Planar Matching in Streams Revisited

Sofya Vorotnikova

University of Massachusetts Amherst

Joint work with Andrew McGregor

Problem Description

Given a large sparse graph, approximate the size of maximum matching using little space. In particular, consider graphs with arboricity α.

Arboricity is the minimum number of forests into which the edges of the graph can be partitioned.
Property: In a graph with arboricity α, every induced subgraph on r vertices has at most αr edges.
Planar graphs have arboricity at most 3.

Note, that we are not computing the matching itself, just its *size*.

Results

Maximum cardinality matching in unweighted graphs:

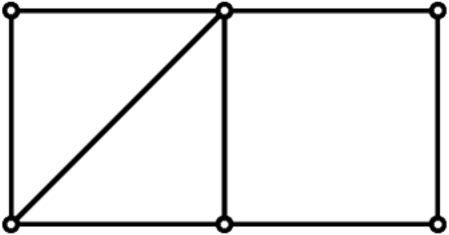
- $(\alpha + 2)$-approximation (improvement over Esfandiari et al. $(5\alpha + 9)$-approximation) using space
  - $\tilde{O}(\alpha n^{2/3})$ in insert-only arbitrary order streams
  - $\tilde{O}(\alpha n^{4/5})$ in insert-delete arbitrary order streams
- $\frac{(\alpha+2)^2}{2}$-approximation using the degree distribution: $O(\log(n))$ space in adjacency list streams.

Maximum weight matching in weighted graphs:

- c-approximation in unweighted graphs $\Rightarrow$ $2c$-approximation in weighted.

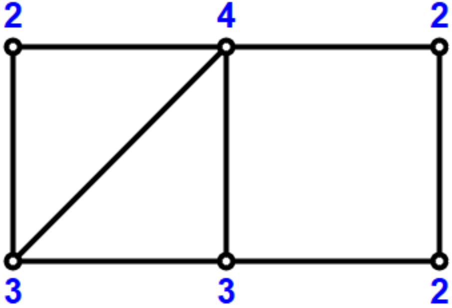
Local Fractional Matchings

$$x_e = \min\left(\frac{1}{deg(u)}, \frac{1}{deg(v)}, \frac{1}{\alpha + 1}\right)$$



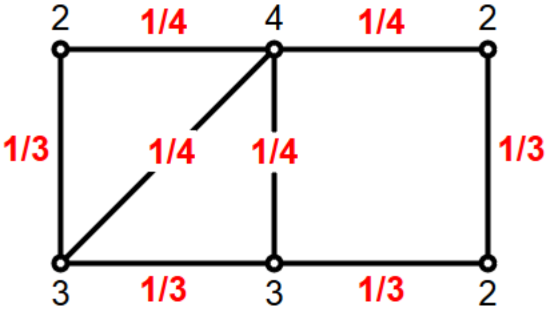
Local Fractional Matchings

$$x_e = \min\left(\frac{1}{deg(u)}, \frac{1}{deg(v)}, \frac{1}{\alpha + 1}\right)$$



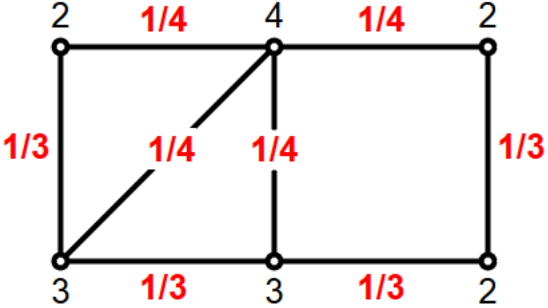
Local Fractional Matchings

$$x_e = \min\left(\frac{1}{deg(u)}, \frac{1}{deg(v)}, \frac{1}{\alpha+1}\right)$$



Local Fractional Matchings

$$x_e = \min\left(\frac{1}{deg(u)}, \frac{1}{deg(v)}, \frac{1}{\alpha+1}\right)$$



- $\mathbf{x}$ is a fractional matching
- x_e is only a function of edges that share an endpoint with e
- $(\alpha+1)\sum_e x_e$ is an $(\alpha+2)$-approx of match(G)

Proof Outline

- $\mathbf{x}$ is a fractional matching
- $\sum_e x_e \leq \frac{\alpha+2}{\alpha+1} \text{match}(G)$
- $\sum_e x_e \geq \frac{\text{match}(G)}{\alpha+1}$

Proof Part 1: Fractional Matching

- From definition, $x_e \geq 0$ for each $e \in E$
- Note that for any $v \in V$,

$$\sum_{e \in E : v \in e} x_e \leq \sum_{e \in E : v \in e} \frac{1}{\deg(v)} = 1$$

Thus, $\mathbf{x}$ is a fractional matching.

