Markov Chain Monte Carlo

6.046 Design and Analysis of Algorithms
MIT

Michael Kapralov

6 March 2014
Random walks on undirected graphs

Markov chains

Examples:
  - pagerank
  - card shuffling

Mixing time
Given undirected graph $G = (V, E)$

- Squirrel stands at vertex $v_0$
- Squirrel ate fermented pumpkin so doesn’t know what he is doing
Given undirected graph $G = (V, E)$

- Squirrel stands at vertex $v_0$
- Squirrel ate fermented pumpkin so doesn’t know what he is doing
- So jumps to a random neighbor $v_1$ of $v_0$
Given undirected graph $G = (V, E)$

- Squirrel stands at vertex $v_0$
- Squirrel ate fermented pumpkin so doesn’t know what he is doing
- So jumps to a random neighbor $v_1$ of $v_0$
- Jumps to a random neighbor $v_2$ of $v_1$
Given undirected graph $G = (V, E)$

Q: where is squirrel after $t$ steps?

A: at some random location!

Q: with what probability is squirrel at vertex $v \in V$ after $t$ steps?

Want to compute $x_t \in \mathbb{R}^n$ such that $x_t(i) = \Pr[squirrel \text{ at node } i \text{ at time } t]$. 

Q: where is squirrel after $t$ steps?
Given undirected graph $G = (V, E)$

Q: where is squirrel after $t$ steps?

A: at some random location!

Q: with what probability is squirrel at vertex $v \in V$ after $t$ steps?

Want to compute $x_t \in \mathbb{R}^n$ such that $x_t(i) = \Pr[\text{squirrel at node } i \text{ at time } t]$.

Let $v_t$ be (random) location at time $t$. 
Given undirected graph $G = (V, E)$

Q: where is squirrel after $t$ steps?

A: at some random location!

Q: with what probability is squirrel at vertex $v \in V$ after $t$ steps?

Want to compute $x_t \in \mathbb{R}^n$ such that

$$x_t(i) = \Pr[\text{squirrel at node } i \text{ at time } t].$$

Distribution of the random walk at time $t$
\( X_t \rightarrow X_{t+1} ? \)

Simplification: all nodes have the same degree \( d \)

\( x_0 = (1,0,0,0,0) \)

If \( u_1, u_2, \ldots, u_d \) are the \( d \) neighbors of \( v_0 \), then \( v_1 = u_i \) with probability \( 1/d \)
$X_t \rightarrow X_{t+1}$?

Simplification: all nodes have the same degree $d$

$x_0 = (1, 0, 0, 0, 0, 0)$

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 $x_1 = (0, 1/2, 0, 0, 1/2)$
$X_t \rightarrow X_{t+1}$?

Simplification: all nodes have the same degree $d$

$x_0 = (1, 0, 0, 0, 0)$

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 $x_1 = (0, 1/2, 0, 0, 1/2)$

$x_2 = (1/2, 0, 1/4, 1/4, 0)$
$X_t \rightarrow X_{t+1}$?

Simplification: all nodes have the same degree $d$

$x_0 = (1, 0, 0, 0, 0)$

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 $x_1 = (0, 1/2, 0, 0, 1/2)$

$x_2 = (1/2, 0, 1/4, 1/4, 0)$
$X_t \rightarrow X_{t+1}$?

Simplification: all nodes have the same degree $d$

$x_0 = (1,0,0,0,0)$

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 $x_1 = (0,1/2,0,0,1/2)$

$x_2 = (1/2,0,1/4,1/4,0)$

Transition matrix

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

$A_{ij} = \Pr[v_{t+1} = j|v_t = i]$
$X_t \rightarrow X_{t+1}$?

Simplification: all nodes have the same degree $d$

$x_0 = (1,0,0,0,0)$

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 $x_1 = (0,1/2,0,0,1/2)$

$x_2 = (1/2,0,1/4,1/4,0)$

**Transition matrix**

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

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

Transition matrix:

$$A = D^{-1} M = \text{adj. matrix with row } i \text{ divided by } d_i$$

Then

$$x_t = x_0 A^t$$
$X_t$

Transition matrix:

$$A = D^{-1} M = \text{adj. matrix with row } i \text{ divided by } d_i$$

Then

$$x_t = x_0 A^t$$

Computing $x_t$ fast?
\( X_t \)

**Transition matrix:**

\[
A = D^{-1} M = \text{adj. matrix with row } i \text{ divided by } d_i
\]

Then

\[
x_t = x_0 A^t
\]

Computing \( x_t \) fast?

