Markov Chains, Mixing Times, and Couplings

James Marshall Reber

Department of Mathematics, Purdue University

July 24, 2018

Motivation

Motivating Question

Preforming a random walk on some graph, how long does it take until you are “sufficiently random?”

Theorem (Diaconis, Bayer ’92)

If you riffle shuffle a deck of size n, it takes approximately $\frac{3}{2}\log_2(n)$ shuffles until the deck is “sufficiently random.”

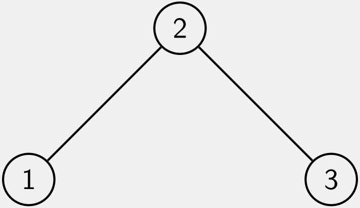
Graph Theory

Graph

We define a **graph** to be a tuple $G = (V, E)$ such that V is a collection of objects called **vertices** and $E \subseteq V \times V$ is a collection of pairs called **edges**.

Example

$V = \{1, 2, 3\}$, $E = \{(1, 2), (2, 3)\}$.



Graph Theory

Degree

We define the **degree** of a vertex to be the number of **neighbors**, or vertices which are connected by an edge, the vertex has. This is generally denoted by $\deg(x)$.

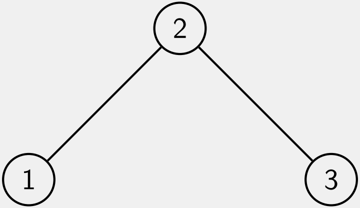
Regular

A graph is said to be n-**regular** if the degree of all the vertices is n.

Graph Theory

Example

$V = \{1, 2, 3\}$, $E = \{(1, 2), (2, 3)\}$.

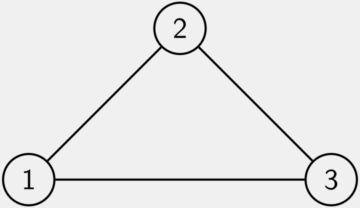


We see $\deg(2) = 2$, $\deg(1) = 1$, and $\deg(3) = 1$. This is therefore **not** regular.

Graph Theory

Example

$V = \{1, 2, 3\}$, $E = \{(1, 2), (1, 3), (2, 3)\}$.



We see $\deg(2) = 2$, $\deg(1) = 2$, and $\deg(3) = 2$. This is therefore **2-regular**.

Markov Chains

Markov Property and Markov Chain

A **Markov Chain** is a series of random variables $(X_0, X_1, \ldots)$ on a common state space Ω satisfying the **Markov Property**:

$$\mathbf{P}\{X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\} = \mathbf{P}\{X_n = x_n \mid X_{n-1} = x_{n-1}\}.$$

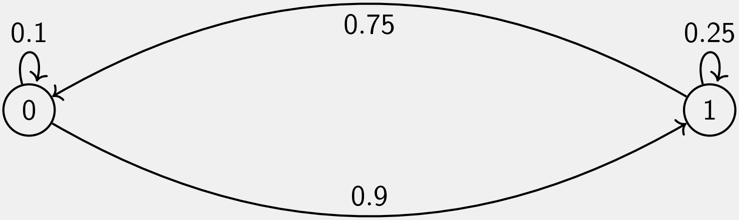
Transition Matrix

We can model Markov Chains using a **transition matrix**, which is a matrix with entries

$$P(x, y) = \mathbf{P}\{X_n = y \mid X_{n-1} = x\}$$

Markov Chains

Example



Markov Chains

Example

This Markov chain has transition matrix

$$P = \begin{matrix} & \begin{matrix} 0 & 1 \end{matrix} \\ \begin{matrix} 0 \\ 1 \end{matrix} & \begin{bmatrix} 0.1 & 0.9 \\ 0.75 & 0.25 \end{bmatrix} \end{matrix}$$

Markov Chains

Aperiodic and Irreducible

We say our Markov Chain is **irreducible** if there exists a $t > 0$ for all $x, y \in \Omega$ such that

$$P^t(x, y) > 0.$$

We say that our Markov Chain is **aperiodic** if

$$\gcd\{t \geq 1 \mid P^t(x, x) > 0\} = 1$$

for all $x \in \Omega$.

Markov Chains

Stationary Distribution

If our Markov chain is **irreducible**, then we have that there exists a unique distribution π such that

$$\pi P = \pi.$$

We call such a distribution a **stationary distribution**.

