MATH IN AI WINTER 2023

Graph Theory

Lecture Notes

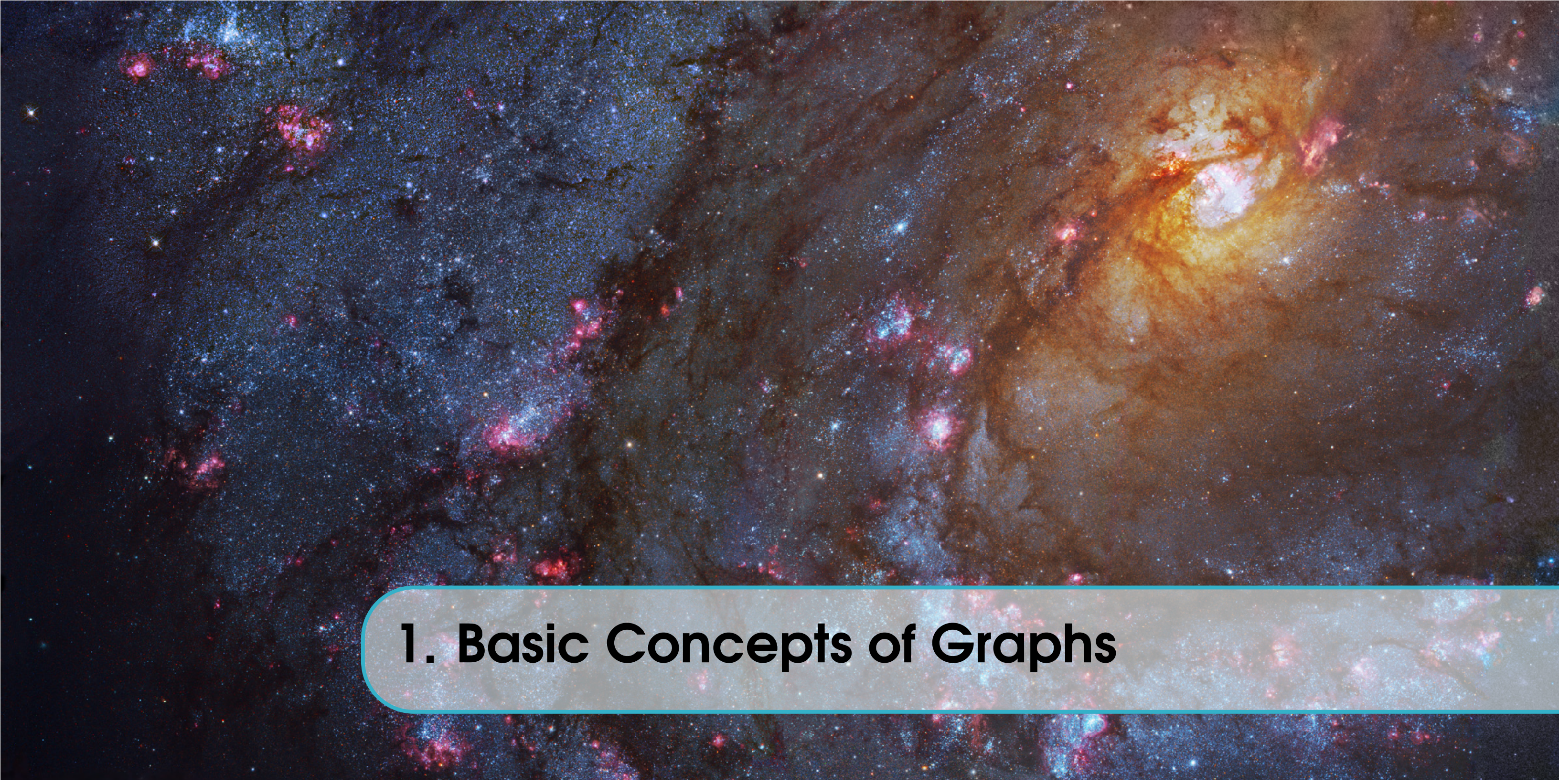
Flaricy

https://github.com/flaricy/math-in-ai-notes

The author hopes to take notes while learning graph theory. Reference books are *Algebraic Graph Theory* and 离散数学基础. Starts from Dec 8th.
released now

Contents

This page is intentionally left blank.



1. Basic Concepts of Graphs

1.1 Simple graph, complete graph, tournament graph

Definition 1.1.1 — simple graph. 有向图或者无向图，如果无平行边（重边）和自环。

1.2 Operations on a graph

点、边的删除；收缩；两个图之间的运算

1.3 Havel-Hakimi algorithm & Erdos-Gallai theorem

给定一个度数序列 $\{d_i\}$，判断是否可以根据这个度数序列构造出简单无向图。

Theorem 1.3.1 — Havel-Hakimi. Let $d = (d_1, d_2, ..., d_n)$, $\sum_{i=1}^{n} d_i = 0(mod 2)$ 且 $n-1 \geq d_1 \geq d_2 \geq ... \geq d_n \geq 0$, then d 简单可图化 $\iff d' = (d_2 - 1, d_3 - 1, ...d_{d_1+1} - 1, ...)$ 简单可图化。

Proof. *to be written* ■

Theorem 1.3.2 — Erdos-Gallai. 设 $d = (d_1, d_2, ...d_n)$ 满足 $n-1 \geq d_1 \geq d_2 \geq ... \geq d_n \geq 0$, 则 d 简单可图化 $\iff \forall 1 \leq r \leq n-1$,

$$(1) \sum_{i=1}^{r} d_i \leq r(r-1) + \sum_{i=r+1}^{n} \min\{r, d_i\} \qquad (2) \sum_{i=1}^{n} d_i = 0(mod 2)$$

