Markov Chains, Random Walks on Graphs, and the Laplacian

CMPSCI 791BB: Advanced ML

Sridhar Mahadevan
Random Walks

- There is significant interest in the problem of random walks
  - Markov chain analysis
  - Computer networks: routing
  - Markov decision processes and reinforcement learning
  - Spectral clustering and semi-supervised learning
  - Physical and biological processes

- There is a natural connection between a graph and a random walk on the vertices
  - Laplacian spectra and random walk convergence
  - How fast a random walk converges depends intimately on the spectra (which in turn depends on the topology of the graph)
Examples of Markov Chains

- **Knight’s tour:**
  - Assume a knight moves on an empty chessboard, starting from some position and randomly selecting one of the possible legal moves.
  - How long does it take the knight to return to its starting position?

- **Shuffling a deck of cards:**
  - What is the optimal algorithm for shuffling a deck of cards?
  - Riffle shuffle: $\sim \log_2 d$; Top-to-random shuffle: $\sim d \log_2 d$

- **Sampling from a complex distribution:**
  - Let $f: \mathbb{R}^d \to (0,1)$ be some arbitrary distribution
  - How do we efficiently sample from $f$ when $d$ is large?
  - Markov chain Monte Carlo (MCMC) methods
Markov Chain

- A Markov chain is a stochastic process on a finite (or infinite) set of states $S$ where
  - $P(s_{t+1} = j \mid s_t = i, s_{t-1} = k, …) = P_{ij}$ (the Markov property holds)
- Let $\alpha_0$ be an initial distribution on $S$
  - $\alpha_0(i) = P(s_0 = i)$
- We are interested in computing the long-term distribution as $t \to \infty$
- Note that
  - $P(s_1 = i) = \sum_k P(s_1 = i \mid s_0 = k) P(s_0 = k)$
  - Or, more generally, $\alpha_1 = \alpha_0 P = P^T \alpha_0^T$
- So, by induction, we get $\alpha_t = (P^T)^t \alpha_0$
Types of Markov Chain

- There are many different types of Markov chains
  - A state $s$ is called recurrent if $\alpha_t(s) > 0$ as $t \to \infty$ (or $P^t(s,s) > 0$)
  - A state $s$ is transient if $P^t(s,s) \to 0$ as $t \to \infty$
- Two states $u$ and $v$ are called communicating if there is a positive probability of reaching each from the other
  - $P^k(u,v) > 0$ and $P^l(v,u) > 0$, for $k > 0, l > 0$
- A set of states is irreducible if they form a communicating class
- A set of states is called ergodic if all states in the set communicate with each other, and don’t communicate with states outside the set
- A Markov chain is ergodic if its state space is irreducible and aperiodic
- A Markov chain is called aperiodic if
  - The gcd of all $t$ such that $P^t(s,s) > 0$ is 1
Examples of Markov Chains

What type of Markov chain does each of these define?

Which states are recurrent or transient?
Which chains are aperiodic?

Does a Knight’s tour define an ergodic Markov chain?
Limiting Distribution

- Let $P$ be a stochastic matrix defining a Markov chain.
- The stationary (or invariant) distribution $P^*$ is defined as
  \[ P^* = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k \]

- If a Markov chain is irreducible, the invariant distribution $P^*$ has identical rows.
- In other words, the long-term probability of being in a state does not depend on the starting distribution.
- If a state is transient, what can be said about the column corresponding to it in the invariant distribution?
Random Walk on Undirected Graphs

- Let \( G = (V, E, W) \) be a weighted undirected graph.
- \( G \) defines a natural random walk, where
  - The vertices of \( G \) are the states of a Markov chain.
  - The transition probabilities are \( p_{uv} = w_{uv} / d_u \)
  - where \( d_u = \sum_v w(u,v) \)
  - More compactly, we can write \( P = D^{-1} W \)
- What sort of Markov chain does this define?
- Does this chain “converge” to a stationary distribution?
- Is the rate of convergence related to the graph topology?
- How can we extend this model to directed graphs?
Reversible Random Walks on Graphs

- If an undirected graph is connected and non-bipartite, the Markov chain defined by the random walk is irreducible (and aperiodic).
- A random walk (or Markov chain) is called reversible if
  \[ \alpha^*(u) P(u,v) = \alpha^*(v) P(v,u) \]
- Random walks on undirected weighted graphs are reversible.
- Note from our earlier analysis that even though the random walk on a graph defines an asymmetric matrix, its eigenvalues are all real!

\[
D^{-1} W = D^{-1/2} \left( D^{-1/2} W D^{-1/2} \right) D^{1/2} = D^{-1/2} \left( I - L \right) D^{1/2}
\]
Perron’s Theorem for Positive Matrices

- A matrix $A$ is positive if all its elements are positive.
- **Theorem:** For any positive square matrix $A$ (of size $n$)
  - $\rho(A) > 0$, where $\rho$ is an eigenvalue.
  - For any other eigenvalue $\lambda$, $|\lambda| < \rho$.
  - The eigenvector (called the Perron vector) associated with $\rho$ has all positive elements.
Irreducible Matrices

- A matrix $A$ is said to be reducible if there exists a permutation matrix using which $A$ can be transformed into the following form:

$$
P^T A P = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}
$$

- A matrix which cannot be reduced is said to be irreducible.

