计算方法

刘景铖

计算机软件新技术国家重点实验室
南京大学

回顾：Random walk on graphs

给定图$G = (V, E)$

图上的随机游走:

- **从一个给定的顶点出发**
- **接下来，**每一步都从当前顶点，移动到一个均匀随机选取的邻居
- **不断重复**

这个随机过程的“长期表现”是怎么样的呢？

1. 重复t步之后，当前顶点是某个顶点u的概率是多少？
2. 是否存在一个极限的随机分布，**随机游走会收敛到它？（稳态）**
3. 多久才会收敛？(混合时间，mixing time)
4. 从点s出发，多久才会到达点t? (hitting time)
5. 多久才会遍历每个顶点至少一次？(遍历时间，cover time)

回顾：幂迭代与无向图上的随机游走

令A为无向图的邻接矩阵, D 为度数的对角矩阵. （注：因为是无向图，所以A是对称的）

记 $p_t(v)$为时间t在点v的概率，则有

$$p_{t+1}(v) = \sum_{u:uv\in E} p_t(u) \cdot \frac{1}{\deg(u)}$$

可以写出矩阵的形式：$\overrightarrow{p_{t+1}} = (AD^{-1})\overrightarrow{p_t}$，进而有$\overrightarrow{p_t} = (AD^{-1})^t\overrightarrow{p_0}$； $\overrightarrow{p_t}$为列向量

注意：之前第9次课里面使用的是行向量；一般Markov chain分析里面习惯使用行向量；这节课我们将会进行谱分析，这里更适合使用列向量的记号

转移矩阵$W := AD^{-1}$

要研究W^t, 只需要研究$\mathcal{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$的幂迭代, 因为:

$$\mathcal{A} = D^{\frac{1}{2}}\,W\,D^{-\frac{1}{2}} \Rightarrow W^t = D^{-\frac{1}{2}}\mathcal{A}^t D^{\frac{1}{2}}$$

对于实数对称矩阵$\mathcal{A} = V\Sigma V^T$的幂迭代：

$$\mathcal{A}^k = V\Sigma^k V^T = \sum_{1\le i\le n} \lambda_i^k v_i v_i^T$$

对称矩阵$\mathcal{A}$的特征值是怎么样分布的？上节课：邻接矩阵的特征值分布

无向图上随机游走的稳态分布

如果分布$\vec{\pi}$满足 $\vec{\pi} = W\vec{\pi}$，则称$\vec{\pi}$ 为稳态分布
稳态分布是一个“平衡态”/ 不动点
特别地：$\vec{\pi} = W^t \vec{\pi}, \forall t$

注意：如果极限分布存在，则一定是一个稳态分布

给定无向图 G，令 $\vec{d} \in \mathbb{R}^n$ 为顶点度数的向量，$m = |E|$

定理. 概率分布$\vec{\pi} = \frac{\vec{d}}{2m}$是$G$上随机游走的一个稳态分布.

换言之，向量$\vec{\pi} = \frac{\vec{d}}{2m}$是$AD^{-1}$的一个右特征向量，且特征值为1

Fundamental Theorem of Markov Chains

无向图上的马尔可夫链基本定理

- **是否对于任意无向图**，不管从什么样的p_0开始，总有$p_t \to \vec{\pi} = \frac{\vec{d}}{2m}$随着$t \to \infty$?
  - **不一定**：非连通的，或者二分图
- **对于无向图，连通性**即强连通，非可约；非二分图，则意味着非周期的

因此无向图上的马尔可夫链基本定理说的是：

对于任意有限的，连通的，非二分图，不管从什么样的p_0开始，总有$p_t \to \vec{\pi} = \frac{\vec{d}}{2m}$随着$t \to \infty$

Lazy Random Walks

对于二分图，可以考虑惰性随机游走：

- 每一步会以1/2的概率停留在当前状态；
- 以1/2的概率，随机地从邻居里面均匀地选取一个，作为下一步的状态。

用矩阵表示的话则有$p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0$.