- repeated squaring!
- compute
  \[
  A \rightarrow A^2 \rightarrow A^4 \rightarrow \ldots \rightarrow A^t
  \]
- then do matrix-vector product
Transition matrix:

\[ A = D^{-1}M = \text{adj. matrix with row } i \text{ divided by } d_i \]

Then

\[ x_t = x_0 A^t \]

Computing \( x_t \) fast?

- repeated squaring!
- compute
  \[ A \rightarrow A^2 \rightarrow A^4 \rightarrow \ldots \rightarrow A^t \]
- then do matrix-vector product

Limiting distribution as \( t \rightarrow \infty \)?

\[ x_\infty := \lim_{t \rightarrow \infty} x_t, \text{ on a 5-cycle?} \]
\( X_t \)

Transition matrix:

\[ A = D^{-1} M = \text{adj. matrix with row } i \text{ divided by } d_i \]

Then

\[ x_t = x_0 A^t \]

Computing \( x_t \) fast?

- repeated squaring!
- compute
  \[ A \rightarrow A^2 \rightarrow A^4 \rightarrow \ldots \rightarrow A^t \]
- then do matrix-vector product

Limiting distribution as \( t \to \infty \)?

\[ x_\infty := \lim_{t \to \infty} x_t, \text{ on a 5-cycle?} \]

\[ x_\infty = \left( \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right) \]
Verifying $x_t \rightarrow (1/5, 1/5, 1/5, 1/5, 1/5)$

Transition matrix:

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

and

$$x_t = x_0 A^t$$

Run in Matlab:

```matlab
z=[1 0 0 0 0];
for t=1:25,
    disp(z);
    z=z*A;
end;
```
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.3750</td>
<td>0.1250</td>
<td>0.1250</td>
<td>0.3750</td>
</tr>
<tr>
<td>0.3750</td>
<td>0.0625</td>
<td>0.2500</td>
<td>0.2500</td>
<td>0.0625</td>
</tr>
<tr>
<td>0.0625</td>
<td>0.3125</td>
<td>0.1562</td>
<td>0.1562</td>
<td>0.3125</td>
</tr>
<tr>
<td>0.3125</td>
<td>0.1094</td>
<td>0.2344</td>
<td>0.2344</td>
<td>0.1094</td>
</tr>
<tr>
<td>0.1094</td>
<td>0.2734</td>
<td>0.1719</td>
<td>0.1719</td>
<td>0.2734</td>
</tr>
<tr>
<td>0.2734</td>
<td>0.1406</td>
<td>0.2227</td>
<td>0.2227</td>
<td>0.1406</td>
</tr>
<tr>
<td>0.1406</td>
<td>0.2480</td>
<td>0.1816</td>
<td>0.1816</td>
<td>0.2480</td>
</tr>
<tr>
<td>0.2480</td>
<td>0.1611</td>
<td>0.2148</td>
<td>0.2148</td>
<td>0.1611</td>
</tr>
<tr>
<td>0.1611</td>
<td>0.2314</td>
<td>0.1880</td>
<td>0.1880</td>
<td>0.2314</td>
</tr>
<tr>
<td>0.2314</td>
<td>0.1746</td>
<td>0.2097</td>
<td>0.2097</td>
<td>0.1746</td>
</tr>
<tr>
<td>0.1746</td>
<td>0.2206</td>
<td>0.1921</td>
<td>0.1921</td>
<td>0.2206</td>
</tr>
<tr>
<td>0.2206</td>
<td>0.1833</td>
<td>0.2064</td>
<td>0.2064</td>
<td>0.1833</td>
</tr>
<tr>
<td>0.1833</td>
<td>0.2135</td>
<td>0.1949</td>
<td>0.1949</td>
<td>0.2135</td>
</tr>
<tr>
<td>0.2135</td>
<td>0.1891</td>
<td>0.2042</td>
<td>0.2042</td>
<td>0.1891</td>
</tr>
</tbody>
</table>
Example 2: 4-cycle