Proof Part 2: Upper Bound

Fun fact:

- From Edmond's theorem, for any fractional matching FM,

$$FM \leq 1.5\,\text{match}(G)$$

Another fun fact:

- For any fractional matching FM with $x_e \leq \epsilon$ for all e,

$$FM \leq (1+\epsilon)\,\text{match}(G)$$

For our fractional matching, all $x_e \leq 1/(\alpha+1)$ and hence

$$\sum_e x_e \leq (1 + \frac{1}{\alpha+1})\,\text{match}(G) = \frac{\alpha+2}{\alpha+1}\,\text{match}(G)$$

Proof Part 3: Lower Bound

Let

- H be the set of *heavy* vertices of degree at least $\alpha + 2$
- E_L be the set of *light* edges with both endpoints not heavy

Weight of light edges $|E_L|/(\alpha + 1)$.

Weight of edges incident to H is at least $|H|/(\alpha + 1)$:

- Increase weight x_{uv} to $1/\deg(u) + 1/\deg(v)$ if $u, v \in H$
- Increases weight by $\leq \alpha|H|/(\alpha + 2)$ and new weight is $|H|$
- Original weight was $\geq |H| - \alpha|H|/(\alpha + 2) \geq |H|/(\alpha + 1)$

Hence,

$$\sum_e x_e \geq \frac{|E_L|}{\alpha + 1} + \frac{|H|}{\alpha + 1} \geq \frac{match(G)}{\alpha + 1}$$

Implementation in Graph Streams

Edges of the graph arrive in an arbitrary order sequence. In an *insert-only* stream we only insert edges into the graph. In an *insert-delete* stream we can both insert and delete edges.

- Sample a set of vertices S with probability p. Let E_S be edges with both endpoints in S.
- In parallel:
  - Compute (approximate) matching if it is small
  - If the matching is large, use $\frac{1}{p^2}\sum_{e \in E_S} x_e$ as estimate

We balance the size of set S and space required to find (approximate) small matching to get space

- $\tilde{O}(\alpha n^{2/3})$ in insert-only streams
- $\tilde{O}(\alpha n^{4/5})$ in insert-delete streams

Approximation Using Degree Distribution

As before, let

- H be the *heavy* vertices of degree at least $\alpha + 2$
- E_H be the *heavy* edges with both endpoints heavy
- E_L be the *light* edges with both endpoints not heavy

Consider the following quantity

$$|E_L| - |E_H| + (\alpha + 1)|H|$$

$|E_H| \leq \alpha|H|$ by the property of arboricity α graphs.
Thus,

$$|E_L| - |E_H| + (\alpha + 1)|H| \geq |E_L| + |H| \geq \text{match}(G)$$

Also

$$|E_L| - |E_H| + (\alpha + 1)|H| \leq (\alpha + 1)(|E_L| + |H|) \leq (\alpha + 1)(\alpha + 2)\,\text{match}(G)$$

With more careful analysis we can show

$$|E_L| - |E_H| + (\alpha + 1)|H| \leq |E_L| + (\alpha + 1)|H| \leq (\alpha + 2)^2/2\,\text{match}(G)$$

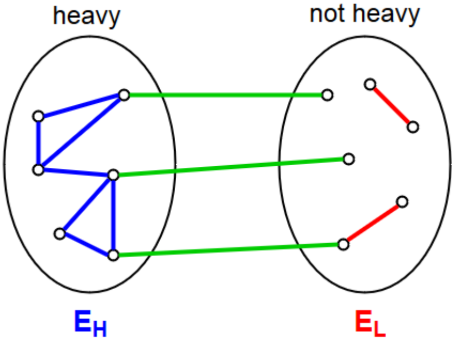
Degree Distribution and Adjacency List Streams

In adjacency list stream edges incident to the same vertex arrive together. We see each edge uv twice: when edges on u arrive and when edges on v arrive.