1.4 通路与回路

Definition 1.4.1 — 通路. 设G为无向标定图,G中顶点与边的交替序列 $\Gamma = v_{i_0} e_{j_1} v_{i_1} e_{j_2} ... e_{j_l} v_{i_l}$ 称为 顶点 v_{i_0} 到 v_{i_l} 的 通路。

Definition 1.4.2 — walk 简单通路. 若Γ中所有边各异。

Definition 1.4.3 — closed walk 简单回路. 若Γ是简单通路且 $v_{i_0} = v_{i_l}$

Definition 1.4.4 — path 初级通路，路径. 如果Γ的所有顶点各异,所有边各异。

Definition 1.4.5 — 初级回路，圈. 若Γ是初级通路且起始点 = 终点。

R

- 初级通路是简单通路。
- 上述定义针对的是无向图。有向图类似。

Definition 1.4.6 — Girth. 围长。简单无向图中最短圈的长度。

Definition 1.4.7 — Perimeter. 周长。简单无向图中最长圈的长度。

2. Eulerian Graph and Hamiltonian Graph

2.1 Eulerian Graph

Definition 2.1.1 — Eulerian trail/path 欧拉通路. an Eulerian trail (or Eulerian path) is a trail in a finite graph that visits every edge exactly once (allowing for revisiting vertices).

Definition 2.1.2 — Eulerian circuit/cycle. an Eulerian trail that starts and ends on the same vertex

Definition 2.1.3 — Eulerian Graph. 具有欧拉回路的图。规定平凡图为欧拉图。

Definition 2.1.4 — 半欧拉图. 具有欧拉通路但没有欧拉回路的图。

Lemma 2.1.1 如果G是欧拉图，那么从G的任何一个顶点出发都可以找到一条简单通路构成欧拉回路。

Proof. 由定义验证即可。 ■

Theorem 2.1.2 Suppose G is a indirected graph. The following 3 propositions are equivalent:
(1) G is Eulerian
(2) The degree of every vertex in G are even. And G is connected.
(3) G is the union of several cycles with no intersected edges. And G is connected.

Proof. (1) $\Longrightarrow$ (2)
如果 G 平凡，结论显然。否则结论也显然。
(2) $\Longrightarrow$ (3)
我们证明: G的任意顶点度数为偶数 $\Longleftrightarrow$ G是若干个边不交的权的并。 对边数归纳：
如果边数 $=2$， 结论显然。 假设$2(m-1)$条边时成立,当$2m$条边时。找一个连通分支$G' =< V', E' >$, 设其中的一条极大路径为$v_1, v_2, ..., v_n$,由于$d(v_1) \geq 2$， 从而存在 $v_i (i \geq 2)$，使得v_1 和 v_i 相邻，从而找到一个简单回路L。从G'中删去L，即可归纳。
(3) $\Longrightarrow$ (1)

对圈的个数归纳。
一个圈时显然成立。假设m-1个圈时成立，则$m \geq 2$个圈时：先任意选一个圈L_1，由于图连通，从而$\exists v_1, v_2 \in L_1, L_2$, v_1, v_2可以是同一个点。由归纳假设和Lemma 1，可以从$v1$出发在剩下的圈中走到$v2$,最后绕L_1这个圈。 ■

Theorem 2.1.3 — 半欧拉图的判定. 设G为连通的无向图，则G是半欧拉图当且仅当G中恰好有两个奇度顶点。

Proof. 在这两个顶点之间连一条边。由Theorem 3.1.2知存在欧拉回路，然后在回路中去掉那条添加的边即可。
假设由欧拉回路，那么由Theorem 3.1.2知没有奇数度的顶点。矛盾。 ■

Corollary 2.1.4 设G为连通的无向图，G中有$2k$个奇度顶点，则G中存在k条边不交的简单通路$P_1, P_2, ...P_k$,使得$E(G) = \cup_{i=1}^{n} E(P_i)$.

Proof. 将2k个点两两配对，每对连一条边。在新图中取一个欧拉回路，然后删去这k条边，一定能得到k段不交的简单通路。 ■

Theorem 2.1.5 — 有向欧拉图的判定. 设D为有向图，则下面三个命题等价：
(1) D is Eulerian.
(2) D is connected and $\forall v \in V(D), d^+(v) = d^-(v)$.
(3) D is connected and is the union of several indirected cycles with no intersected edges.

Proof. (1) $\Longrightarrow$ (2) : trival
(2) $\Longrightarrow$ (3) : similar to Theorem 8.1.1
(3) $\Longrightarrow$ (1) : similar to Theorem 8.1.1 ■

Theorem 2.1.6 — 有向半欧拉图的判定. D中恰有两个奇数度的顶点，且一个点出度比如度大1，一个点入度比出度大1.其他点如度都等于出度。

Definition 2.1.5 — Fleury's algorithm. 求无向图中的欧拉回路。

Definition 2.1.6 — 逐步插入回路算法.

2.2 Hamiltonian Graph

Proposition 2.2.1 — motive. (1859 Willian Hamilton) **Traverse the World Problem**
正十二面体图上是否存在一条初级回路（圈）遍历所有顶点?

Ⓡ [Traveling Salesman Problem, TSP] 在一个赋权的无向图中，去找一个哈密尔顿回路，并且使得该回路的总权值最小。
这是一个NP-complete问题。

Definition 2.2.1 Hamiltonian path

- 经过所有点恰好1次的通路称为哈密顿路。
- 如果上面的通路是回路，那么称为哈密顿回路。
- 具有哈密顿回路的图称为哈密顿图。
- 具有哈密顿通路但没有哈密顿回路的图称为半哈密顿图。
- 规定：平凡图是哈密顿图。

Theorem 2.2.2 $G=<V,E>$ is Hamiltonian. For every $\emptyset \neq V_1 \subsetneq V$ we have

$$p(G-V_1) \leq |V_1|$$

, where $p(G)$ denotes the number of connected components of G.