Transition matrix:

\[
A = \begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
\end{pmatrix}
\]

and

\[x_t = x_0 A^t\]

Run in Matlab:

```matlab
z=[1 0 0 0];
for t=1:25,
    disp(z);
    z=z*A;
end;
```
<table>
<thead>
<tr>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
</tr>
<tr>
<td>0.5000</td>
<td>0</td>
<td>0.5000</td>
<td>0</td>
</tr>
</tbody>
</table>
Example 2: 4-cycle

Transition matrix:

\[
A = \begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix}
\]

and

\[x_t = x_0 A^t\]

Run in Matlab:

```matlab
z=[1 0 0 0 0];
for t=1:25,
    disp(z);
    z=z*A;
end;
Periodicity!
```
Example 2: 4-cycle

Transition matrix:

\[ A = \begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix} \]

and

\[ x_t = x_0 A^t \]

Run in Matlab:

```matlab
z=[1 0 0 0];
for t=1:25,
    disp(z);
    z=z*A;
end;
Periodicity!
```
Example 2: 4-cycle

Transition matrix:

\[
A = \begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix}
\]

and

\[x_t = x_0 A^t\]

Run in Matlab:

```
z=[1 0 0 0 0];
for t=1:25,
    disp(z);
    z=z*A;
end;
```

Periodicity!
Avoiding periodicity

What can we do to make the walk aperiodic?

This is called the lazy random walk (one loop would suffice): 

\[
\begin{bmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{bmatrix} = \left( \frac{2}{3} \right) A_{\text{old}} + \left( \frac{1}{3} \right) I
\]
Avoiding periodicity

What can we do to make the walk aperiodic?

This is called the lazy random walk (one loop would suffice)

$$A_{new} = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix} = (2/3)A_{old} + (1/3)I$$
Avoiding periodicity

What can we do to make the walk aperiodic?

\[
A_{\text{new}} = \begin{pmatrix}
1/3 & 1/3 & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 \\
0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & 1/3 & 1/3
\end{pmatrix} = \left(\frac{2}{3}\right)A_{\text{old}} + \left(\frac{1}{3}\right)I
\]
Proving that $x_t \to (1/4, 1/4, 1/4, 1/4)$?

$$A_{new} = \begin{pmatrix}
1/3 & 1/3 & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 \\
0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & 1/3 & 1/3
\end{pmatrix}$$

and

$$x_t = x_0 A_{new}^t$$

Idea: look at the eigenvalues of $A$ (4 real eigenvalues since $A$ is symmetric)

Matlab calculation: $\lambda_1 = 1, \lambda_2 = \lambda_3 = 1/3, \lambda_4 = -1/3$. 
Proving that $x_t \rightarrow (1/4, 1/4, 1/4, 1/4)$?

$$A_{\text{new}} = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

and

$$x_t = x_0 A_{\text{new}}^t$$

Idea: look at the eigenvalues of $A$ (4 real eigenvalues since $A$ is symmetric)

Matlab calculation: $\lambda_1 = 1, \lambda_2 = \lambda_3 = 1/3, \lambda_4 = -1/3$.

Which of these eigenvalues is most interesting?
Proving that $x_t \to (1/4, 1/4, 1/4, 1/4)$?

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

and

$$x_t = x_0 A_{new}^t$$

Idea: look at the eigenvalues of $A$ (4 real eigenvalues since $A$ is symmetric)

Matlab calculation: $\lambda_1 = 1, \lambda_2 = \lambda_3 = 1/3, \lambda_4 = -1/3$.

Which of these eigenvalues is most interesting?

Why?
Proving that $x_t \to (1/4, 1/4, 1/4, 1/4)$?

$$A_{new} = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

and

$$x_t = x_0 A_{new}^t$$

**Idea:** look at the eigenvalues of $A$ (4 real eigenvalues since $A$ is symmetric)

Matlab calculation: $\lambda_1 = 1, \lambda_2 = \lambda_3 = 1/3, \lambda_4 = -1/3$.

