Дискретные структуры

осень 2013

Александр Дайняк

www.dainiak.com

Гиперграфы

Гипеграф — это пара (V, E), где

- V — конечное множество *вершин*,
- E — набор непустых подмножеств V — (*гипер*-)*рёбра*.

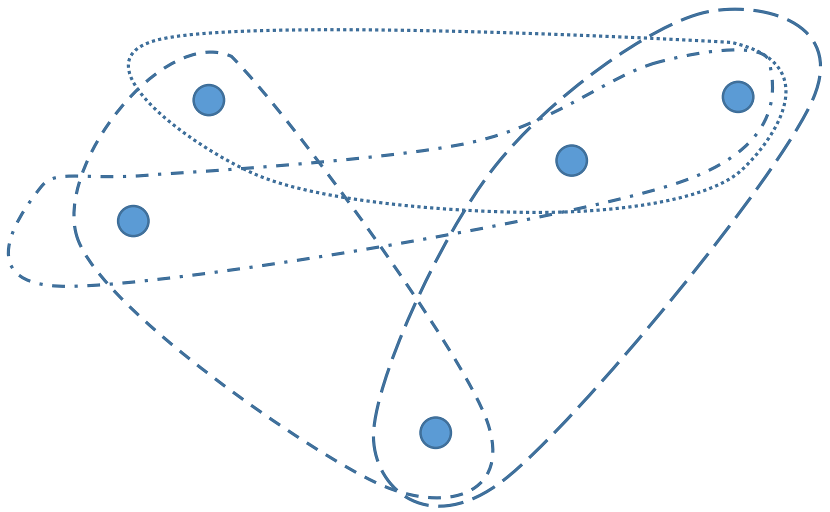
Т.е. гиперграф в общем случае — это просто набор/семейство/совокупность непустых подмножеств конечного множества.

Гиперграфы

Гиперграф h-*однородный*, если $\forall e \in E \quad |e| = h$.

Так что граф — это 2-однородный гиперграф.

Пример 3-однородного гиперграфа:



Аналоги теоретико-графовых понятий

Полный гиперграф — тот, в котором есть все возможные (в данном контексте) рёбра.

Например, полный h-однородный гиперграф — тот, в котором
$$E = \{V' \subseteq V \mid |V'| = h\}$$

Независимое множество — это такое $A \subseteq V$, что $\nexists e \in E\colon \quad e \subseteq A$.

Правильная вершинная раскраска — такая, при которой совокупности вершин одного цвета образуют независимые множества.