- A matrix is nonnegative if $A \geq 0$.

- **Theorem**: A nonnegative matrix $A$ is irreducible if and only if $(I + A)^{n-1} > 0$.
Example

Is this irreducible?
Irreducibility Check

\[(I + P)^2 = \]

\[
\begin{array}{ccc}
1.6900 & 2.0300 & 0.2800 \\
0 & 2.8400 & 1.1600 \\
0 & 2.0300 & 1.9700 \\
\end{array}
\]
Example

Is this irreducible?
Irreducibility Check

\[(I + P)^2 = \]

\[
\begin{array}{ccc}
1.7600 & 1.9600 & 0.2800 \\
0.2800 & 2.6000 & 1.1200 \\
0.0700 & 1.9600 & 1.9700 \\
\end{array}
\]
Perron-Frobenius Theorem

- **Perron-Frobenius Theorem**: Let $A$ be an irreducible non-negative matrix (so that $A \succeq 0$).
  - Then, $A$ has a positive real eigenvalue $\rho$ such that every other eigenvalue $\lambda$ of $A$ has modulus $< \rho$.
  - Furthermore, the eigenvector $x$ corresponding to $\rho$ has all positive entries (and this vector is unique).
  - $\rho$ is a simple eigenvalue (geometrically and algebraically).
  - The largest eigenvalue of a matrix $A$ (in this case $\rho$) is called its spectral radius.
- **Simple corollary**: If $A$ is irreducible, $A^T$ is also irreducible. In this case, the positive eigenvector $x$ is a left eigenvector.
Applying the Perron-Frobenius Theorem

- Let $P$ be a irreducible stochastic matrix, where $\sum_j P_{ij} = 1$
- Since all its entries are non-negative, we can apply the Perron-Frobenius theorem.
- Note however that we know that the constant vector $1$ is an eigenvector of $P$, since
  \[ P \mathbf{1} = \mathbf{1} \]
- Hence, it must be the Perron vector!
- Thus, the spectral radius of $P$ is $\rho = 1$
Convergence of Random Walks

- Theorem: Let $G = (V, E, W)$ be a connected non-bipartite weighted undirected graph. Let $\alpha_0$ be an initial probability distribution on $V$. If

$$\lambda = \max (\lambda_i(P(G)) : \lambda_i(P(G)) \neq 1)$$

(i.e., $1-\lambda$ is the Fiedler eigenvalue)

- Then, for a simple random walk on $G$, we have

$$\| \alpha_t - \alpha_* \| < \lambda^t$$
Proof of Convergence Theorem

- Let $Q = P^T$, and let $y_i, \ 1 < i < n$ be the eigenfunctions of $Q$ which are orthonormal.
- Note that $\lambda_n(Q) = 1$ (why?) and therefore $\lambda < 1$.
- We apply the Perron-Frobenius theorem and deduce that all the components of $y_n(v) > 0$.
- Since the eigenfunctions $y_i$ form a basis for $\mathbb{R}^n$, we get
  \[ \alpha_0 = \sum_{i=1}^{n} w_i y_i \]
- Observe that $w_n = \langle \alpha_0, y_n \rangle > 0$.
- We can now write the distribution at time step $t$ as
  \[ \alpha_t = Q^t \alpha_0 = \sum_{i=1}^{n-1} w_i \lambda_i^t y_i + w_n y_n \]
Proof of Convergence Theorem

- Since $\lambda < 0$, $\alpha_t \rightarrow w_n y_n = \alpha_*$
- Hence, the difference can be written as

$$
\| \alpha_t - \alpha_* \|^2 = \| \sum_{i=1}^{n-1} w_i \lambda_i^t y_i \|^2
$$

$$
= \sum_{i=1}^{n-1} \| w_i \lambda_i^t y_i \|^2
$$

$$
= \sum_{i=1}^{n-1} w_i^2 \lambda_i^{2t} \| y_i \|^2
$$

$$
= \lambda^{2t} \sum_{i=1}^{n-1} w_i^2 < \lambda^{2t} \sum_{i=1}^{n} w_i^2 \leq \lambda^{2t}
$$
Implications of this result

- The theorem quantifies the relationship between the rate at which the random walk approaches the stationary distribution and the first non-zero eigenvalue of the normalized Laplacian (why?)

- By deriving bounds on $\lambda_1$ for example graphs, we can see how fast a random walk will mix on these graphs
  - Spielman’s lectures (2 and 3) derive lower bounds on canonical graphs.
  - For a path graph $P_n$ on $n$ nodes, $\lambda_1 > 4/n^2$

- The mixing rate of a random walk is defined as $1/(1 - \lambda)$
Laplacian for Directed Graphs

- In the next class, we will see how to apply the Perron-Frobenius theorem to define the Laplacian for directed graphs.