Foundations of Data Science & Machine Learning

Summary — Week 09
Devansh Singh Rathore
111701011

B.Tech. in Computer Science & Engineering
Indian Institute of Technology Palakkad

May 6, 2021

Abstract

This week we discuss about Markov Chain Monte Carlo Methods and Metropolis Hasting rule. Next we discuss Normalized Conductance.

1 Markov Chain Monte Carlo Methods

Goal: Sample a point x according to a distribution D. D is on domain X.

$$x \sim D$$

eg. σ is a permutation of [n] chosen uniformly at random.
$x \in \mathbb{R}^n$ is sampled from an n-dimensional gaussian.

Key Idea: (π : distribution over X)
Let the domain X be finite ($|X| = n$)
So, π will be a n-length prob. vector. i.e. $\pi = (\pi_1, \pi_2, ..., \pi_n)$ s.t. $\forall i,\ \pi_i \geq 0$ and $\sum_{i=1}^{n} \pi_i = 1$
Design a directed graph G and transition probabilities P s.t. π is a stationary distribution of P.

→ Desirable properties:
(1) G is strongly connected (unique stationary distribution)
(2) gcd(cycle length) = 1 ($p(t) \to \pi$)
(3) low degrees
(4) symmetries (regular for example)
(5) rapid mixing (convergence to stationary distribution)

Easy Case: π is uniform on X.
G : undirected connected non-bipartite graph with V(G) = X
: k - regular (k ¿ 2)
: $P_{ij} = 1/k \forall i, j$
: Expander.
→ $\pi_i = 1/n$
→ 1 is eigen vector of P & P^T (since $P = P^T$). So $\pi = (1/n, 1/n, ..., 1/n)$ is the stationary distribution of P.

Non-uniform π
If π is a prob. vector & P is a stochastic matrix s.t.,

$$\forall i, j \ \pi_i P_{ij} = \pi_j P_{ji} \quad -(*)$$

then π is a stationary distribution of P(proof given below).
Proof: Let $\sigma = P^T\pi$
then $\pi_j = \sum_{j=1}^{n} P_{ji}\pi_j$
$= \sum_{j=1}^{n} P_{ij}\pi_i \quad$ (using (*))
$= \pi_i \sum_{j=1}^{n} P_{ij}$
$= \pi_i$
So $\sigma_i = \pi_i \ \forall i$
Hence π is the stationary distribution of P.

1.1 Metropolis-Hasting

Input: π, a prob. distribution in [n]. (n length prob. vector)
Design of G:
Pick a "good" connected undirected graph H and replace each edge with 2 opposite arcs and add a self loop at each node to get G.

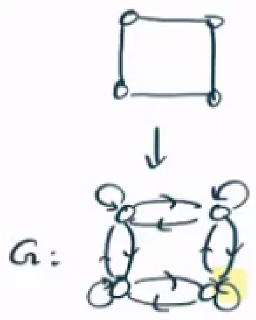

Fig 1.0 Generating graph G

Observation: G is strongly connected, gcd(cycle-length) = 1
Design of P: (Aim: $\forall i, j \ \pi_i P_{ij} = \pi_j P_{ji}$)
This is already satisfied for missing edges of H and self loops.

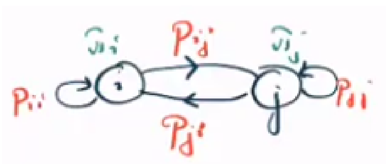

Fig 1.1 Designing P

Attempt 1:

$$P_{ij} = \pi_j \ \forall j \in N^+(i)$$

then,

$$\pi_i P_{ij} = \pi_i \pi_j$$

$$\pi_j P_{ji} = \pi_j \pi_i$$

also,

$$\sum_{j=1}^{n} P_{ij} = \sum_{j \in N^+(i)} \pi_j << 1 \text{ ,if H is sparse}$$

We can fix this by changing P_{ii}

$$P_{ij} = \begin{Bmatrix} \pi_j & , \; j \in N^+(i)\backslash\{i\} \\ 1 - \sum_{k \in N^+(i)\backslash\{i\}} \pi_k & , \; j = i \end{Bmatrix}$$

Issue: Too slow a walk because with max prob. it will travel self loop.