定理. 对于任意有限的，连通的，不管从什么样的p_0开始，$p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0 \to \frac{\vec{d}}{2m}$

转移矩阵的谱分析

记 $W = AD^{-1}$ 为随机游走的转移矩阵，$Z = \frac{1}{2}I + \frac{1}{2}AD^{-1}$为惰性随机游走的转移矩阵

要研究$p_t = W^t p_0$，可以先分析W的特征空间

注意：尽管A是对称的，W不一定是对称的

不过 W与一个对称矩阵是相似的：$D^{-\frac{1}{2}}WD^{\frac{1}{2}} = D^{-\frac{1}{2}}(AD^{-1})D^{\frac{1}{2}} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = \mathcal{A}$

定理. W 与 $\mathcal{A}$ 拥有相同的特征值

注意：W不一定有两两正交的特征向量

d-regular graphs (d-正则图)

对于 d-正则图，有$W = \frac{1}{d}A = I - \frac{1}{d}L$，记$W$的特征值为$\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_n$， 正交正归化的特征向量为$v_1, v_2, \ldots, v_n$

要研究$p_t = W^t p_0$，可以把 $p_0 = c_1 v_1 + \cdots + c_n v_n$展开
那么 $W^t p_0 = c_1 \alpha_1^t v_1 + c_2 \alpha_2^t v_2 + \cdots + c_n \alpha_n^t v_n$

$W = \frac{1}{d}A = I - \frac{1}{d}L$ 的特征空间:

- $1 \ge \alpha_1$， $\alpha_n \ge -1$
- $1 = \alpha_1$，对应的特征向量为$\vec{1}$
- $\alpha_1 > \alpha_2$当且仅当图是连通的
- $\alpha_1 = -\alpha_n$当且仅当图是二分图（课后练习）

可见连通性，和非二分图的性质，分别对应于$\alpha_2 < 1$ 和 $\alpha_n > -1$

这意味着，对连通的非二分图，$W^t p_0 \to c_1 v_1$ as $t \to \infty$
如果$\alpha_2 < 1 - \epsilon$， 并且 $\alpha_n > -1 + \epsilon$，则ϵ越大，收敛越快

Fundamental Theorem of Markov chain, for undirected graph!

$c_1 v_1$是什么？

对于 d-正则图， $\vec{\pi} = \frac{\vec{d}}{2m} = \frac{1}{n}\vec{1}$为特征值为1对应的特征向量，因此$v_1 = \frac{1}{\sqrt{n}}\vec{1}, c_1 = \langle p_0, v_1 \rangle = \frac{1}{\sqrt{n}}\left\langle p_0, \vec{1} \right\rangle = \frac{1}{\sqrt{n}}$
因此$c_1 v_1 = \frac{1}{n}\vec{1} = \vec{\pi}$

Lazy Random Walks

要证明$p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0 \to \frac{\vec{d}}{2m}$随着 $t \to \infty$:

对 d-正则图， $W = \frac{1}{d}A$， $Z = \frac{1}{2}I + \frac{1}{2d}A$

若 W的特征值为$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$

则 Z的特征值为$1 = \frac{1+\alpha_1}{2} \geq \frac{1+\alpha_2}{2} \geq \cdots \geq \frac{1+\alpha_n}{2} \geq 0$

（课后练习）

定理. 对有限的连通无向图, 随着 $t \to \infty$， $p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0 \to \frac{\vec{d}}{2m}$， 不管从任何$p_0$出发.

混合时间 (mixing time)

既然 $p_t \to \vec{\pi} = \frac{\vec{d}}{2m}$ 收敛，收敛速度？

回顾：给定概率分布$\vec{p}$与$\vec{q}$，它们的TV距离定义为

$$d_{TV}(\vec{p},\vec{q}) := \frac{1}{2}\|p-q\|_1 = \frac{1}{2}\sum_{i=1}^{n}|p_i - q_i|$$

注意 $0 \le d_{TV}(\vec{p},\vec{q}) \le 1$

<u>**定义.**</u> ϵ-混合时间为最小的整数t 使得:

$$\|p_t - \pi\|_1 \le \epsilon \quad \forall\, p_0$$

<u>定义.</u> 谱间隔 (spectral gap): $\lambda = \min\{1-\alpha_2, 1-|\alpha_n|\}$

Mixing time (for d-regular graphs)

定理. 随机游走的ϵ-混合时间最多为$\frac{1}{\lambda}\log\left(\frac{n}{\epsilon}\right)$, 其中$\lambda = \min\{1-\alpha_2, 1-|\alpha_n|\}$.

