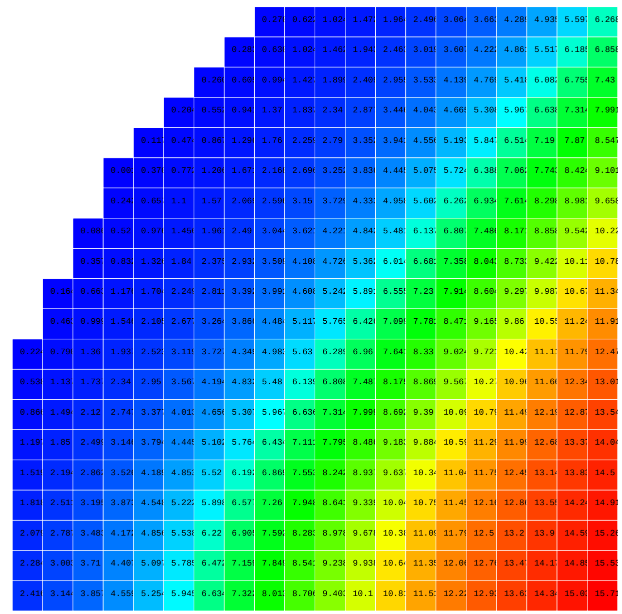
Small world $SW(n,m)$ - n and τ

- For cycle: $\mu = \max\{|\lambda_2|, |\lambda_n|\}) = \left|\cos\left(\frac{(n-1)\pi}{n}\right)\right| = \cos\left(\frac{\pi}{n}\right)$
  Substitute into τ:
  $$\tau(\mu) = \frac{1}{\rho(\mu)} = \frac{-1}{\log(\mu)} = \frac{-1}{\log\cos\left(\frac{\pi}{n}\right)}$$
  Expanded as Taylor expansion:
  $$\tau(n) \approx \frac{2n^2}{\pi^2} - \frac{1}{3} - \frac{\pi^2}{30n^2} - \frac{5\pi^4}{756n^4} + O(n^{-6})$$
  mixing time τ is in direct proportion to n^2 as $n \to \infty$:
  $$\sqrt{\frac{\tau}{2}}\pi \approx n$$

Topology $\stackrel{?}{\rightarrow}$ mixing time $(SW(n,m))$

- Colour plot of $\log(\tau)$ against $\log(m)$ and $\log(n)$



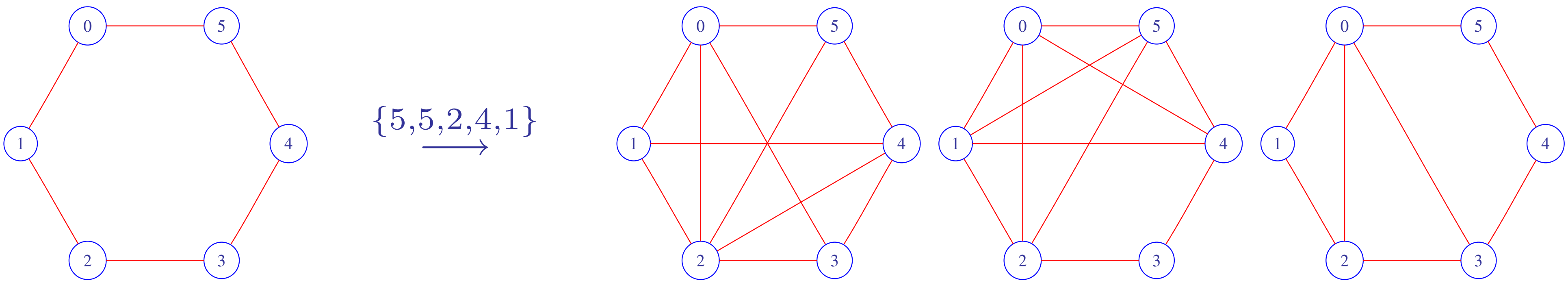


- $\tau \approx O(m^{k_m}) \approx O(n^{k_n})$

Computational results - small world $SW\{n,p\}$

- $SW\{n,p\}$ model:

  n node cycle $\Rightarrow\ \{m_0, m_1, \ldots, m_k\} \in B(\binom{n}{2}-n, p)\ \Rightarrow\ SW(n, m_0), \ldots, SW(n, m_k)$

