6.046- *Design and Analysis of Algorithms*

Lecture 09

Markov Chain Monte Carlo

(supplementary material to these slides has also been posted)

Prof. Constantinos Daskalakis

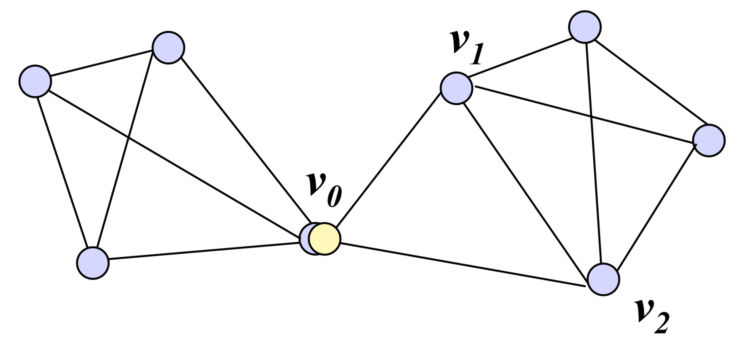
Menu

- Random walks on graphs
- Markov Chains
- Examples:
  - pagerank
  - card-shuffling
  - colorings

Menu

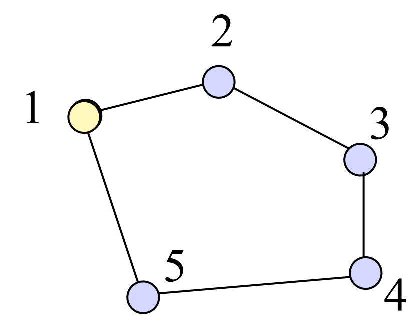
- **Random walks on graphs**
- Markov Chains
- Examples:
  - pagerank
  - card-shuffling
  - colorings

Random Walks



- Given undirected graph $G = (V, E)$
- A squirrel stands at vertex v_0 :
- Squirrel ate fermented pumpkin so doesn't know what he's doing
- So jumps to random neighbor v_1 of v_0
- Then jumps to random neighbor v_2 of v_1
- etc
- Question: Where is squirrel after t steps?
- A: At some random location.
- OK, with what probability is squirrel at each vertex of the graph?
- Want to compute $x_t \in \mathrm{R}^n$, where
- $x_t(i)$: probability squirrel is at node i at time t.
- v_t : random variable representing location at time t.

$x_t \to x_{t+1}$?



- Simplification: all nodes have same degree *d, e.g.*
- $x_0 = (1, 0, 0, 0, 0)$
- $x_0 \to x_1$?
- if $u_1, u_2, \ldots, u_d$ are the d neighbors of v_0, then
- $v_1 = u_i$ with probability $1/d$
- so $x_1 = (0, ½ , 0, 0, ½)$
- $x_2 = (½, 0, ¼, ¼, 0)$
- …
- $A = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$ (adjacency matrix divided by d)

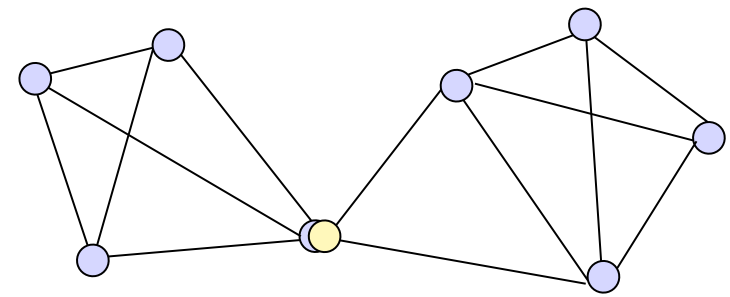
$$x_1 = x_0 A$$
$$x_2 = x_1 A = x_0 A^2$$
$$x_3 = x_2 A = x_0 A^3$$
$$\vdots$$
$$x_{t+1} = x_t A = x_0 A^{t+1} \qquad (1)$$

"the transition matrix"