Limiting Distribution

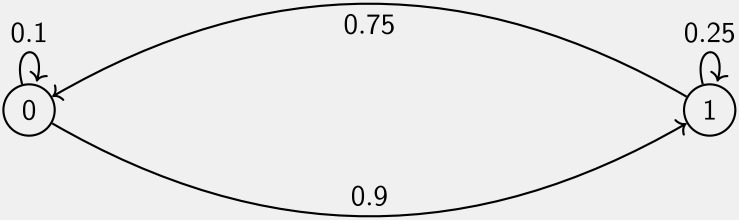
We call a distribution $\hat{\pi}$ a **limiting distribution** if

$$\lim_{t \to \infty} P^t(x, y) = \hat{\pi}(y).$$

If our Markov chain is **aperiodic** and **irreducible**, then we have that the stationary distribution π is the limiting distribution $\hat{\pi}$.

Markov Chains

Example



Markov Chains

Example

This Markov chain has transition matrix

$$P = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{bmatrix} 0.1 & 0.9 \\ 0.75 & 0.25 \end{bmatrix} \end{array}$$

Notice that this is **aperiodic** and **irreducible**, and so we have a stationary distribution. The stationary distribution is

$$\pi = \begin{bmatrix} \frac{5}{11} & \frac{6}{11} \end{bmatrix}$$

Markov Chains

Simple Random Walk

Given some graph G, we can define a **simple random walk on** G to be a Markov chain with state space V and transition matrix

$$P(x,y) = \begin{cases} \frac{1}{\deg(x)} & \text{if x and y are neighbors,} \\ 0 & \text{otherwise.} \end{cases}$$

Lazy Random Walk

Given some graph G, we can define a **lazy random walk on** G to be a Markov chain with state space V and transition matrix

$$P(x,y) = \begin{cases} \frac{1}{2} & \text{if } x = y, \\ \frac{1}{2\deg(x)} & \text{if x and y are neighbors,} \\ 0 & \text{otherwise.} \end{cases}$$

Mixing Times

Total Variation Distance

We define the **total variation distance** between two probability distributions μ and ν on a common state space Ω to be

$$||\mu - \nu||_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

In particular, we care about

$$d(t) := ||P^t(x, \cdot) - \pi(\cdot)||_{TV}$$

Mixing Time

We define the **mixing time** of a Markov chain to be

$$t_{\text{mix}}(\epsilon) := \min\{t \mid d(t) \leq \epsilon\}.$$

Motivation

Motivating Question

Preforming a random walk on some graph, how long does it take until you are "sufficiently random?"

Theorem (Diaconis, Bayer '92)

If you riffle shuffle a deck of size n, it takes approximately $\frac{3}{2}\log_2(n)$ shuffles until the deck is "sufficiently random."

Coupling

Markovian Coupling of Markov Chains

We define a **Markovian coupling of two Markov chains** (X_t) and (Y_t) with common state space Ω and transition matrix P to be the Markov chain $(X_t, Y_t)_{t=0}^{\infty}$ over $\Omega \times \Omega$, with the addendum that

$$\mathbf{P}\{X_{t+1} = x' \mid X_t = x, Y_t = y\} = P(x, x')$$

and

$$\mathbf{P}\{Y_{t+1} = y' \mid X_t = x, Y_t = y\} = P(y, y').$$

We will also require that $X_s = Y_s$ for some s implies $X_t = Y_t$ for all $t \geq s$. A coupling is not a required to be Markovian (and it may not even be the optimal coupling), but in general we want it to be.

Coupling

Theorem

Let

$$\tau := \min\{t \mid X_s = Y_s \text{ for all } s \geq t\}.$$

If (X_t) and (Y_t) evolve according to a coupling, then we have

$$d(t) \leq \max_{x,y\in\Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{TV}$$

$$\leq \max_{x,y\in\Omega} \mathbf{P}\{\tau > t \mid X_0 = x, Y_0 = y\}$$

$$\leq \max_{x,y\in\Omega} \frac{\mathbf{E}(\tau \mid X_0 = x, Y_0 = y)}{t}$$

Coupling

Theorem

Let

$$\tau := \min\{t \mid X_s = Y_s \text{ for all } s \geq t\}.$$

If (X_t) and (Y_t) evolve according to a coupling, then we have

$$d(t) \leq \max_{x,y\in\Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{TV}$$

