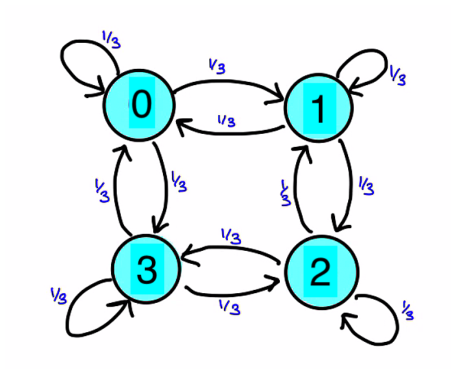
CS265, Winter 2026

Class 13: Agenda and Questions

1 Warm-Up

Consider the Markov chain given by:



Group Work

1. What is the transition matrix for this Markov chain?
2. Suppose that you start in state 0. What is the probability that you are in state 2 after one step? Two steps? Three steps? 100 steps? (Don't actually compute this, just say how you would).
3. As $t \to \infty$, what do you think is $\lim_{t\to\infty} \Pr[X_t = 2 | X_0 = 0]$? (No need for a formal proof here, just use your intuition).

Group Work: Solutions

1. $P = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 1/3 & 1/3 \end{pmatrix}$ is the transition matrix.
2. After 1 step, the probability is 0. After 2 steps, it's 2/9. After 3 steps, it's 6/27. To figure it out after 100 steps, we should compute $e_0 \cdot P^{100} \cdot e_2$. I don't want to do this by hand!
3. It should be 1/4. Intuitively, if we walk for long enough there's no reason we should prefer being at one state over any other. (Next week we'll see that this intuition

can be formalized as "the stationary distribution is uniform.")

2 Questions/Lecture Recap

Any questions from the minilectures and/or the quiz? (Markov chains and a randomized algorithm for 2SAT)

3 Spectral Analysis of Markov Chains

Next, we'll see how we can use linear algebra to help us out in computing things like $\Pr[X_t = 2|X_0 = 0]$ for general t. We'll focus on example in the warm-up, but as we go, keep in mind what you think the general principle should be.

Group Work

1. Let
$$F = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix}$$
where $i = \sqrt{-1}$. (You may recognize F as the 4×4 discrete Fourier matrix, so $F_{jk} = \frac{1}{2}e^{-2\pi ijk/4}$.) Notice that F is a Hermitian matrix, which means that $F^*F = FF^* = I$, where F^* denotes the Hermitian conjugate (e.g., take the transpose and change all of the i's to $-i$'s).

   Convince yourself that
$$P = F \cdot \begin{pmatrix} 1 & & & \\ & 1/3 & & \\ & & -1/3 & \\ & & & 1/3 \end{pmatrix} \cdot F^*,$$
where P is the transition matrix from the warm-up.

   Hint: *Check that the columns of F are eigenvectors for P. What are their eigenvalues?*

   Note: *If your linear algebra is rusty and you trust me, just remind yourself what an eigenvector actually is. The main point here is that you should understand the conclusion of Question 1 so that you can use it in Question 2.*

2. In the warm-up, you came up with an expression for $\Pr[X_t = 2|X_0 = 0]$, which seemed pretty obnoxious to compute. Use the previous part to find a closed-form expression for this. Recall that our intuition was that $\Pr[X_t = 2|X_0 = 0] \to \frac{1}{4}$ as

$t \to \infty$. Quantify this by coming up with a statement like

$$\Pr[X_t = 2 | X_0 = 0] = \frac{1}{4} \pm O(\_\_\_)$$

where the thing in the $O(\_\_\_)$ term should depend on t. What is the best bound you can get?

Group Work: Solutions

1. Following the hint, let's check that the columns of F are eigenvalues for P. We have:
   - $P \cdot (1,1,1,1)^T = 1 \cdot (1,1,1,1)^T$, so the first column is an eigenvector with eigenvalue 1.
   - $P \cdot (1, -i, -1, i)^T = \begin{pmatrix} 1/3 - i/3 + i/3 \\ 1/3 - i/3 + -1/3 \\ -i/3 - 1/3 + i/3 \\ 1/3 - 1/3 + i/3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 \\ -i \\ -1 \\ i \end{pmatrix}$ so the second column is an eigenvector with eigenvalue $1/3$.
   - Similarly we can check that the third column has eigenvalue $-1/3$ and the fourth column has eigenvector $1/3$.

   If we remember our linear algebra, this is enough to conclude that what's written is the eigendecomposition for P.