For our purposes, we treat adjacency stream as a degree sequence of the graph and show that we can calculate $|E_L| - |E_H| + (\alpha + 1)|H|$ from that using space $O(\log n)$.

Degree Distribution and Adjacency List Streams

$|H|$ can be computed easily.



$$|E_L| - |E_H| = |E| - \sum_{v \in H} \deg(v)$$

which is also easy to compute.

Approximating Weighted Matchings

c-approximation of matching in unweighted graphs $\Rightarrow$ $2(1+\epsilon)c$-approximation in weighted. This implies $2(1+\epsilon)(\alpha+2)$ approximation for weighted graphs with arboricity α.

Let

- E_k be the edges of weight at least $(1+\epsilon)^k$
- $\widehat{m}_k$ be the c-approximation of *unweighted* matching in (V, E_k)

Then

$$(1+\epsilon)\widehat{m}_0 + \sum_{k>0} \left((1+\epsilon)^{k+1} - (1+\epsilon)^k\right) \widehat{m}_k$$

is a $2(1+\epsilon)c$-approximation of maximum weight matching in G.

Proof Idea: Lower Bound

Let M be the set of edges in the maximum weighted matching. Note that $M \cap E_k$ is a matching in E_k.

$$
\begin{aligned}
&(1+\epsilon)\widehat{m}_0 + \sum_{k=1}^{r} \left((1+\epsilon)^{k+1} - (1+\epsilon)^{k}\right) \widehat{m}_k \\
&\quad \geq \frac{1}{c}\left((1+\epsilon)|M \cap E_0| + \sum_{k=1}^{r} \left((1+\epsilon)^{k+1} - (1+\epsilon)^{k}\right) |M \cap E_k|\right) \\
&\quad \geq \frac{\text{match}(G)}{c}
\end{aligned}
$$

Proof Idea: Upper Bound

Consider the following matching M':

- take a maximal matching in E_r where $r = \lfloor \log_{1+\epsilon} W \rfloor$
- extend it to a maximal matching in E_{r-1}
- extend it to a maximal matching in E_{r-2}
- etc.

Let $m_k = |M' \cap E_k|$. Note: it is a maximal matching in E_k. Then,

$$\begin{aligned} &(1+\epsilon)\widehat{m}_0 + \sum_{k=1}^{r} \left((1+\epsilon)^{k+1} - (1+\epsilon)^{k}\right) \widehat{m}_k \\ &\quad \leq 2\left((1+\epsilon)m_0 + \sum_{k=1}^{r} \left((1+\epsilon)^{k+1} - (1+\epsilon)^{k}\right) m_k\right) \\ &\quad \leq 2(1+\epsilon)\,\mathrm{match}(G) \end{aligned}$$

Conclusion

- "Local" fractional matchings and other quantities that provide good approximation of maximum matching in graphs of arboricity α.
- Computing those quantities can be done efficiently.

Recent work by Cormode et al. for arbitrary order streams:

- $(22.5\alpha + 6)$-approximation using space $\tilde{O}(\alpha \log^2(n))$ in insert-only streams.
- $(2\alpha + 1)(2\alpha + 2)$-approximation using space $\tilde{O}(\alpha^{10/3} n^{2/3})$ in insert-delete streams. Requires an assumption that the number of deletions in the stream is $O(\alpha n)$.

Thank you for your attention!

Bibliography

- H. Esfandiari, M.T. Hajiaghyi, V. Liaghat, M. Monemizadeh, and K. Onak. *Streaming algorithms for estimating the matching size in planar graphs and beyond*. SODA 2015
- G. Cormode, H. Jowhari, M. Monemizadeh, S. Muthukrishnan. *The Sparse Awakens: Streaming Algorithms for Matching Size Estimation in Sparse Graphs*. `arXiv:1608.03118`