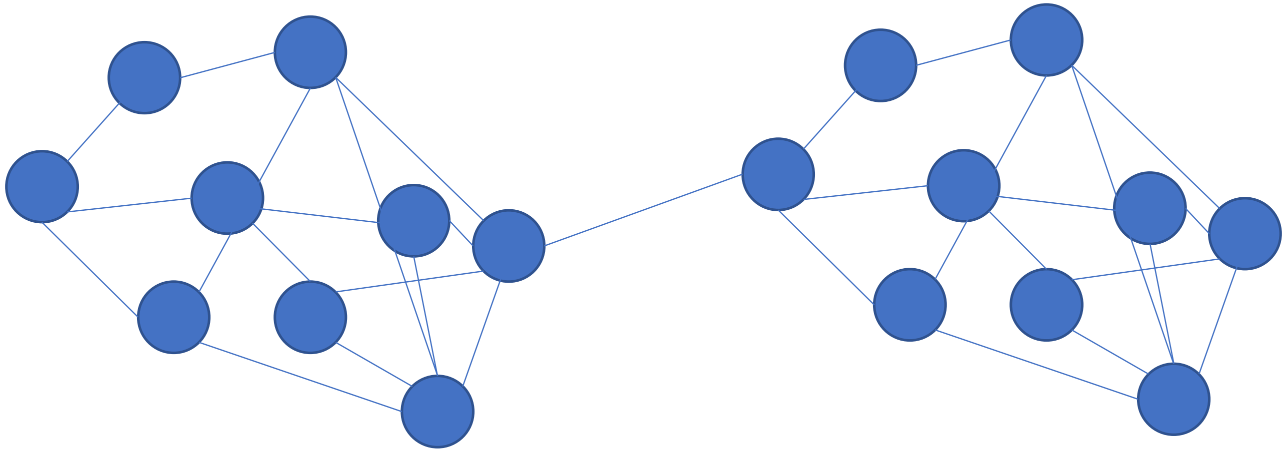
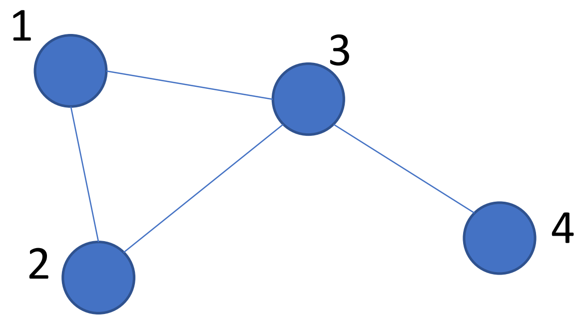
Markov Chains and Mixing Times

VRDI 2019

Sarah Cannon



Example: Random Walk



Can express current state as probability vector:

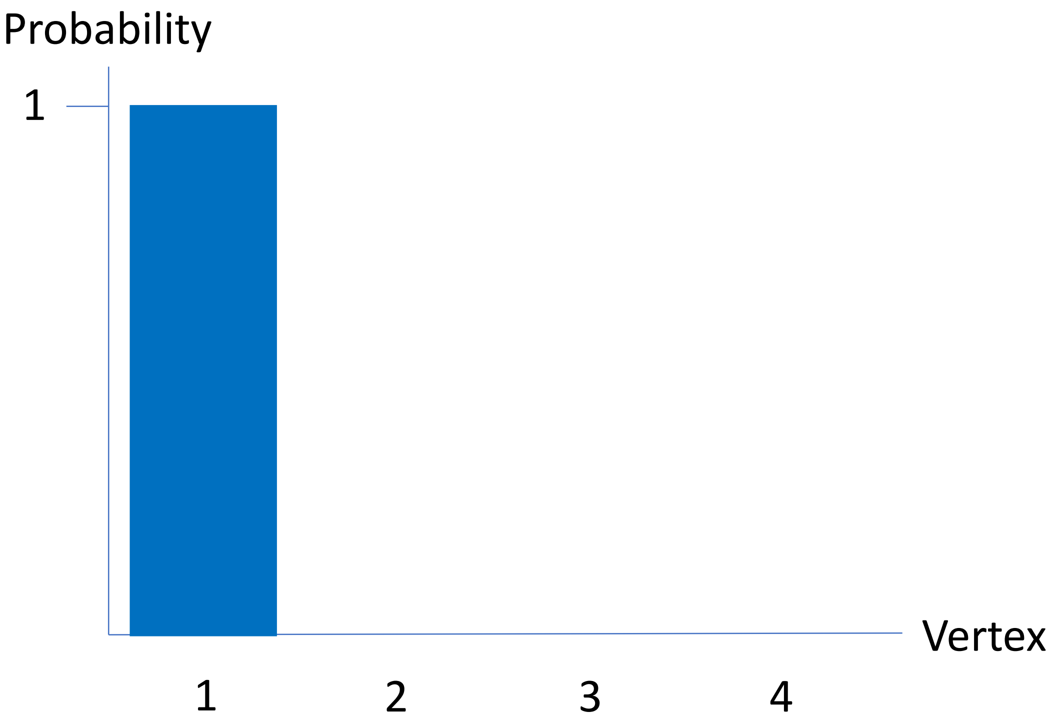
V_0 = [1,	0,	0,	0]
V_1 = [0,	1/2,	1/2,	0]
V_2 = [5/12,	1/6,	1/4,	1/6]
….			
π = [1/4,	1/4,	3/8,	1/8]