Left eigenvector corresponding to $\lambda_1 = 1$?

$$e_1 = (1/4, 1/4, 1/4, 1/4)$$

Interesting...
Proving that $x_t \rightarrow (1/4, 1/4, 1/4, 1/4)$?

$$A_{new} = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$$

and

$$x_t = x_0 A_{new}^t$$

Express $x_0$ in the eigenbasis of $A_{new}$:

$$x_0 = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4, \quad e_i A_{new} = \lambda_i e_i, i = 1, 2, 3, 4$$
Proving that \( x_t \to (1/4, 1/4, 1/4, 1/4) \)?

\[
A_{\text{new}} = \begin{pmatrix}
1/3 & 1/3 & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 \\
0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & 1/3 & 1/3 \\
\end{pmatrix}
\]

and

\[
x_t = x_0 A_{\text{new}}^t
\]

Express \( x_0 \) in the eigenbasis of \( A_{\text{new}} \):

\[
x_0 = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4, \quad e_i A_{\text{new}} = \lambda_i e_i, i = 1, 2, 3, 4
\]

Now

\[
x_t = x_0 A_{\text{new}}^t = a_1 e_1 A_{\text{new}}^t + a_2 e_2 A_{\text{new}}^t + a_3 e_3 A_{\text{new}}^t + a_4 e_4 A_{\text{new}}^t
\]
Proving that \( x_t \to (1/4, 1/4, 1/4, 1/4) \)?

\[
A_{new} = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}
\]

and

\[
x_t = x_0 A_{new}^t
\]

Express \( x_0 \) in the eigenbasis of \( A_{new} \):

\[
x_0 = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4, \quad e_i A_{new} = \lambda_i e_i, i = 1, 2, 3, 4
\]

Now

\[
x_t = x_0 A_{new}^t = a_1 e_1 A_{new}^t + a_2 e_2 A_{new}^t + a_3 e_3 A_{new}^t + a_4 e_4 A_{new}^t
\]

\[
= a_1 \lambda_1^t e_1 + a_2 \lambda_2^t e_2 + a_3 \lambda_3^t e_3 + a_4 \lambda_4^t e_4
\]
Proving that $x_t \rightarrow (1/4, 1/4, 1/4, 1/4)$?

$$A_{\text{new}} = \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\
0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}$$

and

$$x_t = x_0 A_t^{\text{new}}$$

Express $x_0$ in the eigenbasis of $A_{\text{new}}$:

$$x_0 = a_1 e_1 + a_2 e_2 + a_3 e_3 + a_4 e_4, \quad e_i A_{\text{new}} = \lambda_i e_i, i = 1, 2, 3, 4$$

Now

$$x_t = x_0 A_t^{\text{new}} = a_1 e_1 A_t^{\text{new}} + a_2 e_2 A_t^{\text{new}} + a_3 e_3 A_t^{\text{new}} + a_4 e_4 A_t^{\text{new}}$$

$$= a_1 \lambda_1^t e_1 + a_2 \lambda_2^t e_2 + a_3 \lambda_3^t e_3 + a_4 \lambda_4^t e_4$$

$$\rightarrow a_1 e_1, \quad \text{and must have } a_1 = 1$$
Proving that $x_t \to (1/4, 1/4, 1/4, 1/4)$?

$$A_{\text{new}} = \begin{pmatrix}
1/3 & 1/3 & 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0 \\
0 & 1/3 & 1/3 & 1/3 \\
1/3 & 0 & 1/3 & 1/3
\end{pmatrix}$$

and

$$x_t = x_0 A_{\text{new}}^t$$

Matlab calculation: $\lambda_1 = 1, \lambda_2 = \lambda_3 = 1/3, \lambda_4 = -1/3$.

General facts:
- largest eigenvalue equal to 1
- all other eigenvalues have absolute value $< 1$ if connected

Exact same proof for the lazy r.w. on any connected undirected graph
- Random walks on undirected graphs
- **Markov chains**
- Examples:
  - pagerank
  - card shuffling
- Mixing time
General setup

Let $G = (V, E)$ be a

- strongly connected weighted directed graph
General setup

Let $G = (V, E)$ be a

- strongly connected weighted directed graph (i.e. there is a path from every vertex to every vertex)
- with a self-loop on every vertex (to avoid periodicities)

The weights correspond to probabilities:

- $p_{ij}$ of edge $(ij) \in E$ represents probability of transitioning from $i$ to $j$
- for all $i \in V$: $\sum_{j \in V} p_{ij} = 1$

Transition matrix: $A = (p_{ij})$

$$A = \begin{pmatrix}
0.9 & 0.075 & 0.025 \\
0.15 & 0.8 & 0.05 \\
0.25 & 0.25 & 0.5
\end{pmatrix}$$
General setup

Transition matrix: $A = (p_{ij})$
Still true that $x_t = x_0 A^t$.