Sketch.
Fact 2.2.3 往一个图中添加边，那么连通分支数不会减少。
所以只需要考虑G中哈密顿回路所包含的边，如果删去一些节点，对连通分支数的影响。 ■

Corollary 2.2.4 — case of semi-Hamiltonian graph. 设$G=<V,E>$ is semi-Hamiltonian, For every $\emptyset \neq V_1 \subsetneq V$ we have

$$p(G-V_1) \leq |V_1|+1$$

, where $p(G)$ denotes the number of connected components of G.

Proof. 同样考虑在哈密顿通路（不是回路）中删除一些点能产生多少段即可。 ■

■ **Example 2.1 — Peterson Graph.** Show that Peterson Graph is semi-Hamiltonian. ■

Ⓡ 彼得森图满足Theorem 3.2.2，但不是哈密顿图。

Theorem 2.2.5 — a sufficent condition. 设G为$n \geq 1$ 阶简单无向图，若对于G中不相邻的任意两点v_1, v_2，均有

$$d(v_1)+d(v_2) \geq n-1$$

则G中存在哈密顿通路。

Proof. First, show that G is connected.
if G is not connected, then consider two connected components G_1, G_2, pick $v_1 \in G_1, v_2 \in G_2$. We have $d(v_1) \leq |V(G_1)|-1, d(v_2) \leq |V(G_2)|-1$, which implies $d(v_1)+d(v_2) \leq |V|-2$. Contradiction.
考虑极大路径法。
1. 任选一条极大路径 $\Gamma = v_1 v_2 \ldots v_l$。如果 $l=n$, 那么这就是哈密顿通路。如果不是：
2. 证明存在一个圈经过Γ 上所有顶点。为此，只要证明：存在顶点 $v_i \in \Gamma$ 使得 v_{i-1} 与 v_1 相邻，v_i 与 v_n 相邻。这一点可以通过条件证明。

因为图G中有Γ中未出现的点，所以可以找到这样一个点和上面构造的圈中某个点相邻，我们就可以找到一条更长的哈密顿通路。
3. 重复此过程，我们可以在有限步之内得到最长的极大通路，此即哈密顿通路。 ■

Corollary 2.2.6 — Øystein Ore, Norwegian. 设G为$n \geq 1$阶简单无向图，若对于G中不相邻的任意两点v_1, v_2，均有

$$d(v_1) + d(v_2) \geq n$$

则G中存在哈密顿通路。

Proof. 由Theorem3.2.5知，G中存在哈密顿通路$\Gamma = v_1 v_2 \dots v_n$，如果$v_1 v_n$相邻，那么找到了回路。如果不相邻，用类似的方法可以找到一个圈。 ■

Corollary 2.2.7 设G为$n \geq 1$阶简单无向图，若对于G中任意v，均有

$$d(v) \geq \frac{n}{2}$$

则G中存在哈密顿通路。

Theorem 2.2.8 设u, v为无向n阶简单图G中的任意两个不相邻的顶点，且$d(u) + d(v) \geq n$，则
G为哈密顿图 $\iff$ $G \cup e = (u, v)$ 为哈密顿图。

Proof. $\Longrightarrow$ trivial.
$\Longleftarrow$ 如果$G \cup e = (u, v)$ 的一条哈密顿回路中有边 e，那么删去这条边，用同样的方法构造一个圈。 ■

■ **Example 2.2** 对于 $n \geq 4$阶简单无向图G，只要$\delta(G) \geq \frac{n}{2} + 1$，G中至少存在2条不同的哈密顿回路。 ■

Theorem 2.2.9 设D为$n \geq 2$阶竞赛图,则D具有哈密顿通路。

Proof. induction by n.
n = 2 : trivial
Suppose n = k OK. When n = k + 1. WLOG, let $\Gamma = v_1 v_2 \dots v_k$ be a Hamiltonian path of $G - v_{k+1}$, If $\forall 1 \leq i \leq k$, directed edge $(v_i, v_{k+1}) \in E(G)$, then $\Gamma' = v_1 \dots v_k v_{k+1}$ is the desired Hamiltonian path. Otherwise, there $\exists r \in \{1, 2, \dots, k\}$ such that $(v_i, v_{k+1}) \in E(G), \forall i < r$, but $(v_{k+1}, v_r) \in E(G)$. Then $\Gamma' = v_1 \dots v_{r-1} v_{k+1} v_r \dots v_k$ is the desired path. ■

下面探讨竞赛图中何时有哈密顿回路。我们假定竞赛图是强连通的，那么有如下的两个引理。

Lemma 2.2.10 强连通的竞赛图$(n \geq 3)$中存在长度为3的圈。

Proof. Take any vertex $v_0 \in D$. Let $\Gamma_D^+(v_0) = \{v | < v_0, v > \in E(D)\}$, $\Gamma_D^-(v_0) = \{v | <, v_0 > \in E(D)\}$. We claim that both Γ^+ and Γ^- are non-empty. Moreover, there $\exists u \in \Gamma^+, v \in \Gamma^-$ such that $(u, v) \in E(D)$ (otherwise D is not strongly connected). Thus, $v_0 \to u \to v \to v_0$ forms a 3-loop. ■