How long until your current probability vector is close to π?

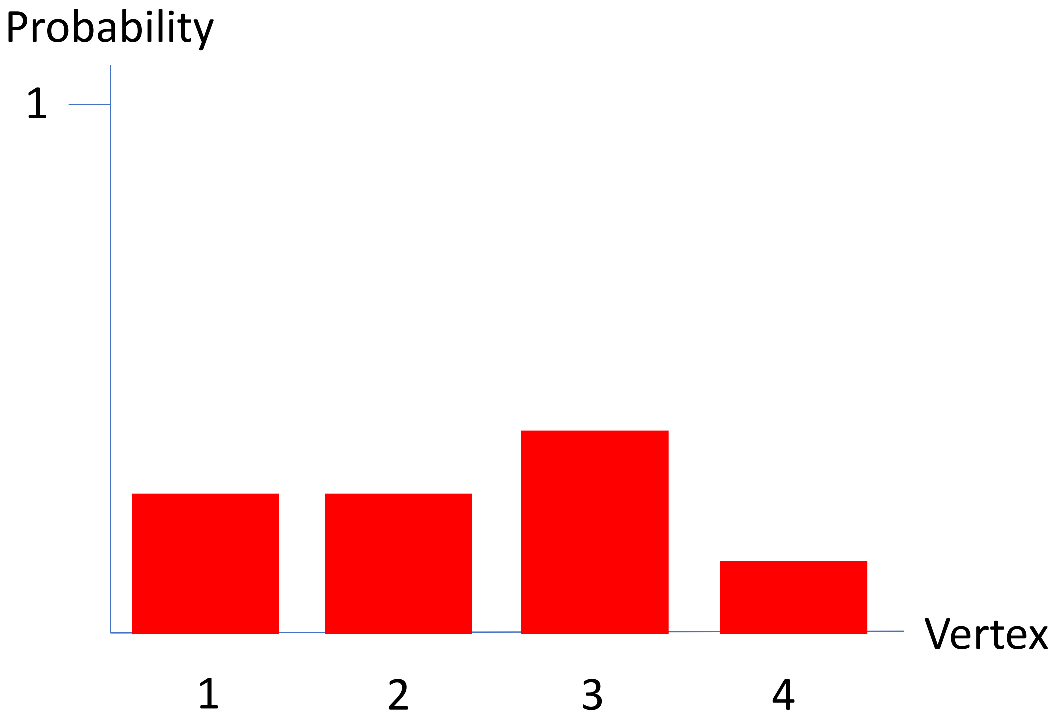
How can you measure how close two probability distributions are?