Scaling P_{ij}:

Subject to -

1. Equal scaling for P_{ij} and P_{ji}.
2. $\sum_{j=1}^{n} P_{ij} \leq 1$. (Deficit $\rightarrow P_{ii}$)

let r = maximum out degree of G not counting self loop

(2.) is ensured if $\forall i, j \; P_{ij} \leq 1/r \; (i \neq j)$

Scaling for $P_{ij} = (1/r)/max\{P_{ij}, P_{ji}\}$

Hence $\forall j \in N^+(i)\backslash\{i\}, \; P_{ij} = ((1/r)/(max\{\pi_i, \pi_j\})) \times \pi_j$

$$= (1/r) \times min\{1/\pi_i, 1/\pi_j\} \times \pi_j$$
$$= (1/r)min\{1, \pi_j/\pi_i\}$$

Thus,

$$\forall i \; P_{ij} = \begin{Bmatrix} (1/r)min\{1, \pi_j/\pi_i\} & , j \in N^+(i)\backslash\{i\} \\ 1 - \sum_{j \in N^+(i)\backslash\{i\} P_{ij}} & , j = i \end{Bmatrix}$$

Sanity Check of Metropolis-Hasting:

$\rightarrow \pi_i P_{ij} = (\pi_i/r)min\{1, \pi_j/\pi_i\}$

$= (1/r)min\{\pi_i, \pi_j\}$

$= (\pi_j/r)min\{\pi_i/\pi_j, 1\}$

$= \pi_j P_{ji}$

$\rightarrow \forall i, \; \sum_{j \in N^+(i)} P_{ij} = 1 \;\; (by \; definition)$

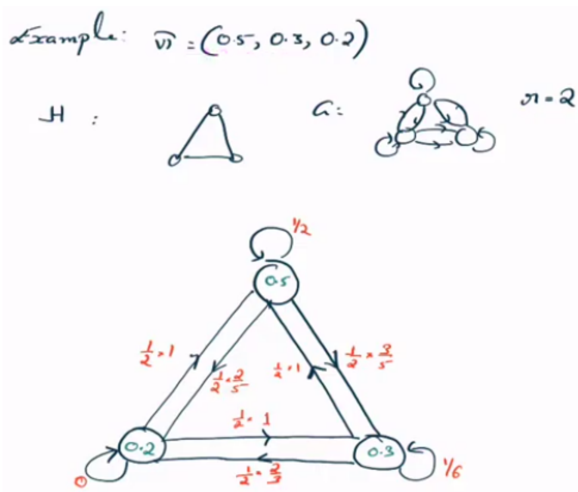

Fig 1.2 Example of using Metropolis-Hasting

Walker's Rule:

When at node i -

1. Select each out edge (i, j) w.p. 1/r. $(i \neq j)$
2. If $(\pi_j \geq \pi_i)$

   move to node j w.p. 1

   Else

   move to node j w.p. π_j/π_i

→ Usually, you can choose graph to be d dimensional lattice i.e. $[m]^d$:

1. $n = m^d \geq |X|$
2. Almost 2d - regular (except boundary vector). r = 2d.

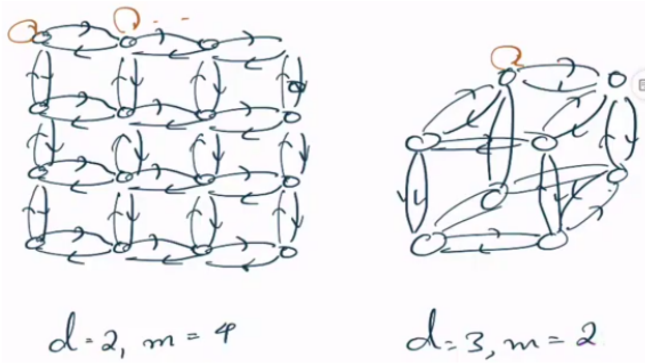

Fig 1.3 Graph Choices

2 Normalized Conductance

Definition: A Markov Chain with TPM P and a stationary distribution π is called **Time Reversible** if

$$\forall i, j\ \pi_i P_{ij} = \pi_j P_{ji}$$

⇒ Underlying digraph G is symmetric

⇒ G can be obtained from an undirected graph H by doubling each edge. (H may have self loops)

⇒ P can be encoded as weights on the edges of H.

$$w_{ij} = \pi_i P_{ij} \ \ (= \pi_j P_{ji})$$

Claim: If we know all w_{ij} values then we can compute all P_{ij} values.
Question: Can we find P from w_{ij}'s?

$$\sum_j w_{ij} = \sum_j \pi_i P_{ij} = \pi_i \sum_j P_{ij} = \pi_i$$

$$So,\ \pi_i = \sum_j w_{ij}$$

$$P_{ij} = w_{ij}/\pi_i$$

→ Sometimes w_{ij}'s may be scaled by an unknown constant ($w_{ij} = c\pi_i P_{ij}$)
In that case,

$$c = \sum_i \sum_j w_{ij}$$

$$w_i = \sum_j w_{ij} \quad (\pi_i = w_i/c)$$

$$p_{ij} = w_{ij}/w_i$$

Hence the weighted undirected graph H (with possibly self loops) defines the random walk.