Аналоги теоретико-графовых понятий

Цепь в гиперграфе (V, E) — это последовательность вершин и рёбер

$$v_1 e_1 v_2 e_2 \ldots e_s v_{s+1}$$

такая, что

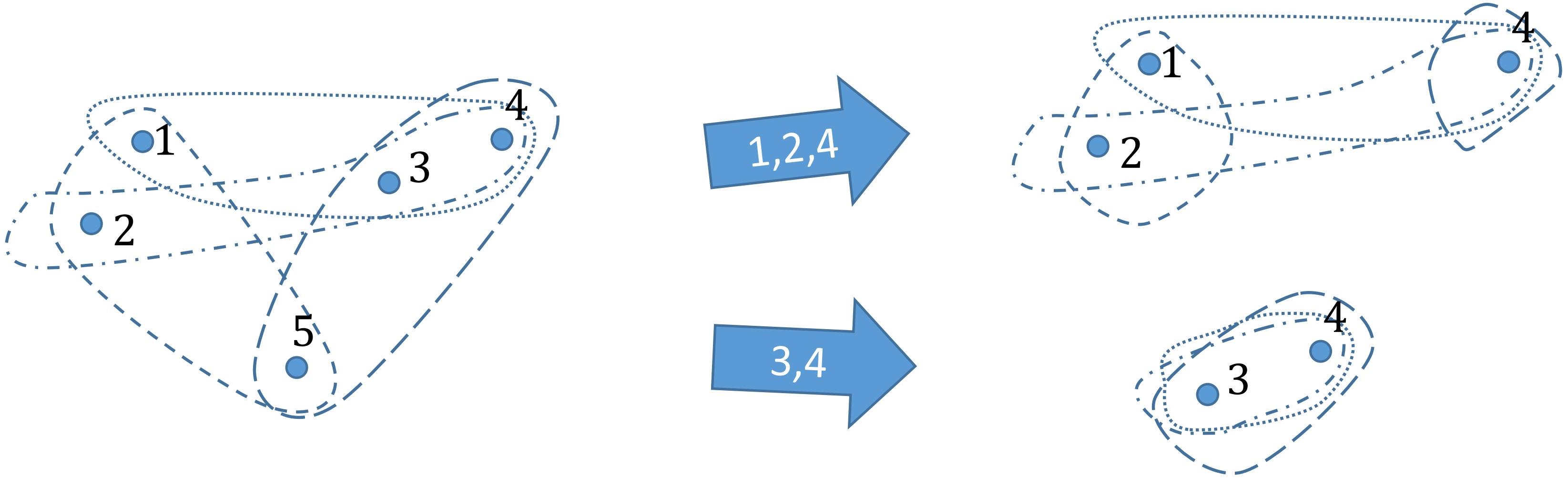
- $\forall i (v_i, v_{i+1} \in e_i)$,
- все v_i различны,
- все e_i различны (не как множества, а как элементы E).

Цикл определяется аналогично, с той разницей, что v_1 и v_{s+1} совпадают.

Аналоги теоретико-графовых понятий

Подгиперграф гиперграфа (V, E), *порождённый множеством вершин* V' — это пара (V', E'), где

$$E' \coloneqq \{e \cap V' \mid e \in E \text{ и } e \cap V' \neq \emptyset\}$$



Аналоги теоретико-графовых понятий

Область связности гиперграфа (V, E) — это такое множество $V' \subseteq V$, что

- $\forall u, v \in V'$ найдётся цепь из u в v,
- к V' нельзя добавить ни одной вершины, так, чтобы предыдущее свойство сохранилось.

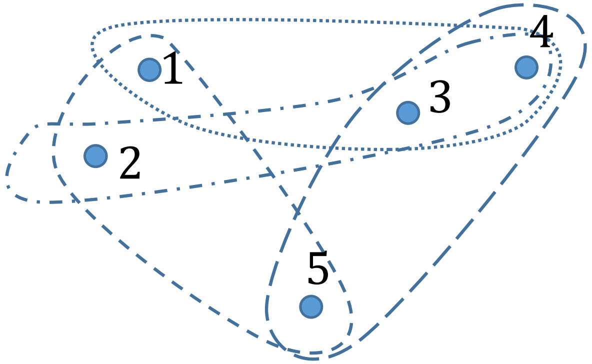
Компонента связности — это подгиперграф, порождённый областью связности.

Матрица инцидентности

Матрица инцидентности гиперграфа (V, E) — это матрица $\left(a_{v,e}\right)_{v\in V, e\in E} \in \{0,1\}^{|V|\times|E|}$, где

$$a_{v,e} := \begin{cases} 1, \text{если } v \text{ и } e \text{ инцидентны} \\ 0, \text{иначе} \end{cases}$$

Пример:



$$\Rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Покрытия

Если вершина v гиперграфа входит в ребро e, то говорим, что v *покрывает* e или *протыкает* e.

Вершинное покрытие гиперграфа — это такое множество вершин $A \subseteq V$, что

$$\forall e \in E \quad e \cap A \neq \emptyset$$

Вершинное покрытие называется ещё

- *протыкающим множеством,*
- *трансверсалью,*
- *системой общих представителей (с.о.п.)*

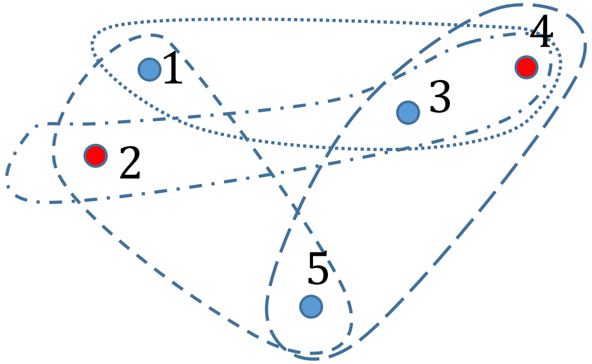
для семейства множеств E.