$$\leq \max_{x,y\in\Omega} \mathbf{P}\{\tau > t \mid X_0 = x, Y_0 = y\}$$

$$\leq \max_{x,y\in\Omega} \frac{\mathbf{E}(\tau \mid X_0 = x, Y_0 = y)}{t}$$

Coupling

Theorem

Let

$$\tau := \min\{t \mid X_s = Y_s \text{ for all } s \geq t\}.$$

If (X_t) and (Y_t) evolve according to a coupling, then we have

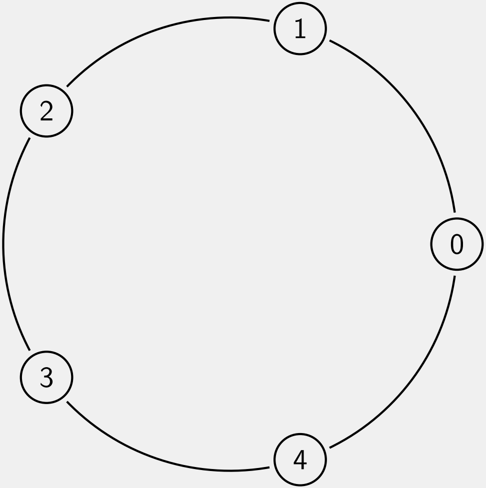
$$d(t) \leq \max_{x,y \in \Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{TV}$$