Total variation distance

- One way to measure how close two probability distributions are



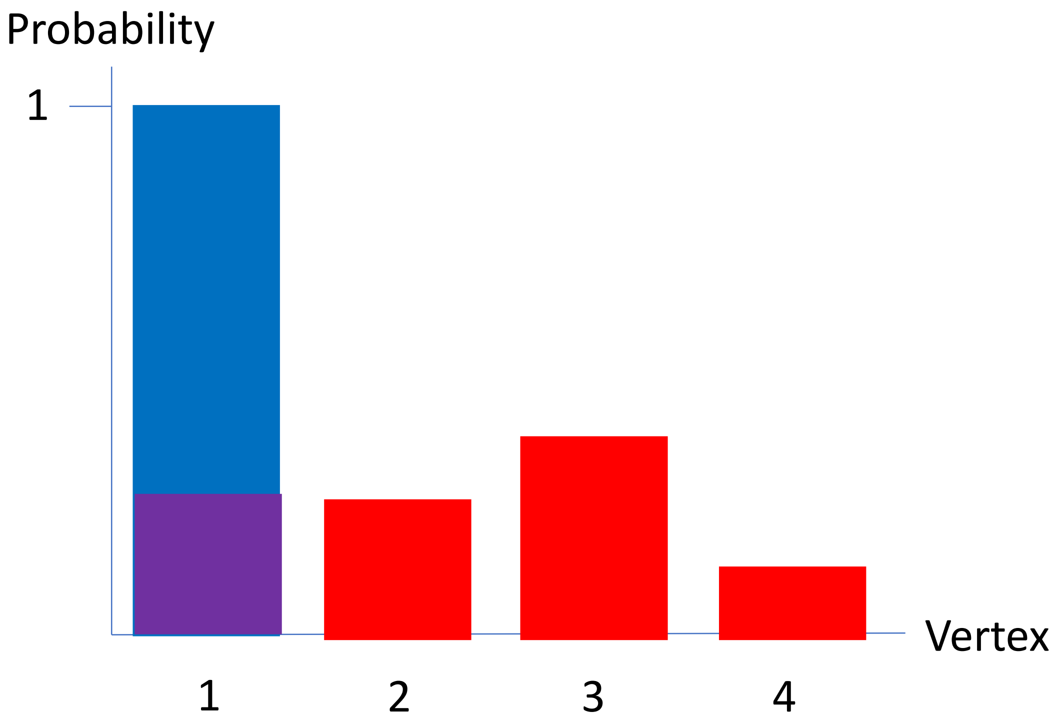
Starting distribution V_0 = [1, 0, 0, 0]



Stationary distribution π = [1/4, 1/4, 3/8, 1/8]

Total variation distance

- One way to measure how close two probability distributions are



Starting distribution V_0 = [1, 0, 0, 0]

Stationary distribution π = [1/4, 1/4, 3/8, 1/8]