A_{ij} : probability of jumping to j if squirrel is at i

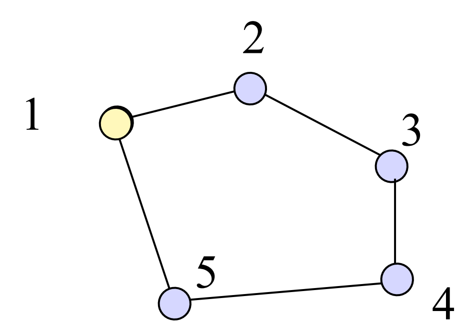
formally $A_{ij} = \Pr[v_{t+1} = j \mid v_t = i]$

x_t



- More general undirected graphs?
- *Transition Matrix:*

  A =adjacency matrix where row i is divided by the degree d_i of i
- $x_t = x_0 A^t$
- Computing x_t fast?
  - repeated squaring!
  - compute $A \rightarrow A^2 \rightarrow A^4 \rightarrow \ldots \rightarrow A^t$
    (if t is a power of 2; if not see lecture 2)
  - then do vector-matrix product
- Limiting distribution x_t as $t \rightarrow \infty$?
- e.g. what is x_∞ in 5-cycle?
- $x_\infty = (1/5, 1/5, 1/5, 1/5, 1/5)$

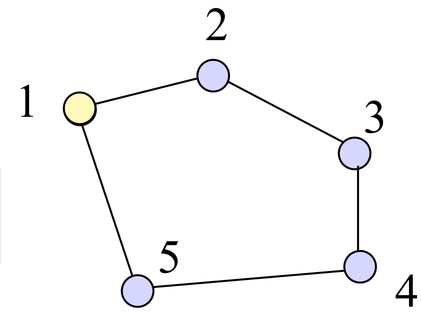


Verifying $x_t \to (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$

- Recall

$$A = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

$$\boxed{x_t = x_0 A^t}$$

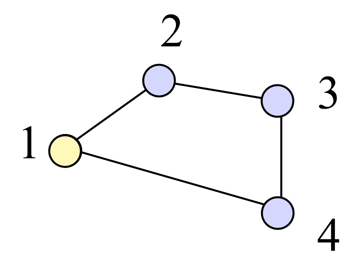


- $x_0 = [1 \quad 0 \quad 0 \quad 0 \quad 0]$
- $x_1 = [0 \quad 0.5000 \quad 0 \quad 0 \quad 0.5000]$
- $x_2 = [0.5000 \quad 0 \quad 0.2500 \quad 0.2500 \quad 0]$
- $x_3 = [0 \quad 0.3750 \quad 0.1250 \quad 0.1250 \quad 0.3750]$
- $x_4 = [0.3750 \quad 0.0625 \quad 0.2500 \quad 0.2500 \quad 0.0625]$
- $x_5 = [0.0625 \quad 0.3125 \quad 0.1562 \quad 0.1562 \quad 0.3125]$
- $x_6 = [0.3125 \quad 0.1094 \quad 0.2344 \quad 0.2344 \quad 0.1094]$
- $x_7 = [0.1094 \quad 0.2734 \quad 0.1719 \quad 0.1719 \quad 0.2734]$
- $x_8 = [0.2734 \quad 0.1406 \quad 0.2227 \quad 0.2227 \quad 0.1406]$
- $x_9 = [0.1406 \quad 0.2480 \quad 0.1816 \quad 0.1816 \quad 0.2480]$
- $x_{10} = [0.2480 \quad 0.1611 \quad 0.2148 \quad 0.2148 \quad 0.1611]$
- $x_{11} = [0.1611 \quad 0.2314 \quad 0.1880 \quad 0.1880 \quad 0.2314]$
- $x_{12} = [0.2314 \quad 0.1746 \quad 0.2097 \quad 0.2097 \quad 0.1746]$
- $x_{13} = [0.1746 \quad 0.2206 \quad 0.1921 \quad 0.1921 \quad 0.2206]$
- $x_{14} = [0.2206 \quad 0.1833 \quad 0.2064 \quad 0.2064 \quad 0.1833]$

