



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



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)



Random walk on graphs

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)

前3个问题可以通过纯概率的方法(coupling)来研究；我们将讨论如何利用与幂迭代的联系，通过马尔可夫链(Markov chains)转移矩阵的特征值来研究。

后2个问题我们将留到之后讨论“图与电阻电路网络”的时候



幂迭代与无向图上的随机游走

例子：考虑无向图3-cycle

从顶点1出发

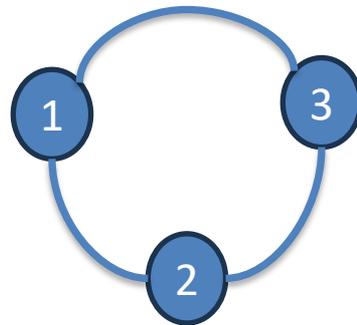
以概率 $\frac{1}{2}$ 走到顶点2

以概率 $\frac{1}{2}$ 走到顶点3

矩阵和向量的形式表示：

$$\text{起始概率分布 } p_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\text{随机游走一步之后，概率分布变成 } p_1 = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix}$$





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: $\vec{p}_{t+1} = (AD^{-1})\vec{p}_t$, 进而有 $\vec{p}_t = (AD^{-1})^t \vec{p}_0$; \vec{p}_t 为列向量

注意: Markov chain里面习惯使用行向量; 我们进行谱分析, 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:

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

幂迭代的收敛性? 对称矩阵 \mathcal{A} 的特征值是怎样分布的? 下节课将讨论。



图的谱

在传统的算法课上，无向图的邻接矩阵往往只是作为一个数据结构

在接下来的几节课里面，会把邻接矩阵作为矩阵研究：特征值，行列式，秩等等，并把这些代数性质与大家习惯的组合性质相联系

这节课，首先给大家介绍随机游走与马尔可夫链的基本概念



Markov chains

考虑一般的有向图上的随机游走，记 X_t 为时间 t 随机游走所处的状态
转移矩阵 $P_{i,j} = \Pr[X_{t+1} = j | X_t = i]$

记 $p_t(i)$ 为时间 t 在状态 i 的概率. 那么, 对于所有 $0 \leq j \leq n - 1$,

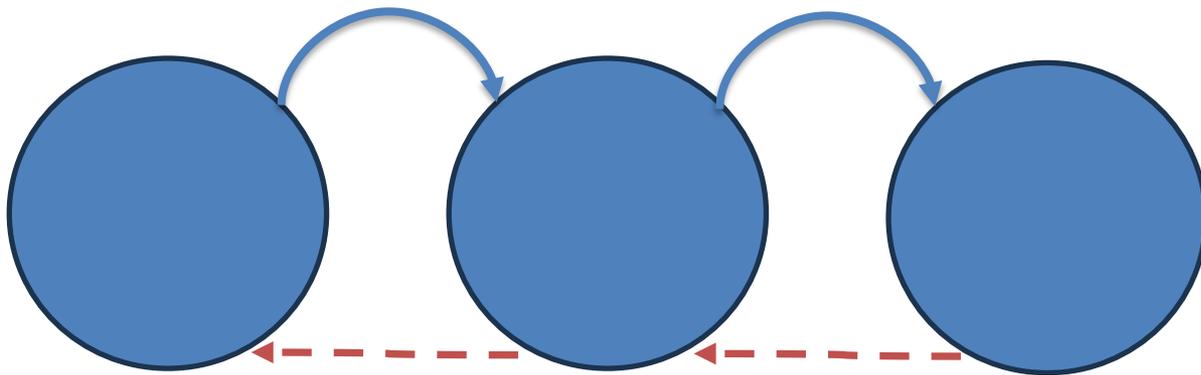
- $p_{t+1}(j) = \sum_{i=0}^{n-1} p_t(i) \cdot P_{i,j}$
- 更一般地, $\overrightarrow{p_{t+1}} = \overrightarrow{p_t} \cdot P$, 并且 $\overrightarrow{p_t} = \overrightarrow{p_0} \cdot P^t$. 这里 $\overrightarrow{p_t}$ 为行向量