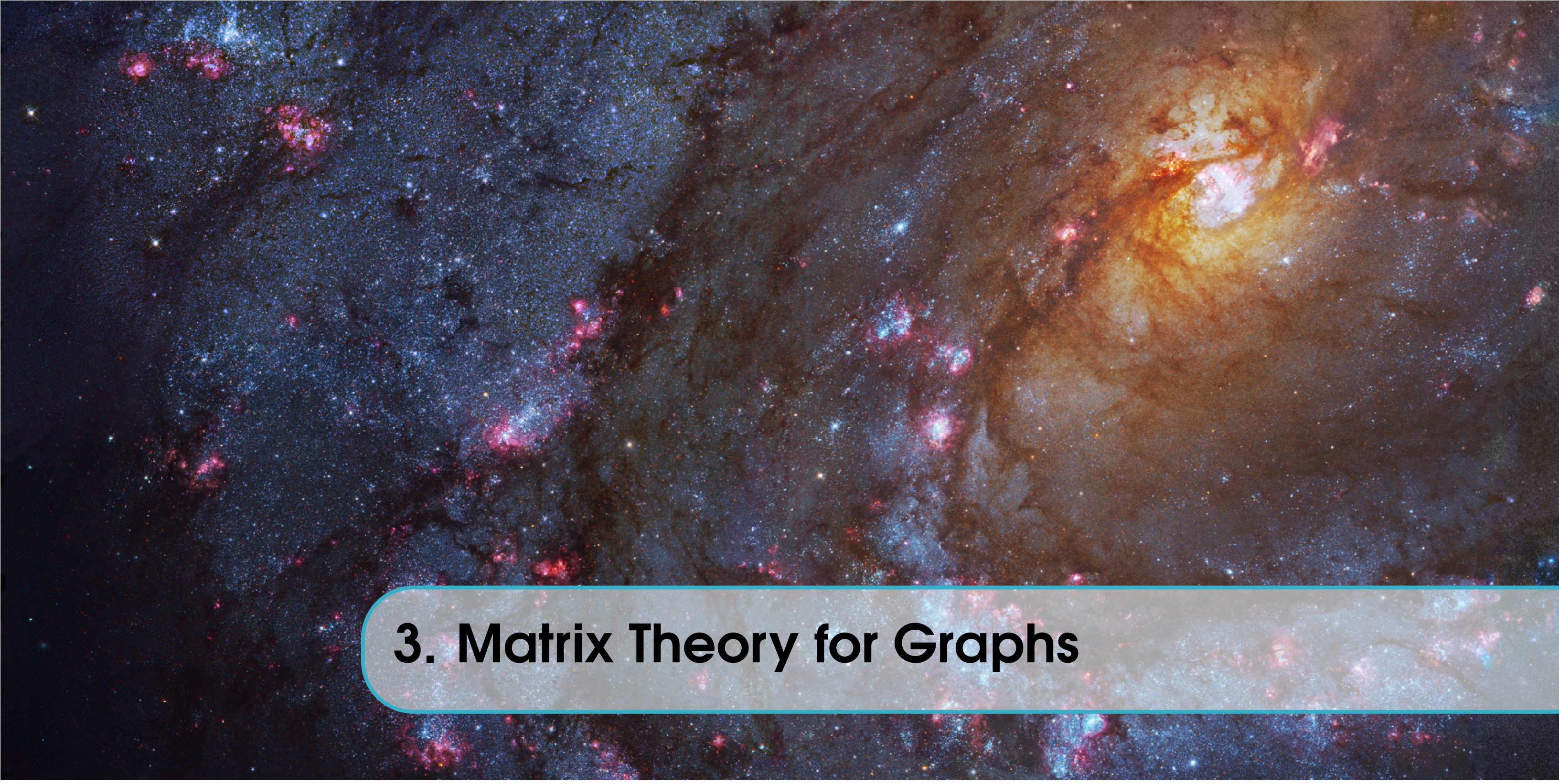
Lemma 2.2.11 强连通的竞赛图$(n \geq 3)$中，如果存在长为$k < n$的圈，则存在长为$k+1$的圈。

Proof. 1. 先考虑是否有点u同时是长为k的圈有关的边的from和to。如果是，那么必然存在$(v_i, u) \in E \wedge (u, v_{i+1}) \in E$, 这一点用之前的套路即可说明。
2. 如果不存在这样的点，那么圈以外的点可以分成两类，一类是圈上的点到该点都是入边，一类是圈上的点到该点都是出边。
由强连通性可知，这两个集合都非空，且任何一个都有指向对方集合的边。那么任意选出圈上的点$v_1 \to v_2 \to v_3$，将其换成$v_1 \to v' \to v'' \to v_3$. ■

Theorem 2.2.12 强连通的竞赛图是哈密顿图。

Proof. Easy by the previous 2 lemmas.
Note that K_2 cannot be strongly connected but other tournament graphs can. ■

Theorem 2.2.13 K_{2n}中有n-1条边不重的哈密顿回路,K_{2n+1}中有n条边不重的哈密顿回路。



3. Matrix Theory for Graphs

以下提到的**graph**指的是无向图，简单起见可以认为没有自环。允许有重边。

3.1 Then Adjacency Matrix

Definition 3.1.1 — adjacency matrix of a directed graph 邻接矩阵. The adjacency matrix $A(X)$ of a directed graph X is the interger matrix with rows and columns indexed by the vertices of X, such that the uv-entry of $A(X)$ is equal to the number of arcs from u to v. If X is simple, then the elements are 0 or 1.

Definition 3.1.2 — adjacency matrix of an undirected graph. We view each edge as a pair of arcs in opposite directions, and $A(X)$ is a symmetric 01-matrix. If the graph has no loops (自环), the diagonal entries of $A(X)$ are 0.