$x_{15} = [0.1833 \quad 0.2135 \quad 0.1949 \quad 0.1949 \quad 0.2135]$
$x_{16} = [0.2135 \quad 0.1891 \quad 0.2042 \quad 0.2042 \quad 0.1891]$
$x_{17} = [0.1891 \quad 0.2088 \quad 0.1966 \quad 0.1966 \quad 0.2088]$
$x_{18} = [0.2088 \quad 0.1929 \quad 0.2027 \quad 0.2027 \quad 0.1929]$
$x_{19} = [0.1929 \quad 0.2058 \quad 0.1978 \quad 0.1978 \quad 0.2058]$
$x_{20} = [0.2058 \quad 0.1953 \quad 0.2018 \quad 0.2018 \quad 0.1953]$
$x_{21} = [0.1953 \quad 0.2038 \quad 0.1986 \quad 0.1986 \quad 0.2038]$
$x_{22} = [0.2038 \quad 0.1969 \quad 0.2012 \quad 0.2012 \quad 0.1969]$
$x_{23} = [0.1969 \quad 0.2025 \quad 0.1991 \quad 0.1991 \quad 0.2025]$
$x_{24} = [0.2025 \quad 0.1980 \quad 0.2008 \quad 0.2008 \quad 0.1980]$
$x_{25} = [0.1980 \quad 0.2016 \quad 0.1994 \quad 0.1994 \quad 0.2016]$

Example 2: 4-cycle

- $x_t \rightarrow$??
- in this case: $A = \begin{pmatrix} 0 & ½ & 0 & ½ \\ ½ & 0 & ½ & 0 \\ 0 & ½ & 0 & ½ \\ ½ & 0 & ½ & 0 \end{pmatrix}$



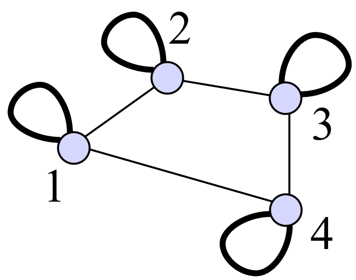
- Let's iterate $x_t = x_0 A^t$
- $x_0 = [1 \quad 0 \quad 0 \quad 0]$
- $x_1 = [0 \quad ½ \quad 0 \quad ½]$
- $x_2 = [½ \quad 0 \quad ½ \quad 0]$
- $x_3 = [0 \quad ½ \quad 0 \quad ½]$
- $x_4 = [½ \quad 0 \quad ½ \quad 0]$
- $x_5 = [0 \quad ½ \quad 0 \quad ½]$
- $x_6 = [½ \quad 0 \quad ½ \quad 0]$
- …
- Ouch:

  $x_{2t} \rightarrow [½ \quad 0 \quad ½ \quad 0]$

  $x_{2t+1} \rightarrow [0 \quad ½ \quad 0 \quad ½]$
- So no limiting distribution…
- Issue: ***Periodicity***

Avoiding Periodicity

- What can we do to this graph to make the walk aperiodic?



Called a ***"lazy* random walk"**

- (one self-loop would suffice)

- in this case: $A_{\text{new}} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix} = \frac{2}{3} A_{\text{old}} + \frac{1}{3}\, I$

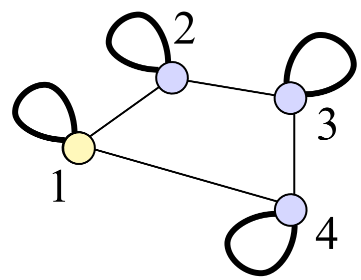
- $x_t \rightarrow$??
- Let's iterate $x_t = x_0 A_{\text{new}}^{\ t}$
- ....
- $x_t \rightarrow [\frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4} \quad \frac{1}{4}]$

$$A_{\text{old}} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Proving $x_t \to (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$?

- Recall

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

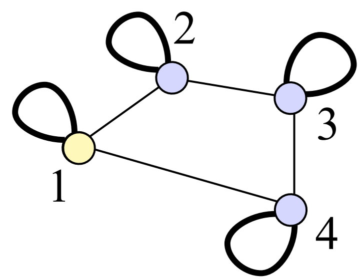


- Idea: look at the ***eigenvalues*** of A
- A symmetric so it has 4 real eigenvalues
- $\lambda_1 = 1,\ \lambda_2 = \lambda_3 = \frac{1}{3},\ \lambda_4 = -\frac{1}{3}$ (thanks Matlab)
- coincidence: $\lambda_2 = \lambda_3$ and $\lambda_4 = -\lambda_2$ (4-cylce is a special graph)
- non-coincidence (true for any ***lazy*** r.w. on an undirected ***connected*** graph):
  - largest eigenvalue =1
  - all others have absolute value <1
- left eigenvector corresponding to $\lambda_1 = 1$?
- $e_1 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ is a left eigenvector for λ_1
- Wow. Why would $x_t \to e_1$ as $t \to \infty$?