注意: 转移只与时间 t 在各个状态的概率相关, 与过去无关



Irreducible Markov chains

对于有限的马尔可夫链，如果它对应的有向图是强连通的，则称它为不可约的(irreducible)。





Aperiodic Markov chains

给定一个马尔可夫链，其中状态 i 的周期定义为

$$period(i) := \gcd\{t \mid P_{i,i}^t > 0\}$$

- 如果一个状态 i 的周期 $period(i) = 1$ ，则称它为非周期性的(aperiodic)
- 如果马尔可夫链中每一个状态都是非周期性的话，则称该马尔可夫链是非周期性的；否则称它是周期性的。

例子：二分图，3-cycle，self-loops



Irreducible & aperiodic Markov chains

引理：如果一个马尔可夫链是不可约，且非周期的话，则一定存在足够大的常数 $T < \infty$ ，使得

$$(P^t)_{i,j} > 0, \quad \forall i, j, \forall t \geq T$$

证明思路：首先证明 $(P^t)_{i,i} > 0$ ，之后使用不可约的性质。



稳态分布 (stationary distribution)

如果分布 $\vec{\pi}$ 满足 $\vec{\pi} = \vec{\pi} \cdot P$, 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: $\vec{\pi} = \vec{\pi} \cdot P^t, \forall t$

注意: 如果极限分布存在, 则一定是一个稳态分布



概率分布的距离度量与收敛

要研究收敛，需要一个距离的度量；这里我们使用的是TV距离(total variation distance)。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 $0 \leq d_{TV}(\vec{p}, \vec{q}) \leq 1$

如果 $\lim_{t \rightarrow \infty} d_{TV}(\vec{p}_t, \vec{q}) = 0$ ，则说 $\{\vec{p}_t\}$ 收敛到 \vec{q}

(选讲) 也可以选用2范数

$$\left\| \frac{p}{\pi} - 1 \right\|_{2,\pi}^2 := \sum_{i=1}^n \pi_i \left(\frac{p_i}{\pi_i} - 1 \right)^2$$

类似于“方差”

由Cauchy-Schwarz不等式,

$$\sum_{i=1}^n |p_i - \pi_i| = \sum_{i=1}^n \pi_i \left| \frac{p_i}{\pi_i} - 1 \right| \leq \sqrt{\sum_{i=1}^n \pi_i \left| \frac{p_i}{\pi_i} - 1 \right|^2} = \left\| \frac{p}{\pi} - 1 \right\|_{2,\pi}$$

这意味着，如果2-norm 是小的，则1-norm 也是小的 (但反过来不一定).



回归时间 (Return time)

从状态 i 出发，第一次返回到状态 i 的时间，称之为状态 i 的回归时间，定义为

$$H_i := \min\{t \geq 1 \mid X_t = i, X_0 = i\}$$

期望回归时间定义为 $h_i := E[H_i]$.



Fundamental Theorem of Markov Chains

马尔可夫链的基本定理

对于有限，不可约，非周期的马尔可夫链，有以下事实：

1. 存在一个稳态分布 $\vec{\pi}$.
2. 随着 $t \rightarrow \infty$, \vec{p}_t 都会收敛到 $\vec{\pi}$, 无论从什么样的 \vec{p}_0 开始.
3. 稳态分布是唯一的
4. $\pi(i) = \frac{1}{h_i}$



Intuition

马尔可夫链的基本定理的一些直观理解

对于有限，不可约，非周期的马尔可夫链，有以下事实：

1. 存在一个稳态分布 $\vec{\pi}$.
2. 随着 $t \rightarrow \infty$ ， \vec{p}_t 都会收敛到 $\vec{\pi}$ ，无论从什么样的 \vec{p}_0 开始.
3. 稳态分布是唯一的
4. $\pi(i) = \frac{1}{h_i}$

考虑两个随机游走，当它们相遇之后，就变得“无法区别”了

回忆之前的引理：存在足够大的常数 T ，使得 $(P^t)_{i,j} > 0 \forall i, j, \forall t \geq T$

这意味着无论它们现在在何方，经过一段时间之后，都有非零的概率相遇

最终会相遇的概率是1，此时将收敛到同一个稳态分布

通过概率的coupling argument，可以把上述想法变成严格的证明

$\pi(i) = \frac{1}{h_i}$ ：期望回归时间为 h_i



Pagerank

考虑一个描述网页超链接的有向图

点集：网页

边集：如果网页 i 到网页 j 存在超链接，则 $i \rightarrow j$

搜索引擎想要给网页一个关于其内容质量/重要程度的打分

Pagerank的想法：

- 如果一个页面被很多人转载/引用，它是比较重要的；
 - 衡量标准：存在超链接指向它；
- 如果一个页面被其它重要的页面所指向，它本身可能也是比较重要的；
 - 衡量标准： ???