证明：取$v_1, v_2, \ldots, v_n$为A的一组正交正规特征向量的基。

设$p_0 = c_1v_1 + c_2v_2 + \ldots + c_nv_n$，则$p_t = W^t p_0 = \underbrace{c_1\alpha_1^t v_1}_{\pi} + c_2\alpha_2^t v_2 + \cdots + c_n\alpha_n^t v_n$

由Cauchy-Schwarz不等式，$\|p_t - \pi\|_1 \le \sqrt{n}\|p_t - \pi\|_2$

$$\|p_t - \pi\|_2^2 = \|c_2\alpha_2^t v_2 + \cdots + c_n\alpha_n^t v_n\|_2^2 = c_2^2\alpha_2^{2t}\|v_2\|_2^2 + \cdots + c_n^2\alpha_n^{2t}\|v_n\|_2^2$$
$$= c_2^2\alpha_2^{2t} + \cdots + c_n^2\alpha_n^{2t} \le (1-\lambda)^{2t}(c_2^2 + \cdots + c_n^2)$$

注意到 p_0本身作为一个概率分布，满足$\|p_0\|_2^2 = \sum_i p_0(i)^2 \le \sum_i p_0(i) = \|p_0\|_1 = 1$

因此 $\|p_t - \pi\|_2^2 \le (1-\lambda)^{2t} \Rightarrow \|p_t - \pi\|_1 \le \sqrt{n}(1-\lambda)^t \le \sqrt{n}\, e^{-\lambda t}$

- 注：非d-regular的情况下$v_1, v_2, \ldots, v_n$需要取作$\mathcal{A}$的正交正规特征向量基，会有额外$\sqrt{n}$的损失
- 当spectral gap为常数时(i.e. $\lambda = \Omega(1)$)， 随机游走在$O\left(\log\frac{n}{\epsilon}\right)$步内收敛.
- 对正则图且$\lambda = \Omega(1)$时，我们可以在$O(\log n)$步内，均匀随机采样到一个顶点.

Mixing time and spectral gap

定理.随机游走的ϵ-混合时间最多为$\frac{1}{\lambda}\log\left(\frac{n}{\epsilon}\right)$, 其中 $\lambda = \min\{1 - \alpha_2, 1 - |\alpha_n|\}$

当谱间隔 (spectral gap)为常数的时候，随机游走在$O\left(\log\frac{n}{\epsilon}\right)$步内收敛

- 回忆PageRank，亦可通过随机游走模拟

对于惰性随机游走，谱间隔变为$\frac{\lambda_2}{2}$，其中λ_2为归一化的拉普拉斯矩阵的第二特征值

直观上看，谱间隔相当于描述了这个图有多么接近不连通的

回顾Cheeger's inequality，有$\lambda_2 \geq \frac{\phi(G)^2}{2}$
因此，Lazy random walk的ϵ-混合时间最多为$\frac{2}{\phi(G)^2}\log\frac{n}{\epsilon}$

随机采样

随机游走的应用之一，在于进行**随机采样**

- **扑克牌中**，如何进行洗牌？
  - 52张牌
  - 把每个排列顺序看成图的一个节点
  - 每次洗牌看成一次随机游走
  - 是否收敛？
  - 多少步收敛？
  - 7次“Rifle” shuffle，New York Times

- **随机的生成树** (Random spanning tree)
  - Basis exchange walk

- **随机的完美匹配**，近似计数等等

Cheeger's inequality

更加量化特征值的刻画：

- λ_2 非常小，当且仅当G 非常接近于不连通(i.e. 存在非常稀疏的割).
- λ_k非常小，当且仅当G 非常接近于存在k个连通分量(i.e. k个不相交的非常稀疏的分割).

一些联系与高级应用：

- **在算法上**，可以通过特征空间的信息，得到图的分割
- 在代数构造中，亦可通过代数构造特征空间，来得到对应的连通性比较好的图 (e.g. expander graph)
- **在分析随机游走的混合时间中**，可以通过分析图的连通性，间接地分析谱间隔 (spectral gap)，进而使用组合数学来分析**随机游走**

Perfect Matchings in d-regular Bipartite Graphs（选讲）

Random walks can also be used to design faster algorithms.

Perfect Matchings in O(n log n) Time in Regular Bipartite Graphs, by Goel, Kapralov, Khanna

- Traditional: find augmenting paths by e.g. BFS/DFS
- Idea: replace BFS/DFS by a random walk (on a different Eulerian directed graph, i.e. indegree=outdegree for every vertex)
- Expected time to find an augmenting path => expected return time of the random walk => stationary distribution value

- Open research problem: Can you extend this to non-regular bipartite graphs?