Proving $x_t \to (¼, ¼, ¼, ¼)$?

- Recall

$$A = \begin{pmatrix} ⅓ & ⅓ & 0 & ⅓ \\ ⅓ & ⅓ & ⅓ & 0 \\ 0 & ⅓ & ⅓ & ⅓ \\ ⅓ & 0 & ⅓ & ⅓ \end{pmatrix}$$



- $\lambda_1 = 1,\ \lambda_2 = \lambda_3 = ⅓, \lambda_4 = -⅓$
- $e_1 = (¼, ¼, ¼, ¼)$
- Proof (see also notes): choose eigenvectors e_2, e_3, e_4 for $\lambda_2, \lambda_3, \lambda_4$, so that $\{e_1, e_2, e_3, e_4\}$ is a basis of R^4 (guaranteed by the spectral theorem, since A is symmetric)
- so $x_0 = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4$, for some a_1, a_2, a_3, a_4
- Now $x_t = x_0 A^t =$

$$= a_1 e_1 A^t + a_2 e_2 A^t + a_3 e_3 A^t + a_4 e_4 A^t$$
$$= a_1 e_1 \lambda_1{}^t + a_2 e_2 \lambda_2{}^t + a_3 e_3 \lambda_3{}^t + a_4 e_4 \lambda_4{}^t$$
$$\to a_1 e_1, \text{ as } t \to \infty$$

exact same proof for any lazy r.w. on any connected undirected graph.

- since $e_1 = (¼, ¼, ¼, ¼)$ is a distribution, it must be that $a_1 = 1$
- Hence $x_t \to (¼, ¼, ¼, ¼)$, as $t \to \infty$.

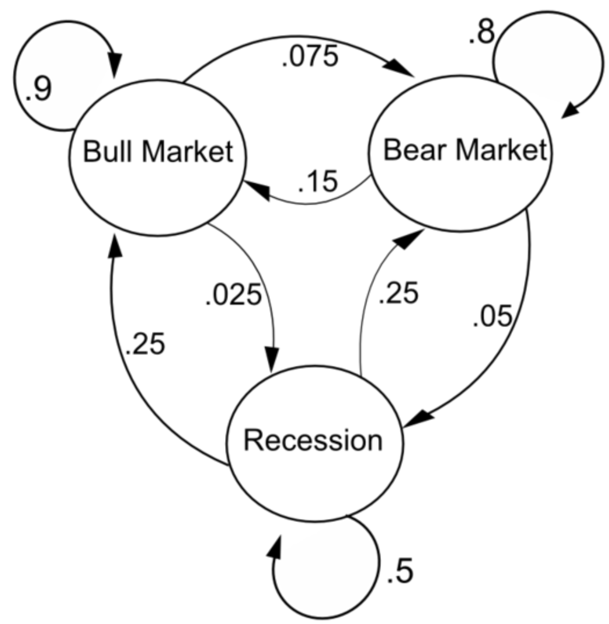
Menu

- Random walks on graphs
- **Markov Chains**
- Examples:
  - pagerank
  - card-shuffling
  - colorings

General Setup

- Let G =(V, E) be a:
  - (i) ***strongly-connected*** weighted directed graph (i.e. there is a path from every vertex to every vertex),
  - with (ii) ***self-loop*** on every vertex (to avoid periodicities)
- The weight p_{ij} of every edge $(i, j) \in E$ represents a probability, namely:
  - p_{ij} : probability of transitioning from i to j.
  - Hence for all i: $\sum_{(i,j)\in E} p_{ij} = 1.$
- Transition matrix: A = [p_{ij}].
- In the example on the left:

$$A= \begin{bmatrix} .9 & .075 & .025 \\ .15 & .8 & .05 \\ .25 & .25 & .5 \end{bmatrix}$$