Pagerank

一个迭代算法:

- 一开始每个页面都有pagerank分数为 $1/n$
- 接下来的**每一步**，每个页面都会把自己的分数平均分开，并发送给它所指向的邻居
- 每个页面会更新自己的分数为，它所收到的分数的和
- 不断重复直到收敛



Pagerank

Pagerank分数收敛到的平衡态，正是随机游走的稳态分布。

$$Pagerank_{t+1}(j) = \sum_{i:i \rightarrow j \in E} Pagerank_t(i) / \deg_{out}(i)$$

$$\text{令 } P_{i,j} = \begin{cases} \frac{1}{\deg_{out}(i)}, & \text{if } ij \in E \\ 0, & \text{otherwise} \end{cases}, \text{ 则有 } \overrightarrow{Pagerank}_{t+1} = \overrightarrow{Pagerank}_t \cdot P$$

Pagerank满足与随机游走相同的迭代关系

- 当图是有限，不可约（强连通），非周期的时候，Pagerank分数是唯一的。（马尔可夫链的基本定理）
- Pagerank分数只与图的结构有关，与初始值无关
- Pagerank分数将会是随机游走期望回归时间的倒数
- 可以通过有向图上的随机游走模拟



Pagerank

实践中的图不一定是不可约（强连通），非周期的。

可以作以下修改：固定 ϵ ，把 $1 - \epsilon$ 的pagerank分数分给邻居， ϵ 的pagerank分数均匀分给所有点

对应的随机游走则是：

以概率 $1 - \epsilon$ ，走到一个均匀随机的邻居；

以概率 ϵ ，走到一个均匀随机的顶点。

可以证明，这个修改将保证图是强连通且非周期的而且不会改变网页的相对分数



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 \Rightarrow expected return time of the random walk \Rightarrow stationary distribution value
- Open research problem: Can you extend this to non-regular bipartite graphs?



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



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



无向图上的马尔可夫链基本定理

是否不管从任意 p_0 出发, 都有 $p_t \rightarrow \vec{\pi} = \frac{\vec{d}}{2m}$?

不一定: 如果Markov chain是reducible (不连通), 或者周期性的 (周期不为1)

对于无向的连通图, 非周期性等价于非二分图

(To see why, recall what happens to an undirected 3-cycle)

因此无向图上的马尔可夫链基本定理如下:

定理. 对任意有限的, 连通的, 非二分图, 都有 $p_t \rightarrow \vec{\pi} = \frac{\vec{d}}{2m}$ 随着 $t \rightarrow \infty$, 与 p_0 的选取无关



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， $p_t = \left(\frac{1}{2}I + \frac{1}{2}AD^{-1}\right)^t p_0 \rightarrow \frac{\vec{d}}{2m}$



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

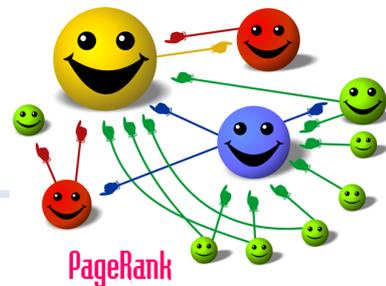
不过 W 与一个对称矩阵是相似的: $D^{-\frac{1}{2}}WD^{\frac{1}{2}} = D^{-\frac{1}{2}}(AD^{-1})D^{\frac{1}{2}} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = \mathcal{A}$

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



谱图理论 (Spectral graph theory)



谱分析: 特征值 + 特征向量
+ 相关的线性代数

图论与组合结构:

- 连通性 (Cheeger不等式)
- 图染色
- 聚类(Clustering)
- Mixing of random walks
- Expander graphs (efficient network, superconcentrators)

计算机科学:

- Pagerank
- 稀疏化Sparsification
- 迭代法解线性方程
- 电阻网络
- Expander codes (LDPC, Tanner codes)
- 不可近似性(Dinur's proof of the PCP theorem)
- 去随机化 (Derandomization)



矩阵的幂

回顾已经多次出现的矩阵的幂

- 对于可以对角化的矩阵 A ，收敛只需要特征值的幂是收敛的（特别地，谱半径 $\rho(A) = 1$ 也有可能是收敛的）

$$\begin{aligned}x &= \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n \\A^k x &= \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n\end{aligned}$$

- 但一般来说，可能需要使用：谱半径 $\rho(A) < 1$ 当且仅当 $\lim_{k \rightarrow \infty} A^k = 0$

– 课后练习：考虑 $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ 的特征值？ $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^k = ?$

– 注意：随机游走的转移矩阵一定不会是 $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ ，对应的只会是 $\begin{bmatrix} 1/2 & 1/2 \\ 0 & 1 \end{bmatrix}$

- 随机游走： $\vec{p}_{t+1} = \vec{p}_t \cdot P$ ，转移矩阵 $P := D^{-1}A$ ，及其相似的对角阵 $W = D^{-1/2}AD^{-1/2}$
 - 这门课：马尔可夫链基本定理在无向图上成立：对于连通的，非二分图，存在唯一的稳态分布，且会收敛
 - 完整证明可参照Olle Häggström的Finite Markov chains and algorithmic applications



Graph spectrum

考虑图的邻接矩阵，它的代数性质（如特征值与特征向量）能否与图本身的组合性质（连通性，是否二分图）相对应？

图的邻接矩阵的例子：完全图 $A = J - I = \vec{1} \vec{1}^T - I$
 J 的特征值和特征向量是什么？

$$J \vec{1} = n \vec{1}$$



Graph spectrum

考虑图的邻接矩阵，它的代数性质（如特征值与特征向量）能否与图本身的组合性质相对应？

图的邻接矩阵的例子：二分图

- 引理：对于二分图 G ，如果 α 是 $A(G)$ 的一个特征值，且重数为 k ，那么 $-\alpha$ 也是 $A(G)$ 的一个特征值，重数也是 k
- 特征值的重数(multiplicity)
 - 代数重数(algebraic multiplicity): 特征多项式里面，根的重复次数
 - 几何重数(geometric multiplicity): 特征值对应的特征空间的维度
 - 对于可对角化的矩阵（特别地，对于无向图的邻接矩阵，它们是一样的）



Graph spectrum

引理：对于二分图 G ，如果 α 是 $A(G)$ 的一个特征值，且重数为 k ，那么 $-\alpha$ 也是 $A(G)$ 的一个特征值，重数也是 k

证明：把 A 表示为
$$U \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} V$$

假设 $\begin{pmatrix} x \\ y \end{pmatrix}$ 是 A 的一个特征向量，对应特征值为 α ：

$$\begin{pmatrix} By \\ B^T x \end{pmatrix} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \alpha \begin{pmatrix} x \\ y \end{pmatrix}$$

因此 $B^T x = \alpha y, By = \alpha x$ 。再考虑 $\begin{pmatrix} x \\ -y \end{pmatrix}$

$$A \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{bmatrix} 0 & B \\ B^T & 0 \end{bmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -By \\ B^T x \end{pmatrix} = \begin{pmatrix} -\alpha x \\ \alpha y \end{pmatrix} = -\alpha \begin{pmatrix} x \\ -y \end{pmatrix}$$

因此 $-\alpha$ 也是 A 的特征值

最后，注意到： α 的重数为 $k \Leftrightarrow$ 存在 k 个线性无关的特征向量对应的特征值 α

对每一个分别取反后依然是线性无关的，因此 $-\alpha$ 的重数也为 k



Graph spectrum

事实上，引理的反方向也是对的。

引理：设 $A(G)$ 的特征值为 $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$ ，如果 $\forall i, \alpha_i = -\alpha_{n-i+1}$ ，那么 G 一定是二分图