Ⓡ 注意是这里的矩阵应该是标定的(顶点指定编号)。同一个顶点集如果采用不同的方式，那么得到的邻接矩阵不同。但是，这些矩阵之间存在关联。具体而言：

Lemma 3.1.1 Let X and Y be directed graphs on the same vertex set. Then they are isomorphic if and only if there is a permutation matrix P such that $P^T A(X) P = A(Y)$.

Proof. 将邻接矩阵视为双线性函数的度量矩阵。在不同的基下双线性函数的度量矩阵是合同的。在此处基的变换矩阵是permutation matrix. ■

Ⓡ Since permutation matrices are orthogonal, i.e. $P^T = P^{-1}$, and so if X and Y are isomorphic, then $A(X)$ and $A(Y)$ are similar matrices.

Definition 3.1.3 — the spectrum of a matrix 矩阵的谱. The *spectrum* of a matrix is the list of its eigenvalues together with their multiplicities.

Definition 3.1.4 — the spectrum of a graph. The spectrum of a graph X is the spectrum of $A(X)$. 我们也称$A(X)$的特征值和特征向量是图X的特征值和特征向量.

R Lemma 4.1.1 shows that the spectrum (or equivalently, the characteristic polynomial of $A(X)$) is an invariant of the isomorphism class of a graph.
但是两个不同构的图也可以有相同的特征多项式，

下面探讨邻接矩阵提供的更多信息。

Definition 3.1.5 — walk 简单通路. A walk of length r in a directed graph X is a sequence of vertices

$$v_0 \sim v_1 \sim \dots \sim v_r$$

A walk is *closed* if $v_0 = v_r$.

R 注意与path区分。path不允许顶点重复。

Lemma 3.1.2 Let X be a directed graph with adjacency matrix A. The number of walks from u to v in X with length r is $(A^r)_{uv}$.

Proof. Induction by n. Consider the meaning of matrix multiplication. ■

R Lemma 4.1.2 shows that the number of closed walks of length r in X is $tr(A^r)$.

Corollary 3.1.3 Let X be a graph (无自环的无向图) with e edges and t triangles. A : adjacency matrix, then
(1) tr $A = 0$
(2) tr $A^2 = 2e$
(3) tr $A^r = 6t$

3.2 The Incidence Matrix

Definition 3.2.1 — incidence matrix of an undirected graph 无向图的关联矩阵. Let $G = < V, E >$, $V = \{v_1, \dots, v_n\}$, $E = \{e_1, \dots, e_m\}$. The incidence matrix $B(X) \in M_{n\times m}(\mathbb{Z})$, such that

$$b_{ij} = \begin{cases} 1, & if\ v_i \in e_j \\ 0, & otherwise \end{cases}$$

Theorem 3.2.1 Let X be a graph with n vertices and c_0 bipartite connected components 连通的二部图分量. Then rk $B = n - c_0$.

Proof. We shall show that the null space of B has dimension c_0. ■

Lemma 3.2.2 Let B be the incidence matrix of the graph X. Then $BB^T = \Delta(X) + A(X)$, where

$$\Delta(X) = diag\{deg(v_1), ..., deg(v_n)\}$$

Proof. 比较平凡。 ■

Definition 3.2.2 — fundamental incidence matrix 基本关联矩阵. 设v_i为参考点，在$B(X)$中删去第i行得到的矩阵。

3.3 The Incidence Matrix of an Oriented Graph

可以理解为有向图。

An *orientation* of a (无向图)graph X is the assignment of a direction to each edge. Recall that an **arc** of a graph is an ordered pair of adjacent vertices (u, v). To put the definition of orientation formally,

Definition 3.3.1 — orientation of a graph. a function $\sigma : E \to \{-1, 1\}$, where E is the arcs of X. The function satisfies that if (u, v) is an arc, then

$$\sigma(u, v) = -\sigma(v, u).$$

If $\sigma(u, v) = 1$, then we will regard the edge uv as $u \to v$.

Now an *Oriented Graph* is a graph together with a particular orientation. We may use X^σ to denote it.

Definition 3.3.2 — incidence matrix of a directed graph. $D(X^\sigma)$ is the $\{0, 1, -1\}$-matrix with rows indexed by vertices, columns indexed by edges of X.

$$d_{uf} = \begin{cases} -1, & u \to \_ \\ 1, & \_ \to u \\ 0, & else \end{cases}$$

R 注意这个和离散数学书上的定义是反的。

Theorem 3.3.1 If X has c connected components. Let σ be any orientation of X and D is the incidence matrix of X^σ, then rk $D = n - c$.

Proof. Show that the null space of D has dimension c. Suppose $z \in \mathbb{R}^n$ such that $z^T D = 0$. 根据定义验证每一个连通分支中的z_i应该取相同值。 ■

Theorem 3.3.2 $DD^T = \Delta(X) - A(X)$.

Proof. to be written. ■

3.4 The Laplacian of a Graph

Definition 3.4.1 — The Laplacian Matrix 无向图的拉普拉斯矩阵. Let σ be an arbitrary orientation of a graph X, and let D be the incidence matrix of X^{σ}. Then the *Laplacian* of X is the matrix $Q(X) = DD^T$.

Ⓡ From Theorem 4.3.2 we know that Q does not depend on the orientation σ, and hence it is well defined.

Lemma 3.4.1 Let X be a graph with n vertices and c connected components. Then rk $Q = n - c$.

Proof. Reduced to show that for any matrix D, we have

$$rk\ D = rk\ DD^T$$

In fact, let $z \in \mathbb{R}^n$ such that $DD^T z = 0$, then $z^T DD^T z = 0$, which implies $D^T z = 0$. Thus any vector in the null space of DD^T is in the null space of D^T. ■

下面探讨拉普拉斯矩阵的特征值。

Since Q is symmetric, its eigenvalues are real. $\mathbb{R}^n$ has an orthogonal basis consisting of eigenvectors of Q. Since $Q = DD^T$, it is positive semi-definite, and therefore its eigenvalues are all nonnegative. By Lemma 4.4.1, the multiplicity of 0 as eigenvalue of Q is equal to the number of components of X.