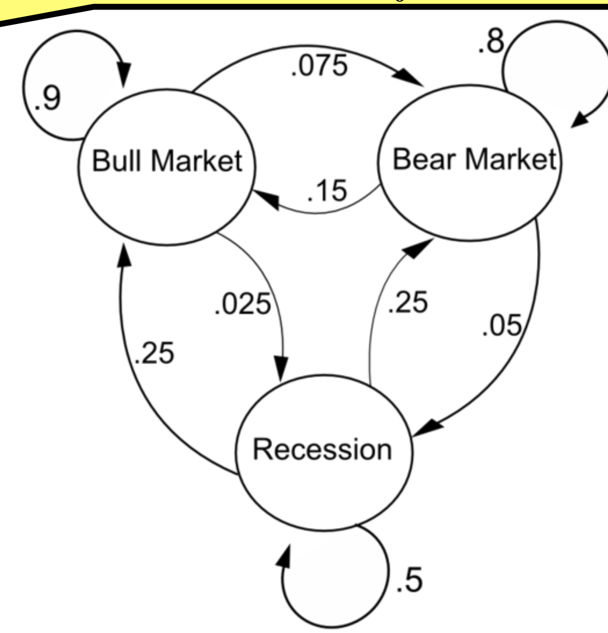
General Setup

- Transition matrix: $A = [\ p_{ij}\ ]$.
- Still true that, if x_0 is the distribution over vertices at time 0, then $x_t = x_0\, A^t$.
- **Claim**: A has eigenvalue 1 (with multiplicity 1), and there is a unique positive vector x such that:
  - $x \cdot A = A$ (*)
  - $\sum_i x_i = 1$ (**)

Note that if a limiting distribution $\boldsymbol{x}$ exists, then it definitely needs to satisfy (*), o.w. it wouldn't be a limiting distribution. The point of the theorem is that the Markov Chain indeed converges to the unique eigenvector of A satisfying (*) and (**), no matter what x_0 is.

Theorem: For any x_0, $x_t \rightarrow \boldsymbol{x}$ *as* $t \rightarrow \infty$.

- $\boldsymbol{x}$ is called the "**stationary distribution of *G***"



Two obvious Questions:

- why is x_∞ interesting?
- how fast does $x_t \rightarrow x_\infty$?

Menu

- Random walks on graphs
- Markov Chains
- Examples:
  - **pagerank**
  - card-shuffling
  - colorings

Pagerank

- No better proof that something is useful than having cool applications ☺
- It turns out that Markov Chains have a famous one: ***PageRank***.
- PageRank of a webpage $w \approx$ Probability that a web-surfer starting from some central page (e.g. Yahoo!) and clicking random links arrives at webpage w.
- How to compute this probability?
- Form graph G = the hyperlink graph;

- Namely, G has a node for every webpage, and there is an edge from webpage w_1 to webpage w_2 *iff* there is a hyperlink from w_1 to w_2.
- All outgoing links from a webpage w are equally probable.
- Compute stationary distribution x_∞, i.e. the left eigenvector of the transition matrix A of G, corresponding to eigenvalue 1.
- Pagerank of page $w = x_\infty(w)$.

Computing Pagerank

- Graph G = the hyperlink graph
- Compute stationary distribution of G, i.e. the left eigenvector of the (normalized by out-degrees) adjacency matrix A of G, corresponding to eigenvalue 1.
- How to compute stationary distribution?
- **Idea 1:** Crawl the web, create giant A, solve eigenvalue problem.
- Runtime $O(n^3)$ using Gaussian elimination
- this is too much for n = size of the web.
- **Idea 2:** Simulate the walk sufficiently many times (theory meets practice)
  - Start at some central page and do random walk for sufficiently many steps;
  - Restart and repeat sufficiently many times;
  - then take PageRank(w) ≈ empirical frequency that random walk ended at w.
- Hope that empirical distribution is good approximation to stationary distribution for the right choice of "sufficiently many" above…
- or at least for the top components of the stationary distribution, which are the most important for ranking the top results.

Menu

- Random walks on graphs
- Markov Chains
- Examples:
  - pagerank
  - **card-shuffling**
  - colorings

Card Shuffling

Q: Why shuffle the deck?

A: well, to start from a uniformly random permutation of the cards

Q2: How many permutations are there?

A: $52! \approx 2^{257} \approx 10^{77}$ – how large is that?