证明：首先注意：对于任意奇数 $k, \sum_i \alpha_i^k = 0$

A 的特征值为 $\alpha_1, \alpha_2, \dots, \alpha_n$ 则 A^k 的特征值为 $\alpha_1^k, \alpha_2^k, \dots, \alpha_n^k$ ，因此对于任意奇数 k ,

$$\text{trace}(A^k) = \sum_i \alpha_i^k = 0$$

组合含义： $(A^k)_{i,j}$ = 长度为 k 的从 i 到 j 的行走方案的数目

$$\text{trace}(A^k) = \sum_i (A^k)_{i,i} = 0, \text{ 又因为 } (A^k)_{i,i} \geq 0, \text{ 所以 } (A^k)_{i,i} = 0$$

即，对于任意奇数 k ，长度为 k 的环路不存在。因此，一定是二分图



Largest eigenvalue of adjacency matrix

图的邻接矩阵A的最大特征值 $\alpha_{max} \leq \deg_{max}(G)$

证明: 设 v 为对应最大特征值的特征向量, 有 $Av = \alpha_1 v$

令 $v_j = \max_i v_i > 0$, 则 $(Av)_j = (\alpha_1 v)_j$

$$\Rightarrow \alpha_1 v_j = \sum_i A_{j,i} v_i \leq \deg_{max}(G) \cdot v_j$$

$$\Rightarrow \alpha_{max} \leq \deg_{max}(G)$$



Laplacian matrix

给定图 G ，它的拉普拉斯矩阵 $L(G)$ 定义为 $L(G) := D(G) - A(G)$ ，其中 $D(G)$ 是对角线上为度数的矩阵

对于 d -正则图， $L(G) = dI - A(G)$ ，因此正则图的拉普拉斯矩阵和邻接矩阵的特征空间是一样的。但一般图上并不成立。

可以写出 $L(G) = \sum_{e \in E} L_e$ ，其中 L_e 是只有 e 这一条边的图的拉普拉斯矩阵

另外： $x^T L x = \sum_{ij \in E} (x_i - x_j)^2$



Laplacian matrix

给定图 G ，它的拉普拉斯矩阵 $L(G)$ 定义为 $L(G) := D(G) - A(G)$ ，其中 $D(G)$ 是对角线上为度数的矩阵

性质： $\vec{1}$ 是 $L(G)$ 的一个特征向量，对应特征值是0

性质： $L(G) \succcurlyeq 0$



Connectedness

定理: 给定图 G , 它是连通的当且仅当 $L(G)$ 的特征值0的重数为1

证明(\Leftarrow): 先考虑 G 是不连通的情况。此时我们需要证明 $L(G)$ 的特征值0的重数大于1.

$$L(G) = \begin{matrix} & V_1 & V_2 \\ \begin{matrix} V_1 \\ V_2 \end{matrix} & \begin{bmatrix} L(G_1) & 0 \\ 0 & L(G_2) \end{bmatrix} \end{matrix}$$

有2组线性无关的特征向量, 对应特征值为0



Connectedness

定理：给定图 G ，它是连通的当且仅当 $L(G)$ 的特征值0的重数为1

证明(\Rightarrow): 接下来考虑 G 是连通的。假设 $Lx = 0$ ，则 $x^T Lx =$

$$\sum_{ij \in E} (x_i - x_j)^2 = 0$$

因此 $(x_i - x_j)^2 = 0, \forall ij \in E$ ，即 $x_i = x_j, \forall ij \in E$

因为 G 是连通的，所以也有 $x_i = x_j, \forall i, j \in V$

因此 $x = c \vec{1}$



Second eigenvalue

把 $L(G)$ 的特征值记为 $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

刚刚的定理证明了： G 是非连通的当且仅当 $\lambda_2 = 0$, 或者说 G 是连通的当且仅当 $\lambda_2 > 0$

更一般地（课后练习）：

$\lambda_k = 0$ 当且仅当 G 有至少 k 个连通分量。



Robust connectedness

连通性和二分图在组合上也非常容易刻画，为什么需要特征值？

事实上，可以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).
- 对于邻接矩阵的特征值 $\lambda_n \approx -\lambda_1$ 当且仅当 G 中存在接近于二分分量



Perron-Frobenius定理

设 A 是非负的，不可约且非周期的矩阵（不一定对称）。

1. 最大特征值的重数是1
2. 对应的特征向量中，每个维度都是非零且同号的
3. $|\lambda_i| < \lambda_1$, 对于 $2 \leq i \leq n$.