Claim

A has eigenvalue 1 (with multiplicity 1), and there is a unique positive vector $x$ such that $xA = x$ and $\sum_i x_i = 1$. 
General setup

Transition matrix: \( A = (p_{ij}) \)
Still true that \( x_t = x_0 A^t \).

Claim

A has eigenvalue 1 (with multiplicity 1), and there is a unique positive vector \( x \) such that \( xA = x \) and \( \sum_i x_i = 1 \).

Theorem

For any \( x_0 \), \( x_t \to x_\infty \) as \( t \to \infty \)
\( (x_\infty \) is called the stationary distribution)
General setup

Transition matrix: \( A = (p_{ij}) \)
Still true that \( x_t = x_0A^t \).

Claim

A has eigenvalue 1 (with multiplicity 1), and there is a unique positive vector \( x \) such that \( xA = x \) and \( \sum_i x_i = 1 \).

Theorem

For any \( x_0 \), \( x_t \to x_\infty \) as \( t \to \infty \)
\( (x_\infty \) is called the stationary distribution)

Q1 Why is \( x_\infty \) interesting?
Q2 How fast does \( x_t \to x_\infty \)?
Random walks on undirected graphs
Markov chains
Examples:
  pagerank
  card shuffling
Mixing time
PageRank

No better proof that something is useful than having cool applications!

$G =$ hyperlink graph

Pagerank of webpage $w$

$\approx$

probability that web-surfer starts at some central page (Yahoo!) and arrives at $w$ by clicking random links
PageRank

No better proof that something is useful than having cool applications!

$G=$ hyperlink graph

Pagerank of webpage $w$

$\approx$

probability that web-surfer starts at some central page (Yahoo!) and arrives at $w$ by clicking random links

Q: in how many steps?
PageRank

Choose

- initial distribution $x_0$ (e.g. Yahoo!)
- reset probability $\varepsilon \in [0, 1]$ (e.g. $\varepsilon = 0.15$)

Random surfer at each step

- Takes a random outgoing edge with probability $1 - \varepsilon$
- Jumps to a random node chosen according to $x_0$ o.w.
PageRank

Choose
  - initial distribution $x_0$ (e.g. Yahoo!)
  - reset probability $\varepsilon \in [0, 1]$ (e.g. $\varepsilon = 0.15$)

Random surfer at each step
  - Takes a random outgoing edge with probability $1 - \varepsilon$
  - Jumps to a random node chosen according to $x_0$ o.w.

Thus, stationary distribution satisfies

$$x(1 - \varepsilon)A + \varepsilon x_0 = x$$
PageRank

Choose

- initial distribution $x_0$ (e.g. Yahoo!)
- reset probability $\varepsilon \in [0, 1]$ (e.g. $\varepsilon = 0.15$)

Random surfer at each step

- Takes a random outgoing edge with probability $1 - \varepsilon$
- Jumps to a random node chosen according to $x_0$ o.w.

Thus, stationary distribution satisfies

$$xA = x$$
Choose
- initial distribution $x_0$ (e.g. Yahoo!)
- reset probability $\epsilon \in [0,1]$ (e.g. $\epsilon = 0.15$)

Random surfer at each step
- Takes a random outgoing edge with probability $1 - \epsilon$
- Jumps to a random node chosen according to $x_0$ o.w.

Thus, stationary distribution satisfies

$$x(1 - \epsilon)A + \epsilon x_0 = x$$
PageRank

Choose

- initial distribution $x_0$ (e.g. Yahoo!)
- reset probability $\epsilon \in [0, 1]$ (e.g. $\epsilon = 0.15$)

Random surfer at each step

- Takes a random outgoing edge with probability $1 - \epsilon$
- Jumps to a random node chosen according to $x_0$ o.w.

