Matroids, Matchings, and Fairness (Supplementary Material)

Flavio Chierichetti
Dipartimento di Informatica
Sapienza University
Rome, Italy

Ravi Kumar
Google Research
1600 Amphitheater Parkway
Mountain View, CA 94043

Silvio Lattanzi
Google Research
Brandschenkestrasse 110
Zurich, ZH 8002

Sergei Vassilvitskii
Google Research
76 9th Ave
New York, NY 10011

1/2-approximation Algorithm for Balanced Matching

We give a simple 1/2-approximation algorithm for the balanced matching problem. This algorithm only works for the maximum matching case (as opposed to any intersection of two matroids), and achieves a worse approximation ratio in the worst case. However, even this limited version is enough to demonstrate the main empirical claim, namely that there are nearly optimal balanced solutions for the problems we consider.

For a color $c \in \{\text{RED}, \text{BLUE}\}$ and for a subset E' of edges, let $\#_c(E')$ denote the number of edges in E' colored c. We call E' to be *fair* if $\#_{\text{RED}}(E') = \#_{\text{BLUE}}(E')$.

To simplify the exposition, we begin with a simple observation. A path $P = x_0, \ldots, x_k$ is called a *zebra path* if the colors on the edges alternate, i.e., $\chi(x_i, x_{i+1}) \neq \chi(x_{i+1}, x_{i+2})$ for each $0 \leq i < k - 1$; by definition, $|\#_{\text{RED}}(P) - \#_{\text{BLUE}}(P)| \leq 1$. Similarly, a cycle $C = x_0, \ldots, x_k, x_0$ is a *zebra cycle* if the colors on its edges alternate; by definition, C is fair and k is even.

Lemma 1.1. *For a zebra path with k edges, $k > 2$, we can find a fair matching of size $\lfloor k/2 \rfloor$. For a zebra cycle with k edges, $k > 2$, we can find a fair matching of size $k/2 - 2$.*

Proof. For a zebra path $x_0, \ldots, x_k$, without loss of generality, assume $\chi(x_0, x_1) = \text{RED}$. We construct a fair matching by picking all the RED edges from the subpath $x_0, \ldots, x_{\lfloor k/2 \rfloor}$ and all the BLUE edges from the remaining subpath $x_{\lceil k/2 \rceil}, \ldots, x_k$. It is easy to see this is indeed a matching (i.e., no picked edges are adjacent to each other), is fair, and is of size $\lfloor k/2 \rfloor$. For a zebra cycle $x_0, \ldots, x_k, x_0$, we first find a fair matching of size $\lfloor k/2 \rfloor = k/2$ in the path $x_0, \ldots, x_k$ and then discard the edges (x_k, x_0) and (x_0, x_1) from this. The resulting subset of edges is a fair matching and is of size $k/2 - 2$. □

Two easy consequence of this are the following.

Corollary 1.2. *For a set $\mathcal{P}$ of zebra paths, we can find a fair matching of size* $\min(\#_{\text{RED}}(\mathcal{P}), \#_{\text{BLUE}}(\mathcal{P}))$.

Proof. Without loss of generality, let $\#_{\text{RED}}(\mathcal{P}) \leq \#_{\text{BLUE}}(\mathcal{P})$. We construct a new set $\mathcal{Q}$ of zebra paths using $\mathcal{P}$. Since each $P \in \mathcal{P}$ is a zebra path, there is a subset $\mathcal{P}' \subseteq \mathcal{P}$, $|\mathcal{P}'| = \#_{\text{BLUE}}(\mathcal{P}) - \#_{\text{RED}}(\mathcal{P})$, such that each zebra path in $\mathcal{P}'$ has one more BLUE edge than a RED edge. (It is also easy to see that the first and the last edge in every such path is BLUE.) We construct $\mathcal{Q}'$ from $\mathcal{P}'$ by discarding the blue edge at an arbitrary end for every $P \in \mathcal{P}'$. Now we let $\mathcal{Q} = \mathcal{Q}' \cup (\mathcal{P} \setminus \mathcal{P}')$. By construction, $\mathcal{Q}$ is fair and the number of edges in $\mathcal{Q}$ is $2\#_{\text{RED}}(\mathcal{P})$.