   If we don't remember our linear algebra, here's one way we could conclude that. (Basically we'll just re-derive why we care about the eigendecomposition). Let $\Lambda = \mathrm{diag}(1, 1/3, -1/3, 1/3)$ be the diagonal matrix in the middle. We want to show that $P = F\Lambda F^*$. It's enough to show that $Pv = F\Lambda F^* v$ for all vectors v; and in particular it's enough to show it for four linearly independent vectors v. Let's choose the vectors v to be the columns of F; say v_i is the i'th column. Since the v_i are orthogonal, we have $F^* v_i = e_i$. Thus, $\Lambda F^* v_i = \lambda_i v_i$, where λ_i is the i'th entry on the diagonal of Λ. Then, $F\Lambda F^* v_i = \lambda_i F e_i = \lambda_i v_i$. But we just established above that $Pv_i = \lambda_i v_i$ also for all i. So the two matrices are the same.

2. Now that we know that $P = F\Lambda F^*$, we can write

   $$e_0^T P^{100} e_2 = e_0^T (F\Lambda F^*)^{100} e_2 = e_0^T F \Lambda^{100} F^* e_2.$$

   Fortunately, Λ^{100} is really easy to compute! Just raise everything on the diagonal to the power of 100. More generally, after t steps, we can write:

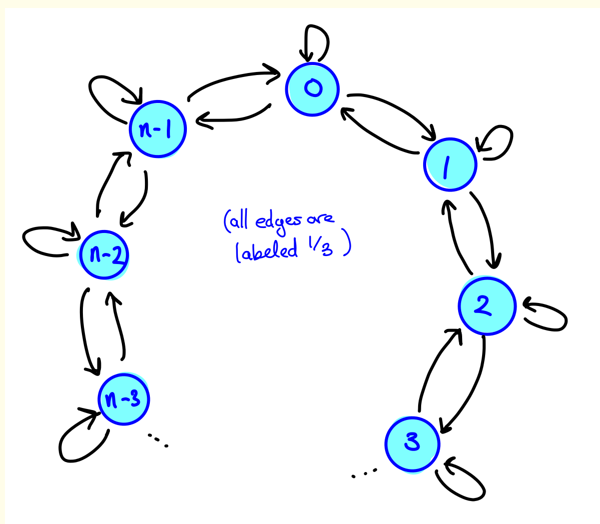
$$e_0^T F \Lambda^t F^* e_2 = \frac{1}{4}(1,1,1,1)\begin{pmatrix} 1 & & & \\ & 3^{-t} & & \\ & & (-3)^{-t} & \\ & & & 3^{-t} \end{pmatrix}\begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = \frac{1}{4}(1 + O(3^{-t})).$$

Before we move on to larger cycles, let's take a minute to reflect on what just went on. *[Insert a bit of lecture about spectral analysis. The point is that if we have a symmetric Markov chain, we can always write the transition matrix as $P = V\Lambda V^*$ for a Hermitian matrix V and a diagonal matrix Λ with real values on the diagonals. Then we can write $P^t = V\Lambda^t V^*$, and as long as the second-largest eigenvalue is strictly less than 1, eventually Λ^t will look like $\mathrm{diag}(1, tiny, tiny, \ldots, tiny)$. This means that we can compute transition probabilities after t steps up to very small error terms.]*

In this next part, you'll generalize what you saw above to larger cycles.

Group Work

1. Consider the analogous Markov chain to the 4-state one that you saw before, except that it has n states. That is, it looks like this:



Let $P \in \mathbb{R}^{n \times n}$ be the transition matrix for this Markov chain. We will see together in class that:

$$P = F_n \Lambda F_n^*,$$

where Λ is a diagonal matrix whose j'th entry is

$$\Lambda_{j,j} = \frac{1 + 2\cos(2\pi j/n)}{3},$$

where $j = 0, \ldots, n-1$. (Importantly, j is zero-indexed here!) Above, F_n is the $n \times n$ DFT, so

$$(F_n)_{j,k} = \frac{1}{\sqrt{n}} e^{-2\pi ijk/n}.$$

(There is no question here, we'll do this bit together.)

2. Come up with an expression for $\Pr[X_t = 0 | X_0 = 0]$. You should get a kind of nasty sum involving some cosines, that's okay.
3. Convince yourself that as $t \to \infty$, $\Pr[X_t = 0 | X_0 = 0] \to 1/n$.

...another part on next page

4. **Bonus, if time:** Try to think about how *fast* this convergence is. That is, how large does t have to be before $\Pr[X_t = 0|X_0 = 0] = \frac{1+o(1)}{n}$? (Don't try to come up with a formal proof, just some back-of-the-envelope calculations).