$$\leq \max_{x,y \in \Omega} \mathbf{P}\{\tau > t \mid X_0 = x, Y_0 = y\}$$

$$\leq \max_{x,y \in \Omega} \frac{\mathbf{E}(\tau \mid X_0 = x, Y_0 = y)}{t}$$

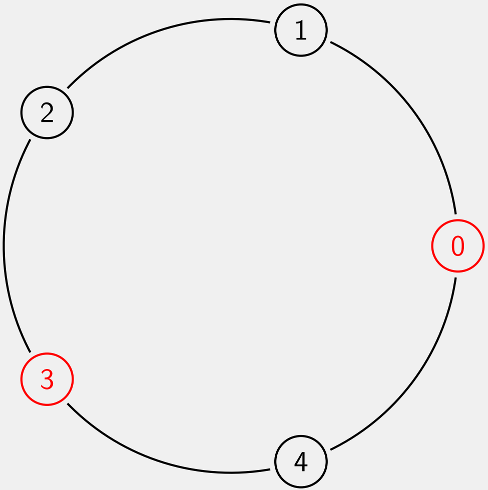
Coupling

Example



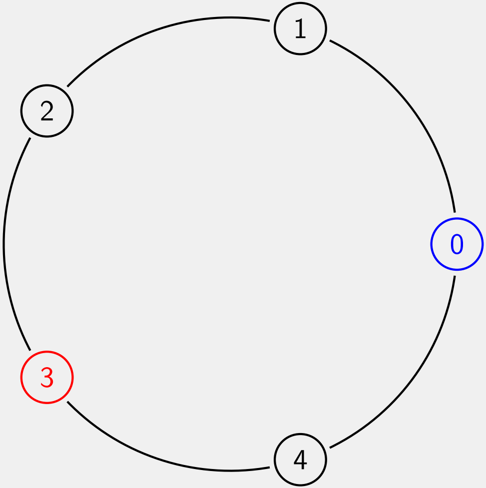
Coupling

Example



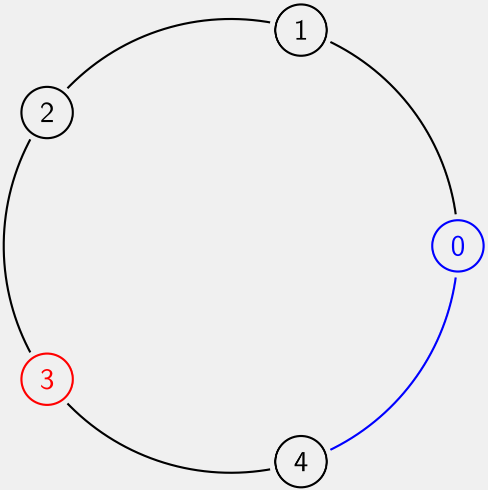
Coupling

Example



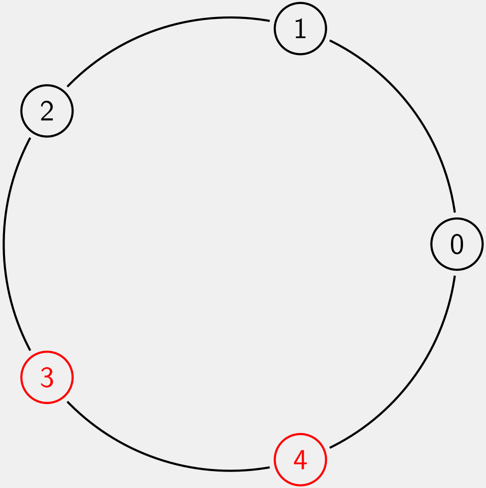
Coupling

Example



Coupling

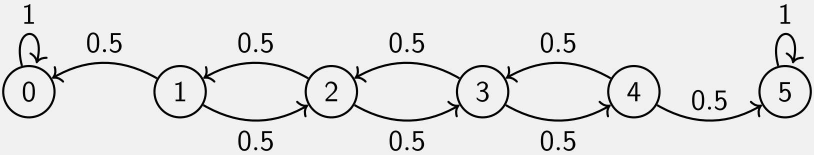
Example



Coupling

Example

If we measure the **clockwise distance** between the two walkers, this coupling gives us a new Markov chain on $\{0, 1, \ldots, n\}$ to study.

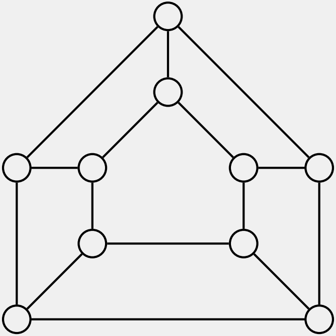


This gives us

$$d(t) \leq \frac{n^2}{4t} \rightarrow t_{\text{mix}}(\epsilon) \leq \frac{n^2}{4\epsilon}$$

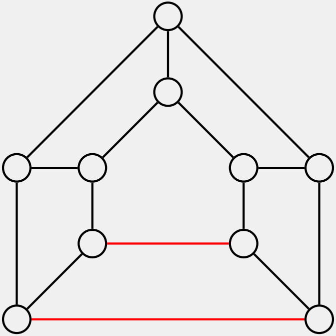
3-Regular Graphs

Prism and Möbius ladder graphs



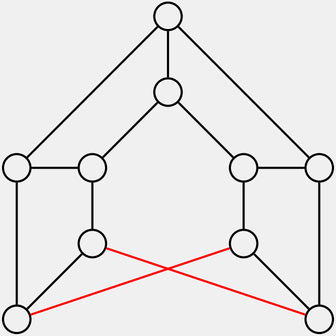
3-Regular Graphs

Prism and Möbius ladder graphs



3-Regular Graphs

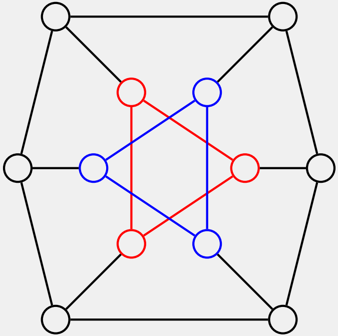
Prism and Möbius ladder graphs



3-Regular Graphs

GP(n,k)

Below is an example of $GP(6, 2)$.



Notice that if we imagined a cycle on the inside, the nodes which are the same color would be distance **2** away from eachother.

3-Regular Graphs

Results (Marshall Reber & White '18)

Using coupling, we were able to determine that for Möbius ladder graphs and prism graphs with n vertices, the mixing time for a (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3n^2}{16\epsilon} + \frac{6}{\epsilon},$$

and for the generalized Petersen graph $\text{GP}(n, k)$, the mixing time of the (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3|k|^2}{2\epsilon} + \frac{3}{2\epsilon}\left(\frac{n}{|k|}\right)^2 + \frac{15}{\epsilon},$$

where $|k| = n/\gcd(n, k)$.

3-Regular Graphs

Results (Marshall Reber & White '18)

Using coupling, we were able to determine that for Möbius ladder graphs and prism graphs with n vertices, the mixing time for a (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3n^2}{16\epsilon} + \frac{6}{\epsilon},$$

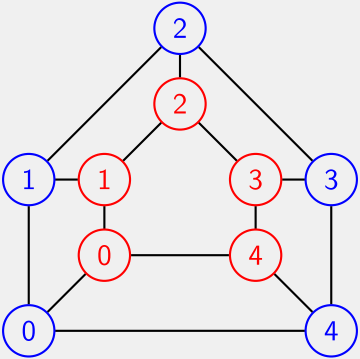
and for the generalized Petersen graph $\text{GP}(n, k)$, the mixing time of the (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3|k|^2}{2\epsilon} + \frac{3}{2\epsilon}\left(\frac{n}{|k|}\right)^2 + \frac{15}{\epsilon},$$

where $|k| = n / \gcd(n, k)$.