Our next goal is to construct a single virtual zebra path out of all the zebra paths in $\mathcal{Q}$. For a color c, let $\mathcal{Q}_{c,c} \subseteq \mathcal{Q}$ be the set of zebra paths in $\mathcal{Q}$ where both the first and last edges have color c. Once again by construction, it is easy to see that $|\mathcal{Q}_{\text{RED,RED}}| = |\mathcal{Q}_{\text{BLUE,BLUE}}|$ and hence we can define an arbitrary bijection between these subsets. Let $\mathcal{Q}_{\text{RED,BLUE}} = \mathcal{Q} \setminus \mathcal{Q}_{\text{RED,RED}} \setminus \mathcal{Q}_{\text{BLUE,BLUE}}$. We now how to create a virtual zebra path $\tilde{P}$ by taking paths one from each of the three partitions of $\mathcal{Q}$. Let $x_0, \ldots, x_i \in \mathcal{Q}_{\text{RED,RED}}$, let $y_0, \ldots, y_j \in \mathcal{Q}_{\text{BLUE,BLUE}}$ be given by the above bijection, and let $z_0, \ldots, z_k \in \mathcal{Q}_{\text{RED,BLUE}}$. By assumption $\chi(x_0, x_1) = \chi(x_{i-1}, x_i) = \text{RED}$, $\chi(y_0, y_1) = \chi(y_{j-1}, y_j) = \text{BLUE}$ and without loss of generality $\chi(z_0, z_1) = \text{RED}, \chi(z_{k-1}, z_k) = \text{BLUE}$. Furthermore, i, j are odd and k is even. Define $\tilde{P} = x_0, \ldots, x_i, y_1, \ldots, y_j, z_1, \ldots, z_k$ where $\chi(x_i, y_1) = \text{BLUE}$ and $\chi(y_j, z_1) = \text{RED}$. The number of edges in $\tilde{P}$ is $i + j + k$ and is even. Furthermore, it is easy to see that a fair matching in $\tilde{P}$ corresponds to a fair matching in the three original zebra paths. We can iterate this construction for all the paths in $\mathcal{Q}$, based on the partition, to obtain a fair matching of size $\#_{\text{RED}}(\mathcal{Q})$. This concludes the proof. □

Corollary 1.3. *For a set $\mathcal{C}$ of zebra cycles, we can find a fair matching of size $\#_{\text{RED}}(\mathcal{C}) - 2$.*

Proof. For two zebra cycles $x_0, \ldots, x_k, x_0$ and $y_0, \ldots, y_\ell, y_0$ in $\mathcal{C}$, we construct a virtual zebra cycle $\tilde{C}$ in the following manner. Without loss of generality, let $\chi(x_0, x_1) \neq \chi(y_0, y_1)$. Then, $\tilde{C}$ is given by $\tilde{C} = x_0, \ldots, x_k, y_0, \ldots, y_\ell, x_0$, where $\chi(x_k, y_0) = \chi(x_k, x_0)$ and $\chi(y_\ell, x_0) = \chi(y_\ell, y_0)$. Note that the number of edges in $\tilde{C}$ is $k + \ell$ and $\#_{\text{RED}}(\tilde{C}) = (k+\ell)/2$. Hence, by Lemma 1.1 we can find a fair matching in $\tilde{C}$ of size $\#_{\text{RED}}(\tilde{C}) - 2$. It is easy to see that a fair matching in $\tilde{C}$ corresponds to a fair matching in the original zebra cycles. We can iterate this construction for all the zebra cycles in $\mathcal{C}$ to conclude the proof. □

It is easy to see that both these bounds are tight. Using these bounds we obtain our main result in this section.

Theorem 1.4. *There is an algorithm to find fair matching of size* $\text{OPT}/2 - 2$ *on edge-colored graphs, where* OPT *is the size of the optimal solution.*

Proof. For a color c, let M_c be a maximum matching on c-colored edges in G. Let $M = M_{\text{RED}} \cup M_{\text{BLUE}}$. By the property of matchings, M is a union of zebra paths $\mathcal{P}$ and zebra cycles $\mathcal{C}$. The algorithm outputs a fair matching that is the union of Corollary 1.2's output when applied to $\mathcal{P}$ and Corollary 1.3's output when applied to $\mathcal{C}$.

For the analysis, first observe that

$$\begin{aligned} \text{OPT} &\leq 2\min(|M_{\text{RED}}|, |M_{\text{BLUE}}|) \\ &= 2\min(\#_{\text{RED}}(\mathcal{P}), \#_{\text{BLUE}}(\mathcal{P})) + 2\#_{\text{RED}}(\mathcal{C}), \end{aligned}$$

where we used $\#_{\text{RED}}(\mathcal{C}) = \#_{\text{BLUE}}(\mathcal{C})$.

Next, by Corollary 1.2 and Corollary 1.3, the size of the fair matching output by the algorithm is at least

$$\min(\#_{\text{RED}}(\mathcal{P}), \#_{\text{BLUE}}(\mathcal{P})) + (\#_{\text{RED}}(\mathcal{C}) - 2) \geq \text{OPT}/2 - 2.$$

□