Q3: Getting a random permutation?

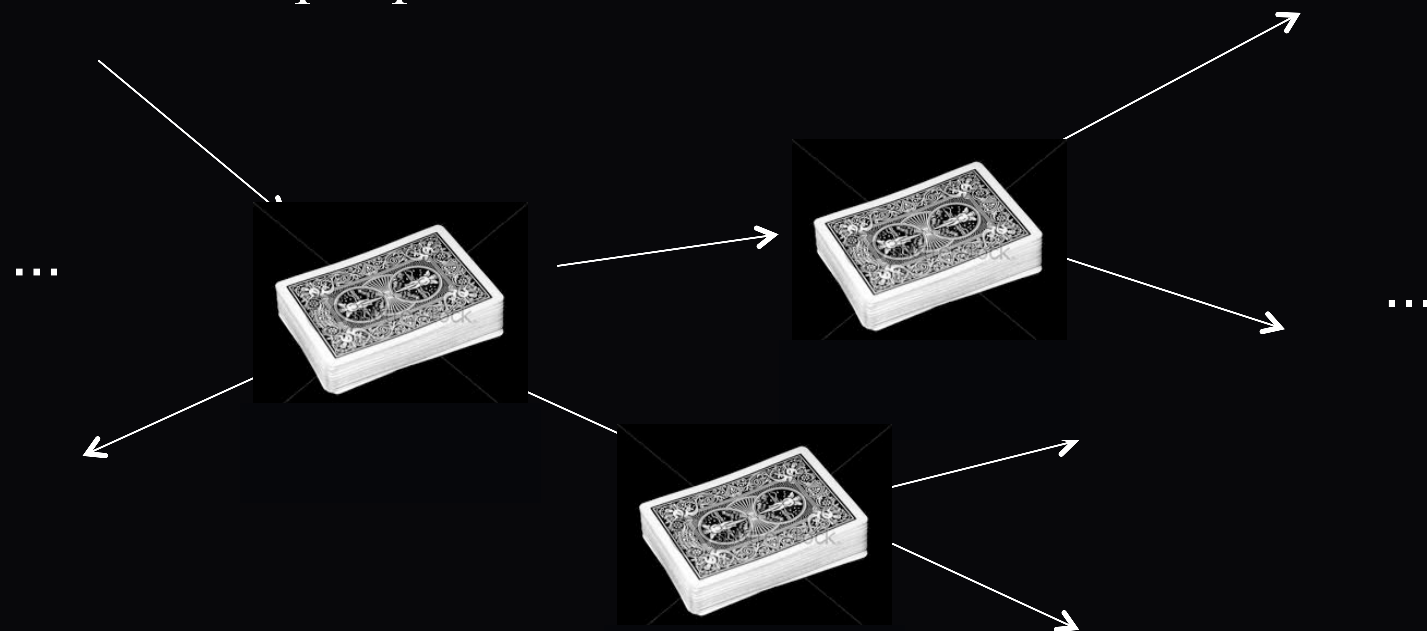
- soln1: dice 10^{77} faces
- soln2: shuffle $\approx$ dice

Example Shuffles:

- top-in-at-random
- riffle-shuffle

Shuffling as a Markov Chain

Graph: - one node per permutation of the deck.



- edge (u,v): v is reachable in one move from u (specific to shuffle)

While performing the shuffle we jump from node to node.

Q: Stationary distribution of a correct shuffle?

Probability 1/52! on each permutation.

Mixing Time

Q: How many steps suffice for uniformity?

A1: The question is meaningless as no matter how long you shuffle there is always a small trace of what permutation you started from in x_t.

Q: OK, how many steps suffice to be close to uniform?

- Top-in-at-Random: ~300 repetitions suffice

  namely distance(x_{300}, *uniform*) < 1%

- Riffle Shuffle: ~10 repetitions suffice

  namely distance(x_{10}, *uniform*) < 1%

i.e. different shuffles have different graphs, and as a result converge to uniform distribution at different speeds.

Mixing Time (formally)

- Let $x_0, x_1, x_2, \ldots, x_t, \ldots$ be a Markov Chain with stationary distribution x_∞.
- **Def:** The *mixing time* of the Markov Chain is the minimum τ such that $d(x_\tau, x_\infty) < .01$.

  where for two distributions x, y: $d(x,y) = \sum_i |x(i) - y(i)|$.