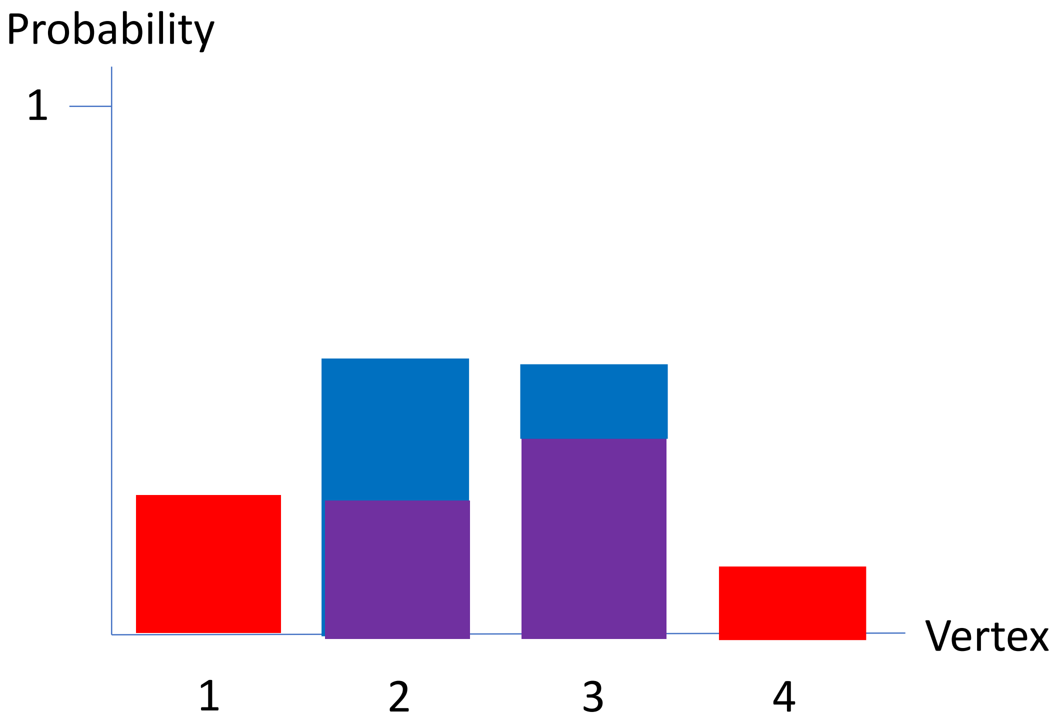
TVD(V_0, π) = half of non-purple area

$$\text{TVD}(V_0, \pi) = \frac{1}{2} \sum_i |V_0(i) - \pi(i)|$$

TVD(V_0, π) = 3/4

Total variation distance

- One way to measure how close two probability distributions are



Distribution after one step V_1 = [0, 1/2, 1/2, 0]

Stationary distribution π = [1/4, 1/4, 3/8, 1/8]

$TVD(V_0, \pi)$ = half of non-purple area

$TVD(V_0, \pi) = \frac{1}{2} \sum_i |V_0(i) - \pi(i)|$

$TVD(V_0, \pi) = 3/4$

$TVD(V_1, \pi) = 3/8$

Under mild conditions, TVD to π is always decreasing

$TVD(V_2, \pi) = 5/24$

Mixing time

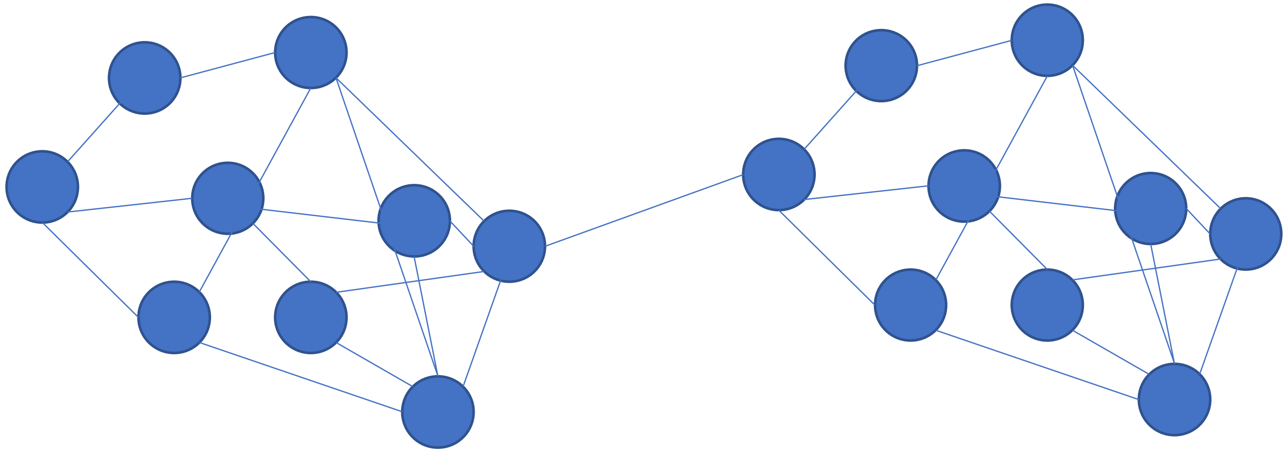
- The *mixing time* is the smallest t such that $TVD(V_t, \pi) < ¼$

Could use any constant less than ½, doesn't affect mixing time very much (once TVD is less than a half, it decreases to zero very quickly)

Mixing time

- The *mixing time* is the smallest t such that $TVD(V_t, \pi) < 1/4$
- Why should we care about mixing time?