Покрытия в терминах матриц

На языке матриц инцидентности, строка *покрывает/протыкает* столбец, если у них на пересечении стоит 1.

Покрытие матрицы — это такое подмножество строк, что каждый столбец покрывается одной из этих строк.

Пример:



$$\begin{pmatrix} 1 & 0 & 1 & 0 \\ \color{red}{0} & \color{red}{1} & \color{red}{1} & \color{red}{0} \\ 1 & 1 & 0 & 1 \\ \color{red}{1} & \color{red}{1} & \color{red}{0} & \color{red}{1} \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

Покрытия в терминах матриц

Число трансверсальности гиперграфа — это минимальный размер покрытия.

Формально, если $H = (V, E)$, то

$$\tau(H) \coloneqq \min\{|A| \quad | \quad \forall e \in E \quad A \cap e \neq \emptyset\}$$

Будем также писать $\tau(E)$.

Аналогично, для любой булевой матрицы M без нулевых столбцов через $\tau(M)$ будем обозначать мощность минимального покрытия (называемую также *глубиной* матрицы).

Системы общих представителей

Пример прикладной задачи.

Есть набор экспертов $V \coloneqq \{v_1, \ldots, v_n\}$. Каждый эксперт, может быть, разбирается не только в одной области.

Пусть $e_1, \ldots, e_m \subseteq V$ — множества экспертов, разбирающихся в 1-й, …, m-й области знания.

Как набрать команду экспертов для решения задачи, требующей владения всеми указанными областями знания?

Жадный алгоритм построения с.о.п.

Для гиперграфа $H \coloneqq (V, E)$ рассмотрим алгоритм:

1. $S \coloneqq \emptyset$
2. $E_{\text{notCovered}} \coloneqq \{e \in E \mid e \cap S = \emptyset\}$
3. if $|E_{\text{notCovered}}| > 0$:
4. $\quad v^* \coloneqq \underset{v \in V}{\operatorname{argmax}} \#\{e \in E_{\text{notCovered}} \mid e \ni v\}$
5. $\quad S \coloneqq S \cup \{v^*\}$
6. $\quad$ goto 2.
7. S — искомая с.о.п.

То есть на каждом шаге добавляем в S любую из вершин, покрывающих наибольшее число из ещё не покрытых рёбер.

Жадный алгоритм построения покрытия матрицы

Для матрицы M рассмотрим алгоритм:

1. $S \coloneqq \emptyset$
2. $C \coloneqq \text{столбцы, непокрытые строками из } S$
3. if $|C| > 0$:
4. $\quad r^* \coloneqq \underset{r\text{ — строка } M}{\operatorname{argmax}} \#\{\text{столбцы из } C, \text{покрываемые } r\}$
5. $\quad S \coloneqq S \cup \{r^*\}$
6. $\quad$ goto 2.
7. S — искомое покрытие матрицы

То есть на каждом шаге добавляем в S любую из строк, покрывающих наибольшее число из ещё не покрытых столбцов.

Теорема о мощности жадного покрытия

Теорема.
Пусть в каждом столбце матрицы $M \in \{0,1\}^{n\times m}$ не менее h единиц, и при этом $mh > n$.

Тогда мощность покрытия, построенного ж.а.,

$$\leq 1 + \frac{n}{h} + \frac{n}{h}\ln\frac{mh}{n}$$

Следствие.
Для таких матриц $\tau(M) \leq 1 + \frac{n}{h} + \frac{n}{h}\ln\frac{mh}{n}$.

Доказательство теоремы о мощности жадного покрытия

Пусть уже сделано k шагов алгоритма, в результате чего $|S| = k$ и остаются непокрытыми $c_k m$ столбцов, где $0 < c_k \leq 1$.

