



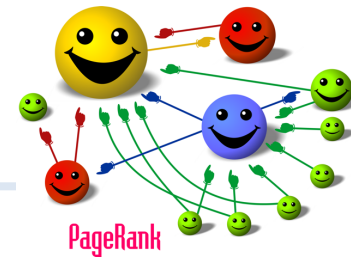
计算方法

刘景铎

计算机软件新技术国家重点实验室
南京大学



谱图理论 (Spectral graph theory)



Spectral theory

eigenvalues + eigenvectors + related linear algebra

Graph structures

- Connectedness
- Coloring / Clustering
- Mixing of random walks
- Expander graphs

In Theoretical CS

- Pagerank
- Sparsification
- Solving linear systems
- Counting / Sampling
- Expander codes
- Hardness of approximation
- Derandomization
- Max flow and more

And Beyond

- Image segmentation
- Electrical networks
- Reliable / Efficient networks
- Epidemic modelling
- [Economic networks](#)



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 $W = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$
 - 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



Complexity of Linear algebra

All the following can be solved in $\tilde{O}(n^\omega)$ arithmetic operations:

- Matrix multiplication
- Matrix inverse
- Determinant
- Characteristic polynomial
- Solving linear equations $Ax = b$
- Singular value decomposition
- Eigen-decomposition of symmetric matrices

In fact, almost linear time (in theory) for matrices that we will care about..



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？

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



Largest eigenvalue of adjacency matrix

图的邻接矩阵 A 的最大特征值 α_1 满足

$$d_{\text{avg}} \leq \alpha_1 \leq \deg_{\text{max}}(G)$$

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 $(Av)_j = (\alpha_1 v)_j$

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

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

Remark. 同样的思路可以证明:

对连通的图 G , $\alpha_1 = d_{\text{max}}$ iff G 是正则图



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 (2-colorability)

- 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
$$\begin{matrix} & U & V \\ U & \begin{bmatrix} 0 & B \end{bmatrix} \\ V & \begin{bmatrix} B^T & 0 \end{bmatrix} \end{matrix}$$

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



Laplacian matrix

给定图 G

- 邻接矩阵 $A(G)$

$$A_{u,v} = 1 \text{ iff } uv \in E$$

- 拉普拉斯矩阵: $D(G)$ 是对角线上为度数的矩阵

$$L(G) := D(G) - A(G)$$

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, 这一区别非常重要

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 $L_e = b_e b_e^\top$

$$L_e = \begin{pmatrix} & u & & v & \\ & \vdots & & \vdots & \\ \cdots & 1 & \cdots & -1 & \cdots \\ & \vdots & & \vdots & \\ \cdots & -1 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \end{pmatrix} \begin{matrix} u \\ \\ v \\ \end{matrix}$$



Laplacian matrix

给定图 G ，它的拉普拉斯矩阵 $L(G)$ 定义为

$$L(G) := D(G) - A(G) = \sum_{e \in E(G)} L_e = \sum_{e \in E(G)} b_e b_e^\top$$

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

证明： $L(G)$ 每行加起来为0.

性质： $L(G) \succcurlyeq 0$

证明： $x^\top L x = \sum_e x^\top b_e b_e^\top x = \sum_e (x_u - x_v)^2 \geq 0$

因此， $L(G)$ 是半正定矩阵，且最小特征值为0



Connectedness

定理： 给定图 G ，它是连通的当且仅当 $L(G)$ 特征值0的重数为1

证明($\Rightarrow$): 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”.



回顾：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)

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

由 $p_t = W^t p_0$ ，需要分析 W 的特征值分布



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 $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$, 正交正归化的特征向量为 $v_1, v_2, \dots, v_n$

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

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

- $1 \geq \alpha_1, \alpha_n \geq -1$
- $1 = \alpha_1$, 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, $W^t p_0 \rightarrow c_1 v_1$ as $t \rightarrow \infty$

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 $v_1 = \frac{1}{\sqrt{n}} \vec{1}, c_1 = \langle p_0, v_1 \rangle = \frac{1}{\sqrt{n}} \langle p_0, \vec{1} \rangle = \frac{1}{\sqrt{n}}$
因此 $c_1 v_1 = \frac{1}{n} \vec{1} = \vec{\pi}$



Lazy Random Walks

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

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 $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$

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

(课后练习)

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



混合时间 (mixing time)

既然 $p_t \rightarrow \vec{\pi} = \frac{\vec{d}}{2m}$ 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 $0 \leq d_{TV}(\vec{p}, \vec{q}) \leq 1$

定义. ϵ -混合时间为最小的整数 t 使得:

$$\|p_t - \pi\|_1 \leq \epsilon \quad \forall p_0$$

定义. 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 $v_1, v_2, \dots, v_n$ 为 A 的一组正交正规特征向量的基。

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

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

$$\begin{aligned}\|p_t - \pi\|_2^2 &= \|c_2 \alpha_2^t v_2 + \dots + c_n \alpha_n^t v_n\|_2^2 = c_2^2 \alpha_2^{2t} \|v_2\|_2^2 + \dots + c_n^2 \alpha_n^{2t} \|v_n\|_2^2 \\ &= c_2^2 \alpha_2^{2t} + \dots + c_n^2 \alpha_n^{2t} \leq (1 - \lambda)^{2t} (c_2^2 + \dots + c_n^2)\end{aligned}$$

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

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

- 注: 非d-regular的情况下 $v_1, v_2, \dots, v_n$ 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 $\frac{2}{\phi(G)^2} \log \frac{n}{\epsilon}$



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