Metagraph:



Districting plans favoring clubs

Districting plans favoring hearts

If don't run Markov chain long enough, only see plans favoring one suit

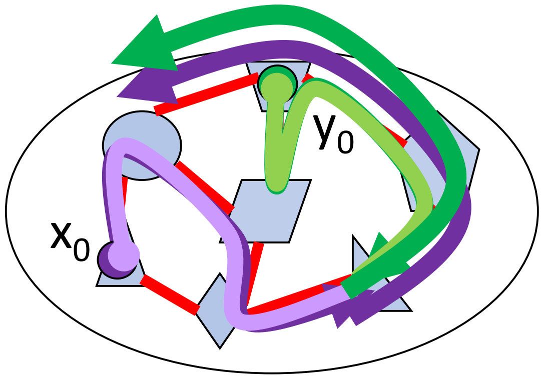
Mixing time

- The *mixing time* is the smallest t such that $TVD(V_t, \pi) < ¼$
- Why should we care about mixing time?
- Finding upper bounds on the mixing time is **HARD**
  - Some techniques known, but rely on metagraph having nice structure

Technique 1: Coupling

Simulate **2** processes:

- Start at any x_0 and y_0
- Couple moves, but each simulates the MC
- Once they agree, they move in sync ($x_t = y_t \rightarrow x_{t+1} = y_{t+1}$)



Expected Coupling Time > Mixing time

Prove chains getting **closer** in expectation in each step

Example: Random to Top Shuffle

- Shuffle a deck of n cards by picking a random card and putting it on top
- How many times do you have to do this until the cards are shuffled?

Coupling Proof:

- Take two decks in arbitrary (different) orders
- At each step, pick the same card in both (e.g 2 of clubs) and put it on top
- After you've picked each card once, the decks are the same
    - Known from probability: expected time it takes to pick each of the n cards once is about n log(n)

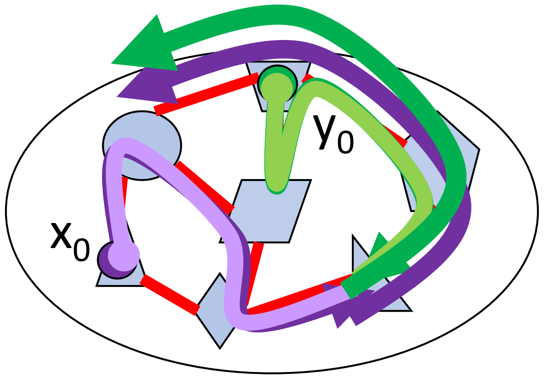
Mixing time < expected coupling time

= expected time to pick each card once = n log(n)

Technique 1: Coupling

Simulate **2** processes:

- Start at any x_0 and y_0
- Couple moves, but each simulates the MC
- Once they agree, they move in sync ($x_t = y_t \rightarrow x_{t+1} = y_{t+1}$)



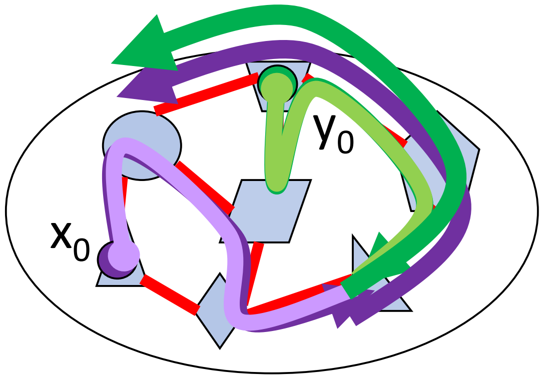
Expected Coupling Time > Mixing time

Prove chains getting **closer** in expectation in each step

Technique 2: Path Coupling

Simulate **2** processes:

- Start at any x_0 and y_0
- Couple moves, but each simulates the MC
- Once they agree, they move in sync ($x_t = y_t \rightarrow x_{t+1} = y_{t+1}$)



Expected Coupling Time > Mixing time

Prove chains getting closer in expectation in each step, **but set up metrics so that you only need to consider starting at states that are adjacent in the metagraph** (often easier to do, but get worse bounds)

More techniques:

- **Coupling**
- **Path coupling**
- **Comparison**: show (in a precise way) your Markov chain is similar to one whose mixing time is known
- **Decomposition**: Break your metagraph into parts, show fast mixing within each part and between the parts
- **Canonical Paths**: Look at flows on the graph, use them to show there are no small cuts, which implies fast mixing

Main points about mixing times:

- All known techniques require there to be some really nice structure in the metagraph that you're unlikely to find in real-world settings
- It's really hard to tell if a chain is mixed or not: be careful with heuristics!