   Also, how does this compare to what we saw in the mini-lectures about the walk on the line?

   ***Hint**: You may find the Taylor expansion* $\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} - \cdots$ *of* $\cos(x)$ *about zero helpful. In particular, when x is small,* $\cos(x) \approx 1 - \frac{x^2}{2}$. *You may also want to use the approximation* $1 - x \approx e^{-x}$ *for small x liberally.*

Group Work: Solutions

1. See posted slides
2. This probability is $\frac{1}{n}\vec{1}^T \Lambda^t \vec{1}$, since $\vec{1}/\sqrt{n} = e_0 F_n$ and $\Pr[X_t = 0|X_0 = 0] = e_0^T P^t e_0 = e_0^T F_n \Lambda^t F_n^* e_0$. Writing that out, it's

$$\frac{1}{n} \sum_{j=0}^{n-1} \Lambda_{j,j}^t = \frac{1}{n} \sum_{j=0}^{n-1} \left(\frac{1 + 2\cos(2\pi j/n)}{3} \right)^t.$$

3. When $j = 0$, $\Lambda_{0,0} = (1 + 2\cos(0))/3 = 1$, so $\Lambda_{0,0}^t = 1^t = 1$ for any t. On the other hand, when $j > 0$, $\Lambda_{j,j} = (1 + 2\cos(2\pi j/n))/3 < 1$, which means that as $t \to \infty$, $\Lambda_{j,j}^t \to 0$. So only the $j = 0$ term survives in the expression from the previous part, and we get

$$\frac{1}{n}(1 + \text{stuff that} \to 0) \to \frac{1}{n}.$$

4. Based on our intuition with the line, maybe we'd think it should be about n^2 steps. That's because we think it will take us about n^2 steps to get all the way around

the circle; and once we get all the way around the circle, surely we're close to the limiting (uniform) distribution. This is pretty hand-wavey, but here's the way to make it rigorous.

Let's think about what happens to $\Lambda_{j,j}^t$ as t gets large. Following the hint, for small j (say, $j < n/10$), we have:

$$\begin{aligned}
\Lambda_{j,j}^t &= \left(\frac{1 + 2\cos(2\pi j/n)}{3}\right)^t \\
&\approx \left(1 + 2(1 - (2\pi j/n)^2/2)\right)^t \\
&= \left(1 - \frac{(2\pi j/n)^2}{3}\right)^t \\
&\approx \exp\left(\frac{-t(2\pi j)^2}{3n^2}\right) \\
&= \exp(-\Omega(tj^2/n^2)).
\end{aligned}$$

For $j \in [n/10, n/2 - n/10]$ (say), we have $|\cos(2\pi j/n)| \le \cos(2\pi/10) \approx 0.8$, so

$$\left(\frac{1 + 2\cos(2\pi j/n)}{3}\right)^t \lesssim (1 + 1.6)/3 \approx 0.86^t = \exp(-\Omega(t)).$$

(The only point here is that 0.86 is some constant that's less than 1 and doesn't depend on n). So no matter what j is, we have

$$|\Lambda_{j,j}|^t \le \exp(-Ctj^2/n^2)$$

for some small enough constant C. Let's set $t = C'n^2$ for some large enough constant C', and return to our nasty expression. We have

$$\begin{aligned}
\frac{1}{n}\sum_{j=0}^{n-1} \Lambda_{j,j}^t &= \frac{1}{n} \cdot 1 + \frac{1}{n}\sum_{j=1}^{n-1} \Lambda_{j,j}^t \\
&= \frac{1}{n} \pm \frac{1}{n}\sum_{j=1}^{n} \exp(-C \cdot C' \cdot j^2) \\
&= \frac{1}{n}(1 \pm 0.01)
\end{aligned}$$

where we are assuming that we've picked our constant C' so that it's large enough that the sum $\sum_{j=1}^n \exp(-CC'j^2) \le 0.01$. (If $C \cdot C' \ge 10$, this should be fine, since $\sum_{j=1}^\infty \exp(-10j^2)$ is much smaller than 0.01).

This was pretty fast and loose with the $\approx$'s, but it gives some intuition for why we should be able to take $t \approx n^2$.

At the end of the day, we conclude that after about n^2 steps (up to some leading constant), there's a very close to $1/n$ chance that we're back at 0. (The same analysis works to show that there's also very close to a $1/n$ chance that we're at, say, $n/2$). One way to say this is that after about n^2 steps, X_t is basically uniform on the cycle.

Next week we'll see another method, called *coupling,* which can also help us bound how fast a Markov chain "mixes." (Aka, gets close to the distribution it should be getting close to).