证明可参见：“Algebraic graph theory” Chapter 8, by
Godsil and Royle



Graph conductance

Recall that G is disconnected if and only if $\lambda_2 = 0$.

Cheeger's inequality will show that G is "close" to be disconnected if and only if λ_2 is "small".

We will first define precisely what it means to be close to be disconnected.

The conductance of a set $S \subseteq V$ is defined as

$$\phi(S) := \frac{|\delta(S)|}{\text{vol}(S)},$$

where $\text{vol}(S) := \sum_{v \in S} \deg(v)$.

When the graph is d -regular, $\phi(S) := \frac{|\delta(S)|}{d|S|}$.

Note: the expansion of a set S is defined as $\frac{|\delta(S)|}{|S|}$; For d -regular graphs, they're basically the same.

The conductance of a graph G is defined as $\phi(G) := \min_{S: \text{vol}(S) \leq m} \phi(S)$.

Note that $0 \leq \phi(G) \leq 1$.



Expander graphs and sparse cuts

A graph G with constant $\phi(G)$ (e.g. $\phi(G) = 0.1$) is called an expander graph.

A set S with small $\phi(S)$ is called a sparse cut.

Both concepts are very useful.

Finding a sparse cut is useful in designing divide-and-conquer algorithms, and have applications in

- image segmentation
- data clustering
- community detection
- VLSI-design
-



Spectral partitioning algorithm

A popular heuristic used in practice.

1. Compute an eigenvector $x \in \mathbb{R}^n$ corresponding to the second smallest eigenvalue of \mathcal{L} .
2. Sort the vertices so that $x_1 \geq x_2 \geq \dots \geq x_n$.
3. Let $S_i := \begin{cases} \{1, \dots, i\} & \text{if } i \leq \frac{n}{2} \\ \{i + 1, \dots, n\} & \text{if } i \geq \frac{n}{2} \end{cases}$

Return $\min_{1 \leq i \leq n} \phi(S_i)$.

This is simple and can be implemented in near-linear time.

It performs very well in practice, especially in image segmentation.



Normalized matrices

To state Cheeger's inequality nicely, we introduce normalized adjacency and Laplacian matrices.

Let $\mathcal{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ be the normalized adjacency matrix.

Let $\mathcal{L} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}}$ be the normalized Laplacian matrix.

Note that $\mathcal{L} = I - \mathcal{A}$.

When the graph is d -regular, $\mathcal{A} = \frac{1}{d}A$ and $\mathcal{L} = \frac{1}{d}L$.

Claim. Let $\alpha_1 \geq \dots \geq \alpha_n$ be the eigenvalues of \mathcal{A} , and $\lambda_1 \leq \dots \leq \lambda_n$ be the eigenvalues of \mathcal{L} .

Then, $1 = \alpha_1 \geq \dots \geq \alpha_n \geq -1$, and $0 = \lambda_1 \leq \dots \leq \lambda_n \leq 2$.



Cheeger's inequality

Cheeger's Inequality [Cheeger 70, Alon-Milman 85]

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}$$

The first inequality is called the easy direction, and the second inequality is called the hard direction. We give some intuition in the case when G is a d -regular graph.

For the easy direction: think of λ_2 as a “relaxation” of the graph conductance problem.

$$\phi(G) \approx \min_{x \perp 1 : x \text{ is binary}} \frac{\sum_{i,j \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2} \quad \text{and} \quad \lambda_2 = \min_{x \perp 1} \frac{\sum_{i,j \in E} (x_i - x_j)^2}{d \sum_{i \in V} x_i^2}.$$

For the hard direction: given an optimizer x for λ_2 , we want to produce a set S with $\phi(S) \leq \sqrt{2\lambda_2}$. The idea is to use the spectral partitioning algorithm: for a “fractional” x , we try to round it to an integral (binary) solution x . This is known as “rounding”.