Sketch of Proof (Prism Graph)

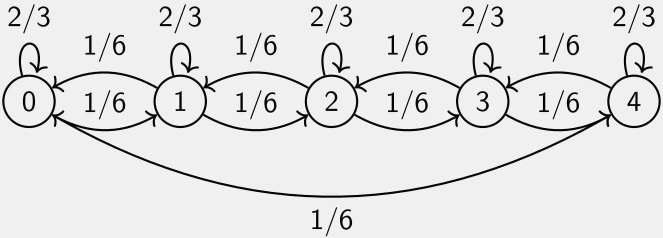
Couple based on inner cycle versus outer cycle.



3-Regular Graphs

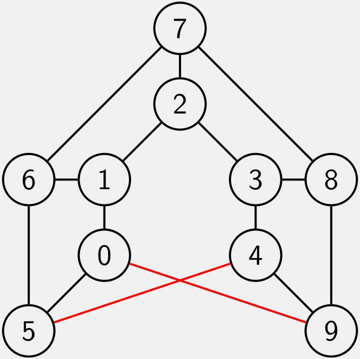
Sketch of Proof (Prism Graph)

The coupling now shifts the problem to the random walk on the cycle, with increased probability of staying in place.



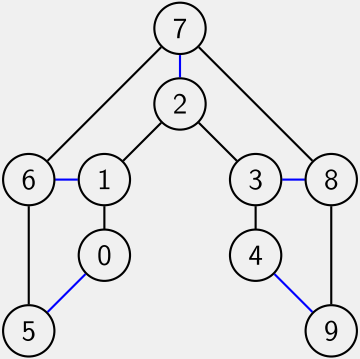
3-Regular Graphs

Sketch of Proof (Möbius Ladder Graph)



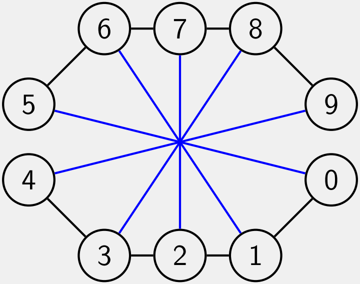
3-Regular Graphs

Sketch of Proof (Möbius Ladder Graph)



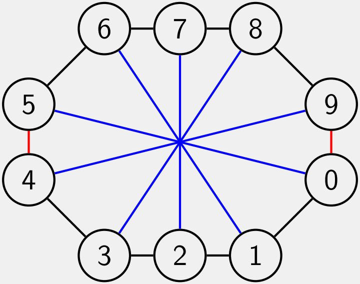
3-Regular Graphs

Sketch of Proof (Möbius Ladder Graph)



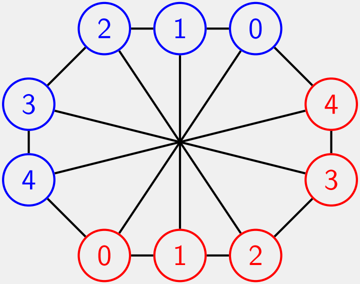
3-Regular Graphs

Sketch of Proof (Möbius Ladder Graph)



3-Regular Graphs

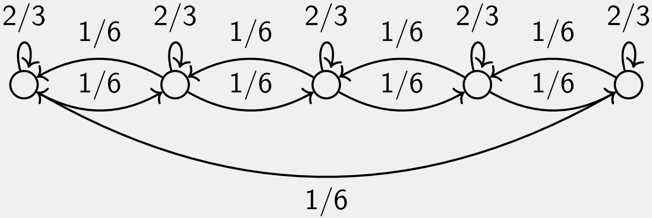
Sketch of Proof (Möbius Ladder Graph)



3-Regular Graphs

Sketch of Proof (Möbius Ladder Graph)

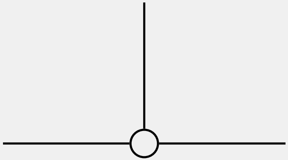
Once on the same cycle, we can shift back to this Markov chain. We combine the results to find that the bound is the same as before.