Expanders

- 组合视角: 有着较好的连通性的图
- 概率视角: 随机游走能迅速混合的图
- <u>代数视角</u>: 有着较大的谱间隔的图

设G为d-正则图, 并令$d = \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n \geq -d$ 为其邻接矩阵的特征值.

接下来，我们关注其<u>*谱半径(spectral radius)*</u>,

$$\alpha := \max\{\alpha_2, |\alpha_n|\}$$

如果 α 远小于d, 我们说图 G 有比较好的<u>spectral expansion</u>.

Expander Mixing lemma

Induced edges: $E(S,T) \coloneqq \{(u,v): u \in S, v \in T, uv \in E\}$

在这个定义中我们也允许相交的S,T, 此时交集中的每条边会被计算两次.

Expander Mixing lemma

设G为n个顶点的d-正则图，其谱半径为α，则对任意$S \subseteq [n], T \subseteq [n]$,

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha\sqrt{|S||T|}.$$

Proof: 注意到$E(S,T) = \chi_S^T A \chi_T$. 令 $\chi_S = \sum_i a_i v_i, \chi_T = \sum_i b_i v_i$, 其中 $\{v_i\}$ 为邻接矩阵A的一组正交正规特征基, 对应特征值$\{\alpha_i\}$.

$$E(S,T) = \frac{d|S||T|}{n} + \sum_{i \ge 2} \alpha_i a_i b_i \,.$$

由Cauchy-Schwarz,

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \le \alpha \|a\|_2 \|b\|_2 = \alpha \|\chi_S\|_2 \|\chi_T\|_2 = \alpha\sqrt{|S||T|}$$

Expander Mixing lemma

Intuition: Expander mixing lemma tells us that a spectral expander looks like a random graph.

Homework: Let G be a d-regular graph with spectral radius α. Show that the size of the maximum independent set of G is at most $\frac{\alpha n}{d}$.

Use this result to conclude that the chromatic number is at least $\frac{d}{\alpha}$.

Converse to Expander Mixing lemma

- Bilu and Linial
- Suppose that for every $S \subseteq [n], T \subseteq [n]$ with $S \cap T = \emptyset$, we have

$$\left| E(S,T) - \frac{d|S||T|}{n} \right| \leq \alpha\sqrt{|S||T|}.$$

Then all but the largest eigenvalue of A in absolute value is at most $O\left(\alpha\left(1 + \log\frac{d}{a}\right)\right)$.

- Based on LP duality, not very intuitive

Existence of expanders （选讲）

- Complete graphs are obviously the best expanders in terms of “expansion” (in all three notions of “expansion”)
- What’s interesting is the existence of sparse expanders: e.g. d-regular expanders for constant d

- A random d-regular graph is a (combinatorial) expander with high probability
- However, deterministic and explicit construction of expanders seems to be much harder to come up with

Alon-Boppana Bound （选讲）

- For d-regular graphs, how small can the spectral radius be?
- Ramanujan graphs: graphs whose spectral radius are at most $2\sqrt{d-1}$

Alon-Boppana Bound

Let G be a d-regular graph with n vertices, and α_2 be the second largest eigenvalue of its adjacency matrix. Then

$$\alpha_2 \geq 2\sqrt{d-1} - \frac{2\sqrt{d-1}-1}{\lfloor \mathrm{diam}(G)/2 \rfloor}$$

Alon-Boppana Bound （选讲）

<u>An easy lower bound on spectral radius</u>

Let G be a d-regular graph with n vertices, and α be its spectral radius. Then $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

Proof: Consider $Tr(A^2)$. Counting length-2 walks we have

$$Tr(A^2) \geq nd$$

On the other hand, $Tr(A^2) = \sum_i \alpha_i^2 \leq d^2 + (n-1)\alpha^2$.

Combined, we have $\alpha \geq \sqrt{d} \cdot \sqrt{\frac{n-d}{n-1}}$.

For the Alon-Boppana bound, one may consider $Tr\left(A^{2k}\right)$.

Trace method/trick

Random walks in expanders （选讲）

- We knew that it mixes rapidly, in time $O\left(\frac{\log n}{1-\epsilon}\right)$ for $\alpha = \epsilon d$.
- Perhaps surprisingly, not just the final vertex is close to the uniform distribution, but the entire sequence of walks looks like a sequence of independent samples for many applications.
- In fact, expander random walks can fool many test functions:
  - **Expander random walks: a Fourier-analytic approach, by Cohen, Peri and Ta-Shma**