Рассмотрим матрицу M_k, образованную строками из $\bar{S}$ и непокрытыми столбцами.

Имеем $M_k \in \{0,1\}^{(n-k)\times(c_k m)}$.

В каждом столбце M_k не менее h единиц, а значит, всего в M_k не менее $c_k m h$ единиц.

Значит, в M_k есть строка, в которой $\geq \frac{c_k m h}{n-k}$ единиц.

Доказательство теоремы о мощности жадного покрытия

Пусть уже сделано k шагов алгоритма, в результате чего $|S| = k$ и остаются непокрытыми $c_k m$ столбцов, где $0 < c_k \leq 1$.

Есть строка, в которой $\geq \frac{c_k mh}{n-k}$ единиц.

Следовательно, ж.а. выберет строку, покрывающую не менее $\frac{c_k mh}{n-k}$ новых столбцов.

После $(k+1)$-го шага непокрытыми останутся

$$\leq c_k m - \frac{c_k mh}{n-k} = c_k\left(1 - \frac{h}{n-k}\right) \cdot m$$

столбцов, то есть

$$c_{k+1} \leq c_k\left(1 - \frac{h}{n-k}\right) \leq c_k\left(1 - \frac{h}{n}\right)$$

Доказательство теоремы о мощности жадного покрытия

Пусть после k шагов алгоритма остаются непокрытыми $c_k m$ столбцов.

Имеем $c_{k+1} \leq c_k\left(1 - \frac{h}{n}\right)$ и $c_0 = 1.$

Отсюда $c_k \leq \left(1 - \frac{h}{n}\right)^k$.

Пусть ж.а. выполнил $k' \coloneqq \left\lceil \frac{n}{h} \ln \frac{mh}{n} \right\rceil$ шагов.

Имеем

$$c_{k'} \leq \left(1 - \frac{h}{n}\right)^{\frac{n}{h} \ln \frac{mh}{n}} = e^{\frac{n}{h} \cdot \ln \frac{mh}{n} \cdot \ln\left(1 - \frac{h}{n}\right)}$$

Доказательство теоремы о мощности жадного покрытия

После $k' \coloneqq \left\lceil \frac{n}{h} \ln \frac{mh}{n} \right\rceil$ шагов,
с учётом неравенства $\ln(1-x) < -x$, имеем

$$c_{k'} \leq e^{\frac{n}{h} \cdot \ln \frac{mh}{n} \cdot \ln\left(1 - \frac{h}{n}\right)} < e^{-\ln \frac{mh}{n}} = \frac{n}{mh}$$

После k'-го шага остаются непокрытыми

$$c_{k'} \cdot m < \frac{n}{h}$$

столбцов. Даже если на покрытие каждого из них потребуется по одному шагу, общее число шагов алгоритма будет

$$\leq k' + \frac{n}{h} \leq 1 + \frac{n}{h} + \frac{n}{h} \ln \frac{mh}{n}$$

Труднопокрываемые матрицы

Теорема о мощности «жадного» покрытия.
Пусть в каждом столбце матрицы $M \in \{0,1\}^{n\times m}$ не менее h единиц, и при этом $mh > n$. Тогда у M мощность «жадного» покрытия $\leq 1 + \frac{n}{h} + \frac{n}{h}\ln\frac{mh}{n}$.

Теорема о существовании «труднопокрываемых» матриц (явная конструкция).
Пусть $m, n, h \in \mathbb{N}$ таковы, что $2 \leq \ln\frac{mh}{n} \leq h \leq \frac{n}{8}$.

Тогда найдётся матрица $M \in \{0,1\}^{n\times m}$, каждый столбец которой содержит $\geq h$ единиц, и мощность минимального покрытия которой $\geq \frac{1}{16}\cdot\frac{n}{h}\ln\frac{mh}{n}$.

Построение труднопокрываемой матрицы

Рассмотрим матрицу $M' \in \{0,1\}^{2a \times \binom{2a}{a}}$, состоящую из всех столбцов высоты $2a$, в каждом из которых ровно a единиц.