3-Regular Graphs

"Triangulating"

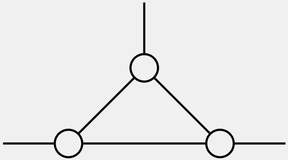
By "triangulating" a 3-regular graph, we mean replace each vertex with a complete graph of size 3.



3-Regular Graphs

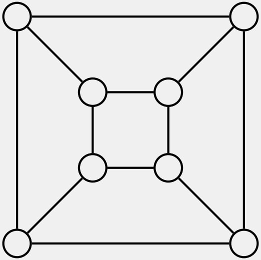
"Triangulating"

By "triangulating" a 3-regular graph, we mean replace each vertex with a complete graph of size 3.



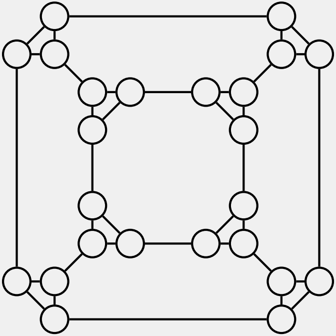
3-Regular Graphs

Example



3-Regular Graphs

Example



3-Regular Graphs

Results (Marshall Reber & White '18)

We found that when you triangulate the Möbius ladder graphs and prism graphs with n vertices, your bound transforms into

$$t_{\text{mix}}(\epsilon) \leq \frac{15n^2}{16\epsilon} + \frac{87}{5\epsilon}.$$

For the generalized Petersen graph GP(n, k), it transforms into

$$t_{\text{mix}}(\epsilon) \leq \frac{15|k|^2}{2\epsilon} + \frac{15}{2\epsilon}\left(\frac{n}{|k|}\right)^2 + \frac{9}{\epsilon}\left(\frac{n}{|k|}\right) + \frac{9}{\epsilon}|k| + \frac{108}{\epsilon},$$

where $|k| = n/\gcd(n, k)$.

3-Regular Graphs

Results (Marshall Reber & White '18)

We found that when you triangulate the Möbius ladder graphs and prism graphs with n vertices, your bound transforms into

$$t_{\text{mix}}(\epsilon) \leq \frac{15n^2}{16\epsilon} + \frac{87}{5\epsilon}.$$

For the generalized Petersen graph GP(n, k), it transforms into

$$t_{\text{mix}}(\epsilon) \leq \frac{15|k|^2}{2\epsilon} + \frac{15}{2\epsilon}\left(\frac{n}{|k|}\right)^2 + \frac{9}{\epsilon}\left(\frac{n}{|k|}\right) + \frac{9}{\epsilon}|k| + \frac{108}{\epsilon},$$

where $|k| = n/\gcd(n,k)$.

3-Regular Graphs

Results (Marshall Reber & White '18)

Using coupling, we were able to determine that for Möbius ladder graphs and prism graphs with n vertices, the mixing time for a (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3n^2}{16\epsilon} + \frac{6}{\epsilon},$$

and for the generalized Petersen graph $\text{GP}(n,k)$, the mixing time of the (slightly modified) lazy random walk is bounded by

$$t_{\text{mix}}(\epsilon) \leq \frac{3|k|^2}{2\epsilon} + \frac{3}{2\epsilon}\left(\frac{n}{|k|}\right)^2 + \frac{15}{\epsilon},$$

where $|k| = n/\gcd(n,k)$.

Remaining Questions

- Can we generalize this result to all vertex transitive 3-regular graphs?
- Can we extend it to all 3-regular graphs?
- Are the bounds we found above tight, or can we improve them?
- Does a similar result apply to lower bounds on these mixing times?

References

- David Aldous and Persi Diaconis, *Shuffling cards and stopping times*, The American Mathematical Monthly **93** (1986), no. 5, 333–348.
- Richard Durrett and R Durrett, *Essentials of stochastic processes*, vol. 1, Springer, 1999.
- David A. Levin, Yuval Peres, and Elizabeth L. Wilmer, *Markov chains and mixing times*, American Mathematical Society, 2006.
- Sidney I Resnick, *Adventures in stochastic processes*, Springer Science & Business Media, 2013.

Acknowledgements

Thanks to Indiana University, Dr. Chris Connell, and the NSF for the REU and the opportunity to work on the project. Special thanks to Dr. Graham White.