Thus, stationary distribution satisfies

$$x(1 - \epsilon)A + \epsilon x_0 = x$$

Solve for $x$?

$$x = \epsilon x_0 (I - (1 - \epsilon)A)^{-1}$$
Computing PageRank

Idea 1: use

$$x = \varepsilon x_0 (I - (1 - \varepsilon)A)^{-1}$$

Crawl the web, create giant $A$, invert giant $A$. An $n \times n$ matrix for $n \approx 1.9$ billion...

$\Theta(n^2)$ space, $\Theta(n^3)$ for Gaussian elimination...
Idea 1: use

$$x = \varepsilon x_0 (I - (1 - \varepsilon)A)^{-1}$$

Crawl the web, create giant $A$, invert giant $A$. An $n \times n$ matrix for $n \approx 1.9$ billion...

$\Theta(n^2)$ space, $\Theta(n^3)$ for Gaussian elimination...

Idea 2: Note

$$(I - (1 - \varepsilon)A)^{-1} = I + (1 - \varepsilon)A + (1 - \varepsilon)^2 A^2 + \ldots,$$

so

$$x = \varepsilon x_0 + \varepsilon (1 - \varepsilon)x_0 A + \varepsilon (1 - \varepsilon)^2 x_0 A^2 + \ldots,$$
Computing PageRank

**Idea 1:** use

\[ x = \varepsilon x_0 (I - (1 - \varepsilon)A)^{-1} \]

Crawl the web, create giant \( A \), invert giant \( A \). An \( n \times n \) matrix for \( n \approx 1.9 \text{billion} \)... 

\( \Theta(n^2) \) space, \( \Theta(n^3) \) for Gaussian elimination...

**Idea 2:** Note

\[ (I - (1 - \varepsilon)A)^{-1} = I + (1 - \varepsilon)A + (1 - \varepsilon)^2 A^2 + \ldots, \]

so

\[ x = \varepsilon x_0 + \varepsilon (1 - \varepsilon) x_0 A + \varepsilon (1 - \varepsilon)^2 x_0 A^2 + \ldots, \]
Computing PageRank

\[ x = \varepsilon x_0 + \varepsilon (1 - \varepsilon) x_0 A + \varepsilon (1 - \varepsilon)^2 x_0 A^2 + \ldots, \]

Let \( L \) be a geometric random variable with parameter \( \varepsilon \):

\[ \Pr[L = k] = \varepsilon (1 - \varepsilon)^k \quad \text{for all} \quad k \geq 0. \]

Then

\[ x = \mathbf{E}[x_0 A^L] \]
Computing PageRank

\[ x = \varepsilon x_0 + \varepsilon (1 - \varepsilon) x_0 A + \varepsilon (1 - \varepsilon)^2 x_0 A^2 + \ldots, \]

Let \( L \) be a geometric random variable with parameter \( \varepsilon \):

\[ \Pr[L = k] = \varepsilon (1 - \varepsilon)^k \quad \text{for all} \quad k \geq 0. \]

Then

\[ x = \mathbb{E}[x_0 A^L] \]

Q: How to approximate \( x \)?

- Sample length \( L \) from geometric distribution
- Run random walk of length \( L \), record where it ended up
- Repeat sufficient number of times
Computing PageRank

\[ x = \varepsilon x_0 + \varepsilon (1 - \varepsilon) x_0 A + \varepsilon (1 - \varepsilon)^2 x_0 A^2 + \ldots, \]

Let \( L \) be a geometric random variable with parameter \( \varepsilon \):

\[ \Pr[L = k] = \varepsilon (1 - \varepsilon)^k \quad \text{for all} \quad k \geq 0. \]

Then

\[ x = \mathbb{E}[x_0 A^L] \]

Q: How to approximate \( x \)?

- Sample length \( L \) from geometric distribution
- Run random walk of length \( L \), record where it ended up
- Repeat sufficient number of times

Definition of ‘sufficient’ depends on quality of approximation.

Can approximate top few elements of \( x \) efficiently (what is needed for ranking).
Random walks on undirected graphs
Markov chains
Examples:
  - pagerank
  - card shuffling
Mixing time
Card shuffling

**Q1:** Why shuffle a deck?
Q1: Why shuffle a deck?

A1: well, to start from a uniformly random permutation of the cards
**Card shuffling**

**Q1:** Why shuffle a deck?

**A1:** well, to start from a uniformly random permutation of the cards

**Q2:** How many permutations are there?

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

**Examples of shuffles:**

- top-in-at-random
- riffle-shuffle
Card shuffling

Q1: Why shuffle a deck?