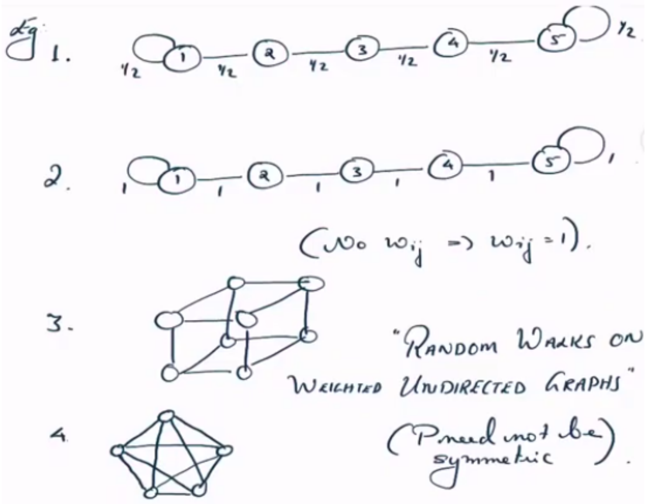

Fig 2.0 Examples

Definition: The **Normalised Conductance** $\phi(G)$ of a *weighted undirected graph* G is defined as

$$\phi(G) = min_{S \subseteq V(G),\, 0<W(S) \leq W(S')} \{ \sum_{x \in S,\, y \in S'} w_{xy}/W(S) \}$$

$$where\ W(S) = \sum_{x \in S} w_x$$

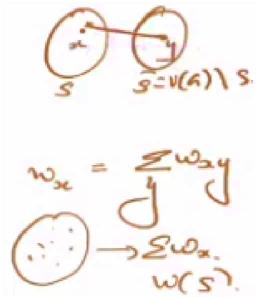

Fig 2.1 S, S', W(S)

Intution: (Assume $\sum_{x,y} w_{xy} = 1$ (i.e. c = 1), so that $\sum_y w_{xy} = \pi_x$ and $w_{xy} = \pi_x P_{xy}$)

$$P_{xy} = Pr[\text{next state} = y/\text{current state} = x]$$
$$= Pr[\text{next move is along the arc is xy/current state} = x]$$

and,

$$w_{xy} = \pi_x P_{xy} = Pr[\text{next move is along the arc xy}] \text{ in steady state}$$
$$\sum_{x \in S, y \in S'} w_{xy} = Pr[\text{next move is an escape from S}] \text{ in SS}$$
$$(1/W(S)) \sum_{x \in S, y \in S'} w_{xy} = Pr[\text{Next state is outside S/Current state is in S}] \text{ in SS}$$
$$= \text{Escape prob. from S}$$

ϕ = minimum escape prob. among all sets $S \subseteq V(G)$ with $o < W(S) \leq 1/2$

Definition: Norm. Cond. of a Cut -

for $S \subseteq V(G)$ with W(S) ¿ 0,

$$\phi(S, S') := (\sum_{x \in S, y \in S'} w_{xy}/min\{W(S), W(S')\}) \leq 1$$

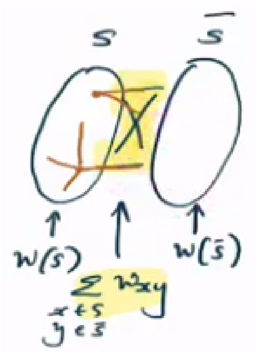

Fig 2.2 Normalized Conductance of a cut

$\phi(G) = \min \phi(S, S')$ where the minimum is over all non-trivial cuts of G.

Examples:

1. K_n with self loops
$w_{ij} = 1 \; \forall$ edges
$W_i = n \; \forall$ vertices
$W(S) = n|S|$

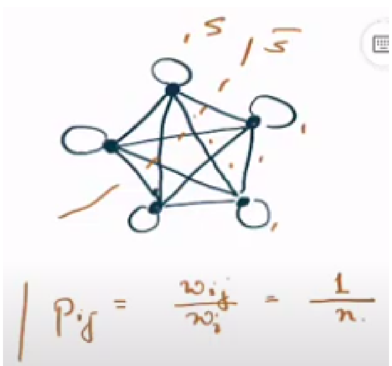

Fig 2.3 Example 1 graph

Let (S, S') be arb. with $k = |S| \leq |S'|$
$\phi(S, S') = k \times (n-k)/min\{nk, n(n-k)\}$
$\quad\quad\quad\quad = k \times (n-k)/nk = (n-k)/n$
$\phi(G) = min_{k \leq n/2}((n-k)/n) \approx 1/2$