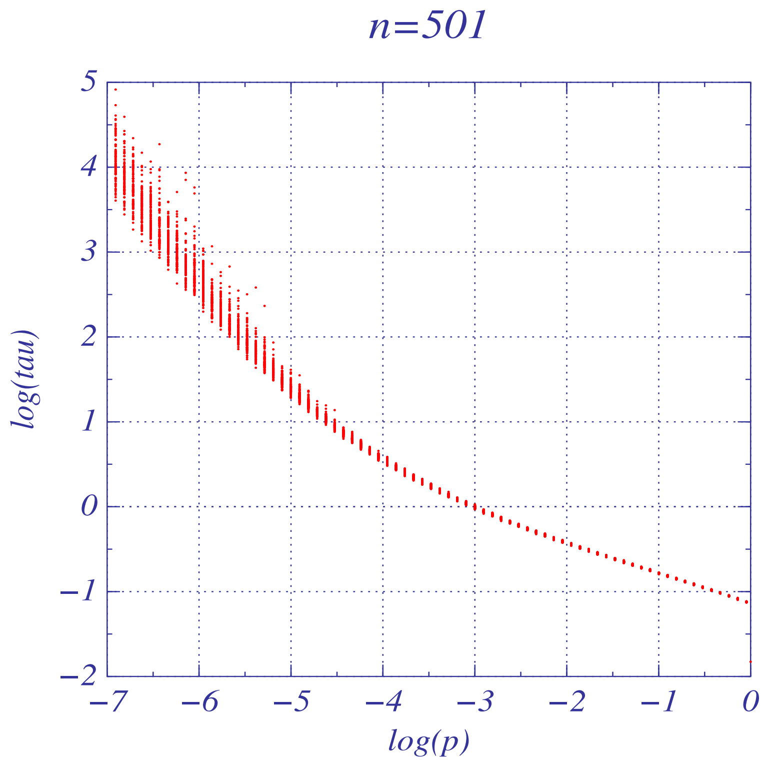


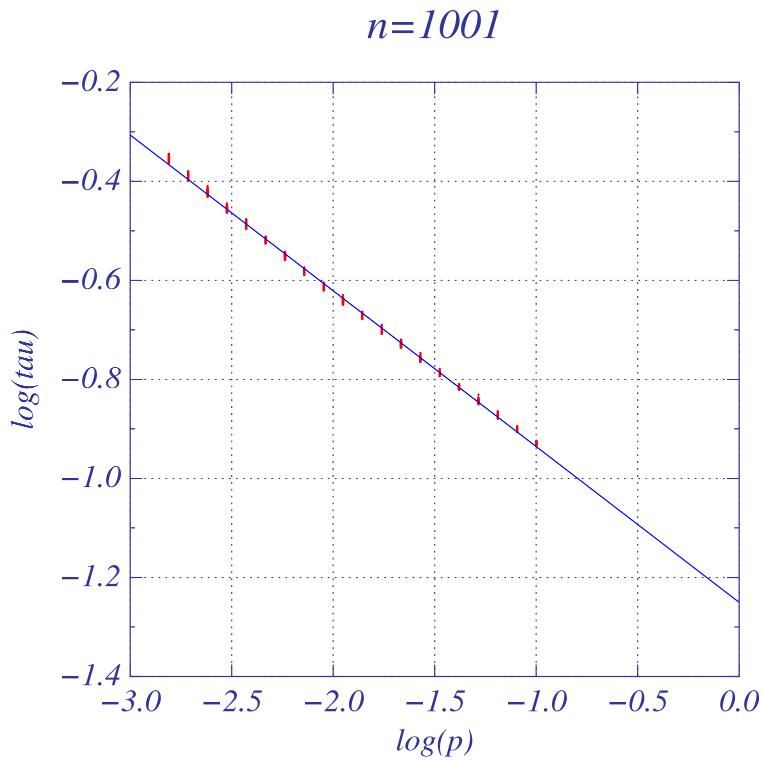
  $SW\{6, 0.2\} \Rightarrow\ SW(6,5), SW(6,5), SW(6,2), \ldots$

- What's the relation between p and mixing time?

Topology $\stackrel{?}{\longrightarrow}$ mixing time $(SW\{n,p\})$

- Mixing time of $SW\{n,p\}$ graphs for $n = 501$ and $n = 1001$





- $\tau \approx O(p^{k_p}) \approx O(m^{k_m}) \approx O(n^{k_n})$

Applications

- Information exchange

  —bounds of the transformation time

- Search engine (Google)

  —How fast can google rank the pages?

- Sampling problem

  —rapid simulation with good results

Future work

- different graph model

$$\begin{cases} \text{Scale-free graph} \\ \text{Geometric random graph } G(n,r) \\ \text{Grid (mesh) graph} \end{cases}$$

- directed graph

- relation between *average hitting time* and *mixing time*

References

- R. Diestel, Graph Theory, Springer 2000
- B. Bollobás, Random Graphs, CUP 2001
- S. Boyd, P. Diaconis & L. Xiao, Fastest Mixing Markov Chain on A Graph
- L. Lovász, Random walks on graphs: a survey
- Fan Chung & S-T Yau 2000 Discrete Green's functions
- N. Biggs, Algebraic graph theory, CUP 1993
- B. Bollobás, Modern graph theory, Springer-Verlag 2002

typeset 2004 November 10 in pdfLaTeX on a linux system