A1: well, to start from a uniformly random permutation of the cards

Q2: How many permutations are there?

A2: $52! \approx 2^{257} \approx 10^{77}$ – how large is that?

Q3: Getting a random permutation?
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.
Random walks on undirected graphs

Markov chains

Examples:
- pagerank
- card shuffling

Mixing time
Mixing time

Q1: How many steps suffice for uniformity?

A1: Meaningless question: no matter how long you go on for, there is a trace of the permutation you started with.

Q2: How many steps suffice to be close to uniform?

A2: Depends on the shuffle

- Top-In-At-Random: \( \approx 300 \) repetitions suffice, i.e. distance \((x^{300}, \text{uniform})\) < 1%

- Riffle-Shuffle: \( \approx 10 \) repetitions suffice: distance \((x^{10}, \text{uniform})\) < 1%

Different shuffles have different graphs, and results converge to uniform at different speeds.
Mixing time

**Q1:** How many steps suffice for uniformity?

**A1:** Meaningless question: no matter how long you go on for, there is a trace of the permutation you started with.

**Q2:** How many steps suffice to be close to uniform?

**A2:** Depends on the shuffle

- **Top-In-At-Random:** \( \approx 300 \) repetitions suffice, i.e. distance \( (x^{300}, \text{uniform}) < 1\% \)

- **Riffle-Shuffle:** \( \approx 10 \) repetitions suffice: distance \( (x^{10}, \text{uniform}) < 1\% \)

Different shuffles have different graphs, and results converge to uniform at different speeds.
Mixing time

Q1: How many steps suffice for uniformity?

A1: Meaningless question: no matter how long you go on for, there is a trace of the permutation you started with.

Q2: How many steps suffice to be close to uniform?
Mixing time

**Q1:** How many steps suffice for uniformity?

**A1:** Meaningless question: no matter how long you go on for, there is a trace of the permutation you started with.

**Q2:** How many steps suffice to be close to uniform?

**A2:** Depends on the shuffle
- Top-In-At-Random: \( \approx 300 \) repetitions suffice, i.e.
  \[
  \text{distance}(x_{300}, \text{uniform}) < 1\%
  \]
- Riffle-Shuffle: \( \approx 10 \) repetitions suffice:
  \[
  \text{distance}(x_{10}, \text{uniform}) < 1\%
  \]

Different shuffles have different graphs, and results converge to uniform at different speeds.
Mixing time (formally)

Let $x_0, x_1, x_2, \ldots$ be a Markov chain with stationary distribution $x_\infty$.

**Definition**

The *mixing time* of the Markov chain is the minimum $\tau$ such that $d(x_\tau, x_\infty) \leq 0.01$. Where $d(x, y)$ captures the rate at which $x_t \to x_\infty$. When we design Markov chains, we would like $\tau_{\text{mix}}$ to be small. $\tau_{\text{mix}}$ depends on the connectivity of the Markov chain.
Mixing time (formally)

Let $x_0, x_1, x_2, \ldots$ be a Markov chain with stationary distribution $x_\infty$.

**Definition**
The *mixing time* of the Markov chain is the minimum $\tau$ such that $d(x_\tau, x_\infty) \leq 0.01$.

For two distributions $x, y$: $d(x, y) = \sum_i |x_i - y_i|$ (≈ total variation distance)

Captures rate at which $x_t \rightarrow x_\infty$. 
Mixing time (formally)

Let $x_0, x_1, x_2, \ldots$ be a Markov chain with stationary distribution $x_\infty$

**Definition**
The *mixing time* of the Markov chain is the minimum $\tau$ such that $d(x_\tau, x_\infty) \leq 0.01$.

For two distributions $x, y$: $d(x, y) = \sum_i |x_i - y_i|$ (≈ total variation distance)

Captures rate at which $x_t \to x_\infty$.

When we design Markov chains, we would like $\tau_{\text{mix}}$ to be small. $\tau_{\text{mix}}$ depends on the connectivity of the Markov chain.
Summary

- random walks on undirected graphs
- Markov chains
- Mixing time: how fast Markov chains converge to stationary distribution