Lemma 3.4.2 — 拉普拉斯二次型. Let X be a graph on n vertices with Laplacian Q. Then for any vector x,

$$x^T Qx = \sum_{uv \in E(X)} (x_u - x_v)^2$$

Proof. This follows from the obeservation that

$$x^T Qx = x^T DD^T x = |D^T x|^2$$

If $uv \in E(X)$, then the entry of $D^T x$ corresponding to uv is $\pm(x_u - x_v)$. ■

3.4.1 Trees

这一节介绍矩阵树定理。一个图的生成树的个数可以被拉普拉斯矩阵决定。 首先是一个准备工作。

Definition 3.4.2 — delete an edge. Let X be a graph, and let $e = uv$ be an edge of X. The graph $X - e$ with vertex set $V(X)$ and edge set $E(X)$
e is said to be obtained by **deleting** the edge e.

Definition 3.4.3 — contract an edge. The graph X/e constructed by identifying the vertices u and v and the deleting e is said to be obtained by **contracting** e.

Definition 3.4.4 If M is a symmetric matrix with rows and columns indexed by the set V and if $S \subseteq V$, then let $M[S]$ denote the submatrix obtained by deleting rows and columns indexed by elements of S.

Theorem 3.4.3 — Matrix Tree Theorem. Let X be a graph with Laplacian matrix Q. If u is an arbitrary vertex of X, then $\det Q[\{u\}]$ is equal to the number of spanning trees of X.

Proof. Induction on the number of edges of X.
First, Let $\tau(X)$ be the number of spanning trees of X. we observe that

$$\tau(X) = \tau(X-e) + \tau(X/e)$$

more ... ■

Corollary 3.4.4 The number of spanning trees of K_n is n^{n-2}.

Proof. This follows directly from the fact that $Q[u] = nI_{n-1} - J$ for any vertex u. ■

下面两个定理的证明都需要一些矩阵的技巧。矩阵树定理表明，对角元的余子式(cofactor)都相同，事实上拉普拉斯矩阵的伴随(adjoint)中所有元素都相同。

Theorem 3.4.5 Let $\tau(X)$ denote the number of spanning trees in the graph X and let Q be its Laplacian. Then adj(Q) = $\tau(X) \cdot J$.

Theorem 3.4.6 Let X be a graph on n vertices, and let $\lambda_1 \leq ... \leq \lambda_n$ be the n eigenvalues of Laplacian of X. Then the number of spanning trees in X is $\frac{1}{n}\prod_{i=2}^{n}\lambda_i$.

Proof. The result clearly holds if X is not connected, so we may assume without loss that X is connected.
Let $\phi(t)$ denote the characteristic polynomial $\det(tI - Q)$. Since $\lambda_1 = 0$, the coefficient of t is

$$(-1)^{n-1}\prod_{i=2}^{n}\lambda_i$$

On the other hand, the coefficient of the linear term in $\phi(t)$ is

$$(-1)^{n-1}\sum_{u\in V(X)}\det Q[u]$$

This yields the theorem. ■

4. Matrix Foundations

4.1 线性映射的伴随

4.2 Singular Value Decomposition

设V,W为$\mathbb{R}$上的有限维内积空间，其内积记作$(\cdot|\cdot)_V$ 和 $(\cdot|\cdot)_W$. 设 $m=\dim W$, $n=\dim W$.

Theorem 4.2.1 — SVD. 对于任意线性映射 $T:V\to W$, 记$p=\min\{m,n\}$, 则存在

V 的单位正交基$v_1,\ldots,v_m$,

- W 的单位正交基$w_1,\ldots,w_m$,
- 非负实数$\sigma_1\geq\sigma_2\geq\ldots\geq\sigma_p$,

使得

$$Tv_i=\begin{cases}\sigma_i w_i, & 1\leq i\leq p\\ 0, & i>p\end{cases}$$

此处的$\sigma_1\geq\ldots\geq\sigma_p$ 由T唯一确定，称为T的奇异值。

Proof. to be written ■

Theorem 4.2.2 — 矩阵版本. 设 $V=\mathbb{R}^m, W=\mathbb{R}^n$, 各自配备标准内积，并且将$T$等同于矩阵$A\in M_{n\times m}(\mathbb{R})$. 对于定理5.1.1中的单位正交基，以列向量定义正交矩阵:

$$P:=(v_1|\ldots|v_m)\in M_{m\times m}(\mathbb{R}),\ Q:=(w_1|\ldots|w_n)\in M_{n\times n}(\mathbb{R}),$$

再用奇异值定义

$$\Sigma := \mathrm{diag}\{\sigma_1, \sigma_2, \dots \sigma_r, 0, 0 \dots\} \in M_{n\times m}(\mathbb{R})$$

其中 $r = \mathrm{rank}(T)$. 则奇异值分解化为矩阵等式

$$AP = Q\Sigma$$

亦即

$$A = Q\Sigma P^T$$

4.3 Kronecker Product

Definition 4.3.1 设 $A \in M_{n\times m}(\mathbb{R}), B \in M_{p\times q}(\mathbb{R})$, 则定义

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & & & \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix}$$

R

- 如果 $A = a \in \mathbb{R}$, 那么 $A \otimes B = aB$
- 如果 $B = b \in \mathbb{R}$, 那么 $A \otimes B = bA$

退化为矩阵空间的纯量乘法。

Property 4.3.1 — 混合乘积.