2. P_n with end loop S

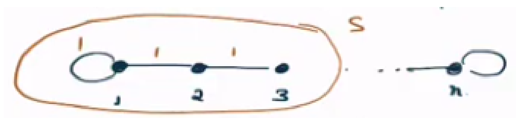

Fig 2.4 Example 2 graph

$w_{ij} = 1 \, \forall edges$
Let $S \subseteq [n]$
$W(S) = 2|S|$

Hence $|S| \leq |S'| \Rightarrow W(S) \leq W(S')$ $(|S| \leq n/2)$
$\phi(S,S') = \sum_{x \in S, y \in S'} w_{xy}/W(S)$
$\quad\quad\quad = |E(S,S')|/2|S|$
$W(G) = min_{|S| \leq n/2} \phi(S,S') \approx 1/(2.n/2) = 1/n$

3. G disconnected $\Rightarrow \phi(G) = 0$

Fig 2.5 Example 3 graph

Pick S to be V(C_i) with $W(V(C_i))$ smallest.

4. Dumbbell

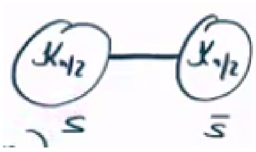

Fig 2.6 Example 4 graph

$(w_{ij} = 1 \; \forall edges)$
$W(S) = (n/2) \times (n/2) \times 1 \approx n^2/4$
$E(S,S') = 1$
$\phi(S,S') = 4/n^2$
$\phi(G) \leq 4/n^2$ (worse than path)
No monotonicity.

5. 2D - lattice with end loops

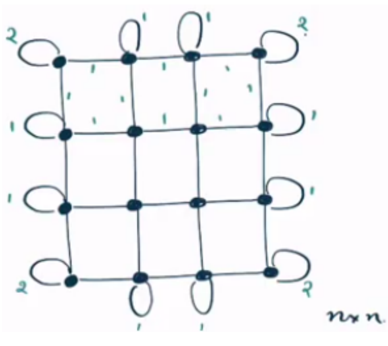

Fig 2.7 Example 5 graph

$W(S) = 4|S|$ hence $|S| < |S'| \Rightarrow W(S) \leq W(S')$. $(|S| = k)$

Observation 1:
If $|S| \leq 4/n^2$ then a corner square has least number of escape edges relative to $|S|$.
$E(S,S') \approx 2\sqrt{|S|}$
Hence, $\phi(S,S') = E(S,S')/W(S)$
$\approx (2\sqrt{|S|})/(4|S|) \leq 1/(2\sqrt{n^2/4}) = 1/n$

Observation 2:
If $n^2/4 < |S| \leq n^2/2$, then a margin strip has least no. of escape edges.
$E(S,S') \approx n$

Hence $\phi(S, S') \approx n/(4|S|) \leq n/(4.n^2/2) = 1/2n$
So, $\phi(G) = 1/2n = 1/2\sqrt{|V(G)|}$

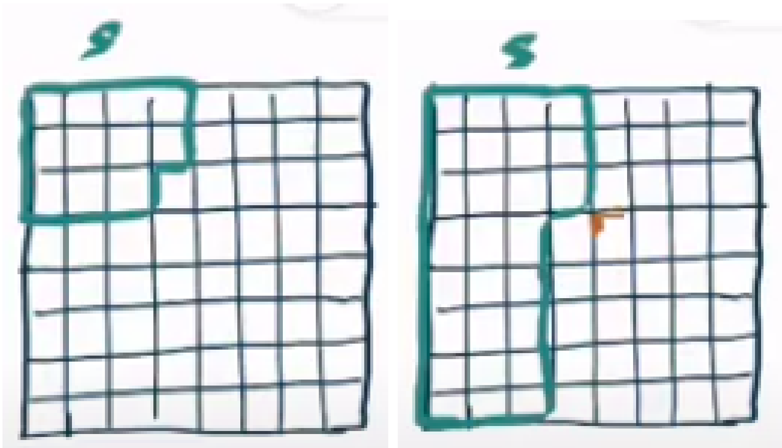

Fig 2.8 Example 5 graphs: observations 1 vs 2 (respectively)

6. d-dimensional lattice with end loops
$|V(G)| = n^d$
$\phi(G) = 1/dn = 1/(d|V(G)|^{1/d})$ (Exercise)

$\rightarrow$ **Inference:** Higher connectivity $\Rightarrow$ Higher conductance.