Сравнительный анализ методов оценивания бикликового покрытия

Попов Максим
6 Курс ФИВТ
Кафедра: Анализ данных

23 июня 2016

Основные результаты

1) Сравнил методы оценивания на графах, построенных при помощи геометрических конфигураций;
2) Доказал, что метод трудного множества для произвольного графа дает оценки не хуже метода Куликова-Юкны;
3) С помощью метода случайных графов показал, что техника трудного множества может давать оценки значительно сильнее, чем метод Куликова-Юкны;
4) Получил обобщение метода информационных неравенств для задачи коммуникационной сложности с $m > 2$ участниками.

Определения и понятия

Определение

Бикликой неориенторованного графа называется подмножество его вершин, образующих полный двудольный подграф.

Определение

Бикликовым покрытием $bcc(G)$ графа G будем называть наименьшее число, возможно, пересекающихся биклик, которыми можно покрыть все ребра графа G.

Мотивировка

1) Сам по себе вопрос бикликового покрытия вполне естественен в теории графов и играет центральную роль во многих вычислительных задачах.
2) Бикликовое покрытие играет важную роль в коммуникационной сложности. Для детерминированной коммуникационной сложности $CC(f)$ верно

$$CC(f) \geq \log_2 \left(\sum_{z \in Z} bcc(G_z) \right)$$

Для недетерминированной коммуникационной сложности $NCC(f)$ верно

$$\left\lceil \log_2 \left(\sum_{z \in Z} bcc(G_z) \right) \right\rceil + 1 \geq NCC(f) \geq \log_2 \left(\sum_{z \in Z} bcc(G_z) \right)$$

Метод трудного множества

Определение

Пусть $G = (V, E)$ произвольный неориентированный граф. Будем называть подмножество ребер $S \subseteq E$ трудным, если для любых двух различных ребер $(x, y) \in S$ и $(x', y') \in S$ имеем $(x', y) \notin E$ или $(x, y') \notin E$. Размер максимального трудного множества будем обозначать через $fool(G)$.

Теорема

Если $S \subseteq E$ трудное множество графа G, то $bcc(G) \geq |S|$. В частности, наилучшая оценка по методу трудного множества:

$$bcc(G) \geq fool(G).$$

Метод Куликова-Юкны

Данный метод был описан в статье "Jukna S., Kulikov A. S. On covering graphs by complete bipartite subgraphs. 2009".

Теорема

Для произвольного неориентированного графа $G = (V, E)$ верно:

$$bcc(G) \geq \frac{\nu(G)^2}{|E|},$$

где $\nu(G)$ – размер максимального паросочетания графа G.

Метод информационных неравенств

Данный метод был описан в статье "Kaced T., Romashchenko A.E., Vereshchagin N.K. Conditional Information Inequalities and Combinatorial Applications. 2015".

Теорема

Пусть ребра двудольного графа $G = (L, R, E)$ раскрашены по следующему правилу:

(*) *для произвольной биклики $C \subseteq G$ и для произвольной пары ребер (x, y') и (x', y) из C одного цвета a, цвет ребра (x, y) тоже a.*

Пусть на ребрах задано произвольное вероятностное распределение. Определим случайные величины (X, Y, A) следующим образом:

- $X =$ [*левый конец ребра*]; $Y =$ [*правый конец ребра*];
- $A =$ [*цвет ребра*].

Тогда $bcc(G) \geq 2^{\frac{1}{2}(H(A|X)+H(A|Y)-H(A))}$.

Основные теоремы

Теорема

Пусть имеется произвольный неориентированный граф $G = (V, E)$, тогда среди ребер максимального паросочетания можно найти трудное множество размера по крайней мере $\frac{\nu(G)^2}{|E|}$.

Теорема

Для произвольных $\alpha \in [0, \frac{1}{3})$ и $\beta \in (\alpha, \frac{1+\alpha}{2})$ при достаточно больших n существует двудольный граф $G = (L, R, E)$ такой, что $|L| = |R| = n$, на котором метод трудного множества дает оценку хотя бы n^β, а оценка Куликова-Юкны не превосходит $n^\alpha + o(1)$.

Обобщение информационного метода

Теорема

Пусть ребра гиперграфа $G = (X_1, X_2, \ldots, X_m, E)$ *раскрашены по следующему правилу:*

(*) *для произвольного полного* m*-дольного гиперграфа* $C \subseteq G$ *и для произвольного набора ребер* $(x_{1,1}, \ldots, x_{1,m}), \ldots, (x_{m,1}, \ldots, x_{m,m})$ *одного цвета a, цвет ребра* $(x_{1,1}, x_{2,2}, \ldots, x_{m,m})$ *тоже a.*

Пусть на ребрах задано произвольное вероятностное распределение. Определим случайные величины $(X_1, \ldots, X_m, A)$ *следующим образом:*

- $X_i =$ [i*-ая вершина ребра*];
- $A =$ [*цвет ребра*].

Тогда выполняется неравенство:

$$bcc(G) \geq 2^{\frac{1}{m}(H(A|X_1)+\ldots+H(A|X_m)-(m-1)H(A))}$$

Используемые теоремы и техники

1) При доказательстве первой теоремы была придумана конструкция графа четырехсторонников $\widetilde{G}$. Доказано, что $fool(G) = w(\widetilde{G})$, $\max\limits_{K_{r,s} \subseteq G} \{r \cdot s\} = \alpha(\widetilde{G})$ и $bcc(G) = \chi(\widetilde{G})$. Окончательное утверждение теоремы получается применением теоремы Турана к графу $\widetilde{G}$.
2) Во второй теореме использовались случайные графы Эрдеша-Реньи. При оценивании максимального паросочетания использовалась лемма Холла, а при оценивании количества ребер использовалось неравенство Хефдинга.
3) В последней теореме использовалось не Шенноновское информационное неравенство:

$$H(A|X_1, B) + H(A|X_2, B) + \ldots + H(A|X_m, B) \leq (m-1)H(A|B)$$

Открытые вопросы

1) Не удалось доказать, что метод трудных множеств работает почти наверное намного лучше, чем оценка Куликова-Юкны. А именно не получилось доказать

$$\sum_{T:T\sim S_0} P\{I_k(T)=1 \mid I_k(S_0)=1\} = o(\mathbb{E}[f_k(G)])$$

2) Не получилось сравнить в общем случае метод трудных множеств с методом информационных неравенств.

Спасибо за внимание!