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

Proof. 利用分块矩阵乘法即可验证。 ■

4.4 Matrix Norm

参考click here

4.4.1 Matrix norms induced by vector p-norms

4.4.2 "Entry-wise" matrix norms 矩阵元范数

Frobenius norm

4.4.3 Schatten norms

proof of the triangular inequality is hard for Schatten norms.

4.5 Matrix Derivative

本部分内容可以参考click here

Definition 4.5.1 — 标量函数对矩阵求导. Let $y : M_{m\times n}(\mathbb{R}) \to \mathbb{R}$. Then

$$\frac{\partial y}{\partial X} = \left(\frac{\partial y}{\partial x_{ij}}\right)_{ij} \in M_{m\times n}(\mathbb{R})$$

Ⓡ 这里和传统的梯度定义的行列顺序是反的。

Definition 4.5.2 — 因变量是矩阵，自变量是标量. Let $y : \mathbb{R} \to M_{m\times n}(\mathbb{R})$, Then

$$\frac{\partial Y}{\partial x} = \left(\frac{\partial Y_{ij}}{\partial x}\right)_{ij} \in M_{m\times n}(\mathbb{R})$$

■ **Example 4.1** 求 $y(X) = ||X||_F^2$ 的导数。 ■

■ **Example 4.2** 求 $y(X) = tr(X)$ 的导数。 ■

下面定义矩阵对矩阵的导数，这个定义并不是统一的。

Definition 4.5.3 设 X, Y 为矩阵，定义

$$\frac{\partial Y}{\partial X} = \left(\frac{\partial}{\partial x_{ij}}\right) \otimes Y$$

Ⓡ 如果 X，Y之一为标量，根据5.3.1的remark，这个定义等同于5.4.1和5.4.2.

Property 4.5.1 如果自变量为标量t.

- $\frac{\partial (X+Y)}{\partial t} = \frac{\partial X}{\partial t} + \frac{\partial Y}{\partial t}$
- $\frac{\partial X\otimes Y}{\partial t} = \frac{\partial X}{\partial t} \otimes Y + \frac{\partial Y}{\partial t} \otimes X$
- $\frac{\partial XY}{\partial t} = \frac{\partial X}{\partial t} Y + \frac{\partial Y}{\partial t} X$

Proof. 后两条归结为标量函数乘积的导数。 ■

Property 4.5.2 如果自变量为向量x。

- $\frac{\partial (a^T x)}{\partial x} = a$
- $\frac{\partial (x^T Ax)}{\partial x} = (A + A^T)x$
- $\frac{\partial (Ax)^T}{\partial x} = A^T$

Proof. 第二条：Let $y = x^T Ax = \sum_{i,j} a_{ij} x_i x_j$, then $\frac{\partial y}{\partial x_i} = 2a_{ii}x_i + \sum_{j\neq i} a_{ij} x_j + \sum_{j\neq i} a_{ji} x_j$.
第三条：考虑结果的第i行$\frac{\partial (Ax)^T}{\partial x_i}$，设$A = (a_1|a_2|...|a_n)$，则 $Ax = (x_1a_1 + ... + x_na_n)$. 所以 $\frac{\partial (Ax)^T}{\partial x_i} = a_i^T$ ■

Corollary 4.5.3 $\frac{\partial (x^T A)}{\partial x} = A$. x可以退化成标量。

Property 4.5.4 自变量为矩阵X.
$\frac{\partial \mathrm{tr}(AX)}{\partial X} = A^T$

- *Proof.* Asume $A \in M_{M\times n}(\mathbb{R})$ Let $y = tr(AX) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_{ji}$, which easily implies the above identity. ■

4.5.1 chain rule

to be written

■ **Example 4.3** 计算 $y = ||BX - C||_F^2$ 关于X的导数。 ■

Proof. Note that

$$y = tr\left((BX-C)(BX-C)^T\right) = tr\left((BX-C)(X^TB^T-C^T)\right) = tr\left(BXX^TB^T\right) - tr\left(BXC^T\right) - tr\left(CX^TB^T\right) + tr\left(CC^T\right)$$

First, we have $tr\left(BXC^T\right) = tr\left(CX^TB^T\right) = tr\left(C^TBX\right)$, and

$$\frac{\partial (C^TBX)}{\partial X} = B^TC$$

Consider that

$$tr\left(B^TB(X+\varepsilon Y)(X+\varepsilon Y)^T\right) = tr\left(B^TB(XX^T)\right) + \varepsilon\, tr\left(B^TBXY^T\right) + \varepsilon\, tr\left(B^TBYX^T\right) + O(\varepsilon)$$

Thus,

$$\lim_{\varepsilon\to 0} \frac{\Delta tr(B^TBXX^T)}{\varepsilon} = tr(YX^TB^TB) + tr(X^TB^TBY) = tr\left((X^TB^TB + X^TB^TB)Y\right)$$