- Captures rate at which $x_t \rightarrow x_\infty$. Namely:
- **Claim:** $d(x_t, x_\infty) \leq \exp(-t/\tau_{mix})$.

- When we design Markov chains we desire τ_{mix} to be small.
- τ_{mix} depends on the connectivity of the Markov Chain.

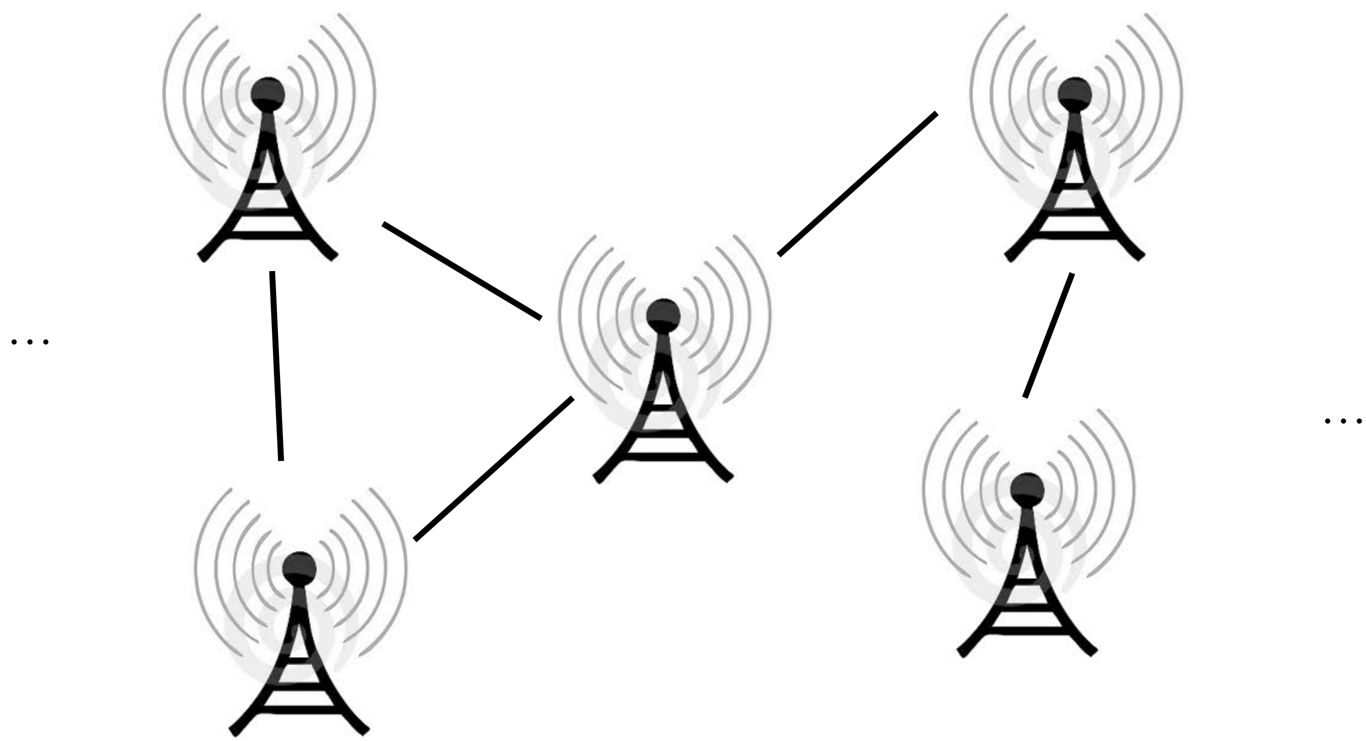
Menu

- Random walks on graphs
- Markov Chains
- Examples:
  - pagerank
  - card-shuffling
  - **colorings**

Allocating Frequencies to Stations

Input: - Graph $G = (V, E)$, V: radio stations, E: interferences
- A set $F=\{f_1, f_2, \ldots, f_q\}$ of frequencies.

Goal: Assign frequencies to stations so that no two interfering stations get the same frequency.



continued on the board (see notes)