Имеем $\tau(M') = a + 1$.

Положим

$$M'' := \begin{pmatrix} M' \\ \vdots \\ M' \end{pmatrix} \in \{0,1\}^{2ab \times \binom{2a}{a}}$$

Получаем $\tau(M'') = \tau(M') = (a + 1)$, в каждом столбце M'' ровно ab единиц.

Построение труднопокрываемой матрицы

Для матрицы $M'' \in \{0,1\}^{2ab\times\binom{2a}{a}}$ имеем

$$\tau(M'') = a + 1$$

Рассмотрим матрицу

$$M''' \coloneqq \begin{pmatrix} M'' & 0 & \dots & 0 \\ 0 & M'' & \dots & 0 \\ \vdots & & \ddots & \\ 0 & \dots & 0 & M'' \end{pmatrix} \in \{0,1\}^{2abc\times\binom{2a}{a}c}$$

Имеем $\tau(M''') = c \cdot \tau(M'') = (a+1)c$,
и в каждом столбце M''' ровно ab единиц.

Построение труднопокрываемой матрицы

Итак, для любых $a, b, c \in \mathbb{N}$ существует матрица $M''' \in \{0,1\}^{2abc \times \binom{2a}{a} c}$,
в каждом столбце которой ab единиц, и для которой $\tau(M''') = (a+1)c$.

По условию, $2 \le \ln \frac{mh}{n} \le h \le \frac{n}{8}$.

Пусть $a \coloneqq \left\lfloor \frac{1}{2} \ln \frac{mh}{n} \right\rfloor$, $b \coloneqq \left\lfloor \frac{2h}{a} \right\rfloor$ и $c \coloneqq \left\lfloor \frac{n}{2ab} \right\rfloor$. Тогда в M'''

- $\#\text{строк} = 2ab \cdot \left\lfloor \frac{n}{2ab} \right\rfloor \le n$
- $\#\text{столбцов} < 4^a \cdot c \le 2^{\ln \frac{mh}{n}} \cdot \frac{n}{2a \cdot \lfloor 2h/a \rfloor} < \frac{mh}{n} \cdot \frac{n}{2a \cdot h/a} < m$
- $\#\text{"1" в столбце} = a \cdot \left\lfloor \frac{2h}{a} \right\rfloor \ge a \cdot \frac{h}{a} \ge h$
- $\tau(M''') \ge \frac{1}{2} \ln \frac{mh}{n} \cdot \left\lfloor \frac{n}{2ab} \right\rfloor \ge \frac{1}{2} \ln \frac{mh}{n} \cdot \left\lfloor \frac{n}{4h} \right\rfloor \ge \frac{1}{16} \cdot \frac{n}{h} \ln \frac{mh}{n}$

Построение труднопокрываемой матрицы

Итак, при $2 \leq \ln\frac{mh}{n} \leq h \leq \frac{n}{8}$ найдётся матрица M''', такая, что

- $\#\text{строк} \leq n, \quad \#\text{столбцов} \leq m$
- $\#"1" \text{ в столбце} \geq h$
- $\tau(M''') \geq \frac{1}{16} \cdot \frac{n}{h} \ln\frac{mh}{n}$

Дополним M''' нулевыми строками, а затем единичными столбцами, до матрицы размера $n \times m$.

В каждом столбце полученной матрицы M не менее h единиц, и $\tau(M) = \tau(M''') \geq \frac{1}{16} \cdot \frac{n}{h} \ln\frac{mh}{n}$, что и требовалось.

Соотношение мощностей жадного и оптимального покрытий

Теорема (без д-ва).
Пусть M — произвольная матрица, в каждой строке которой не более k единиц. Тогда покрытие, построенное ж.а., имеет размер *не более* $(1 + \ln k) \cdot \tau(M)$.

Теорема (без д-ва).
Для любого $k \geq 2$ существует матрица M, в каждой строке которой не более k единиц, а покрытие, построенное для M с помощью ж.а., имеет размер *не менее* $\frac{(\log_2 k)-1}{2} \cdot \tau(M)$.