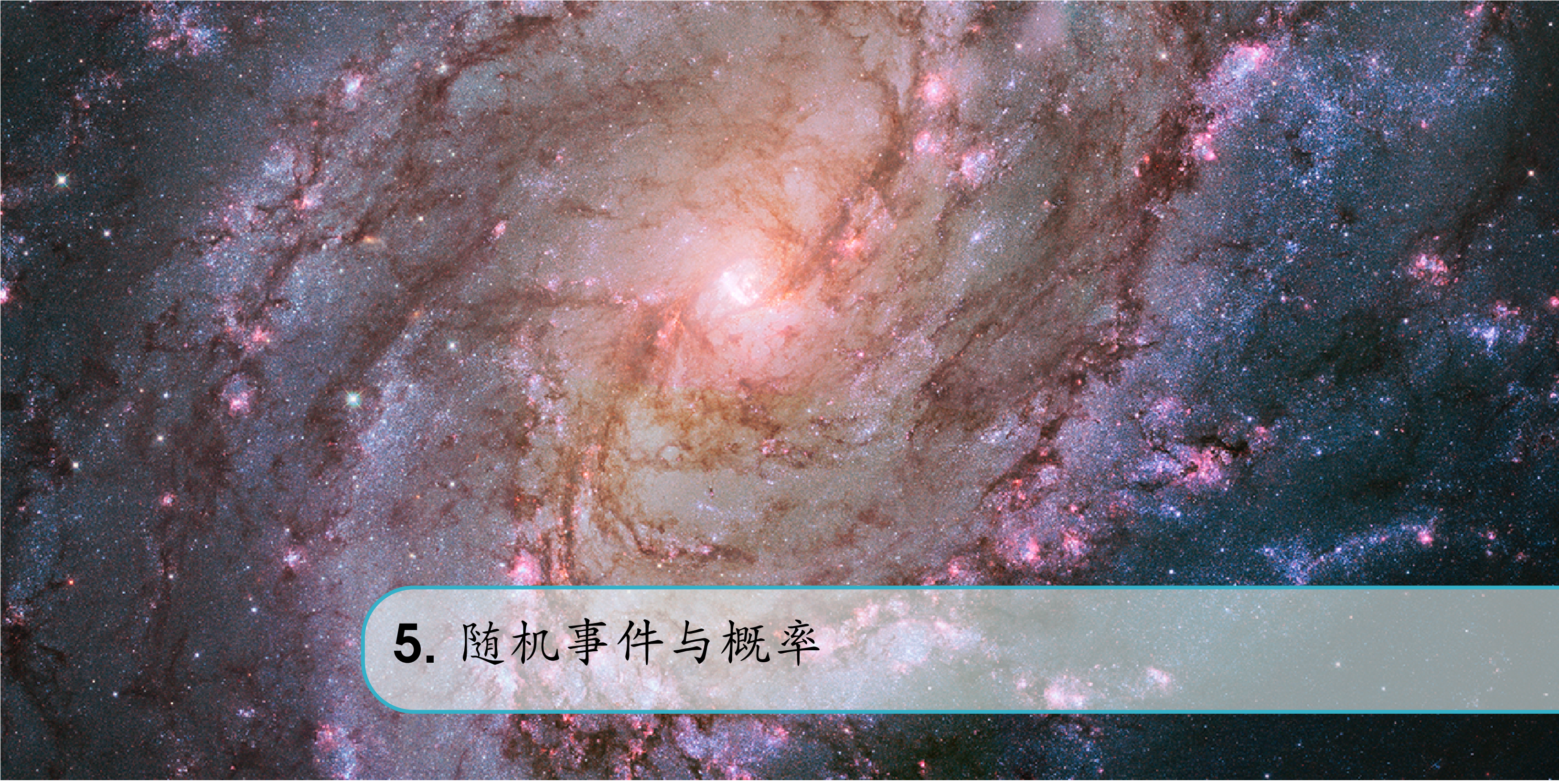
Then,

$$\frac{\partial tr\left(BXX^TB^T\right)}{\partial X} = \frac{\partial 2\cdot tr\left(X^TB^TBY\right)}{\partial Y} = 2\cdot B^TBX$$

To sum up,

$$\frac{\partial y}{\partial X} = 2(B^TBX - B^TC)$$

■



5. 随机事件与概率

5.1 Kolmogorov 公理系统

Let Ω be any non-empty set, called 基本事件空间，$\mathscr{P}(\Omega)$表示Ω的幂集

Definition 5.1.1 Let $\mathscr{F} \subset \mathscr{P}(\Omega)$. $P = P(\cdot) : \mathscr{F} \to R$, 如果满足
(1) $\Omega \in \mathscr{F}$
(2) $A \in \mathscr{F} \implies A^c = \Omega - A \in \mathscr{F}$
(3) $A_n \in \mathscr{F}(n = 1, 2.., .)$, then $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$
(4) $\forall A, P(A) \geq 0$
(5) $P(\Omega) = 1$
(6) If $A_i (i = 1, 2, ...)$ are disjoint, $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
资料$(\Omega, \mathscr{F}, P)$称为概率空间。P是$\mathscr{F}$上概率测度。对一个事件(集合)A而言，$P(A)$为A的概率。

Definition 5.1.2 — $\sigma - algebra$. Let $\mathscr{F} \subset \mathscr{P}(\Omega)$, 如果满足
(1) $\Omega \in \mathscr{F}$
(2) $A \in \mathscr{F} \implies A^c = \Omega - A \in \mathscr{F}$
(3) $A_n \in \mathscr{F}(n = 1, 2.., .)$, then $\cup_{i=1}^{\infty} A_i \in \mathscr{F}$
则称$\mathscr{F}$为Ω中的σ域或者σ代数。

Property 5.1.1 Let $\mathscr{F}$ be a σ-algebra of Ω, we have
(1) If $A_i \in \mathscr{F}$, then $\cup_{i=1}^{n} A_i \in \mathscr{F}$. $\cap_{i=1}^{n} A_i \in \mathscr{F}$
(2) If $A_i \in \mathscr{F}$, then $\cap_{i=1}^{\infty} \in \mathscr{F}$
(3) If $A, B \in \mathscr{F}$, then $A - B \in \mathscr{F}$

Proof. 归结为集合运算的性质。比如 $(\cap_{i=1}^{\infty} A_i)^c = \cup_{i=1}^{\infty} A_i^c$ ∎