Fairness in Graph-theoretical Optimization Problems*

Christopher Hojny[1], Frits Spieksma[1] and Sten Wessel[1,*]

[1]*Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands*

Abstract

There is arbitrariness in optimum solutions of graph-theoretic problems that can give rise to unfairness. Incorporating fairness in such problems, however, can be done in multiple ways. For instance, fairness can be defined on an individual level, for individual vertices or edges of a given graph, or on a group level. In this work, we analyze in detail two individual-fairness measures that are based on finding a probability distribution over the set of solutions. One measure guarantees uniform fairness, i.e., entities have equal chance of being part of the solution when sampling from this probability distribution. The other measure maximizes the minimum probability for every entity of being selected in a solution. In particular, we reveal that computing these individual-fairness measures is in fact equivalent to computing the fractional covering number and the fractional partitioning number of a hypergraph. In addition, we show that for a general class of problems that we classify as independence systems, these two measures coincide. We also analyze group fairness and how this can be combined with the individual-fairness measures.

Keywords

fairness, column generation, matching, independent set, set systems

1. Introduction

Traditionally, when confronted with an instance of an optimization problem, the instance is considered solved when a provably optimum solution has been found. After all, what more can be wished for? We should acknowledge that there can be a large amount of arbitrariness in selecting an optimum solution, and that this can be perceived as a source of unfairness.

Imagine one is a patient with end-stage renal disease, and that there is a donor who is willing to donate a kidney but who is incompatible with you. Imagine further that you and your incompatible donor enter a kidney exchange program; we will refer to the pair entering the program as a node. Entering this program means that at regular moments, runs of software are made, and the output of such a run consists of a set of exchanges between nodes. A patient from a node involved in such an exchange receives a kidney, other patients do not. Clearly, it is quite relevant whether one is included in such an exchange or not. While the software may find an optimum solution (according to some criterion, say maximizing the number of exchanges), it can be the case that some node is not selected for an exchange while being part of *some* optimum solution. In such cases it can be hard to explain that the software simply selects an optimum solution which somehow favors some nodes at the expense of others. In this

EWAF'24: European Workshop on Algorithmic Fairness, July 01–03, 2024, Mainz, Germany

*This extended abstract is based on the preprint of our full paper [1].

*Corresponding author.

c.hojny@tue.nl (C. Hojny); f.c.r.spieksma@tue.nl (F. Spieksma); s.wessel@tue.nl (S. Wessel)

https://stenwessel.nl (S. Wessel)

contribution, we aim to pursue this matter from a fairness point of view, using kidney exchange as a running example.

We analyze an approach that has been used by, among others, Farnadi et al. [2] and Flanigan et al. [3]. Instead of proposing a fixed solution, the idea is to consider a pool of solutions from which a solution is sampled according to some probability distribution, which we call *procedural* or *ex-ante* fairness. In this way, different entities of the optimization problem (e.g., patients in the kidney exchange example) are contained in a selected solution with a certain probability. To model fairness, the probability distribution is required to satisfy some additional criteria.

Although this sampling approach is not new, it has mostly been studied computationally. In this article, we take a fresh look into this approach. While existing results in the literature mostly considered the approach by Flanigan et al. [3] for particular problems, we aim to understand the approach on a structural level. For more details on our results, we refer to [1].

2. A generic framework for modeling fairness in graphs

We formulate a model for finding a probability distribution over the set of feasible solutions of a graph-theoretical optimization problem. This model resembles fairness models for concrete problems that exist in the literature [2, 3]. Let A be a given *ground set*, and let M be a family of subsets of the ground set A. We assume that $\varnothing \in M$ and that every element $a \in A$ is contained in some set $m \in M$. We call the tuple (A, M) a *set system*. In the kidney exchange example, the ground set A coincides with the nodes, whereas M coincides with the collection of subsets of nodes that are covered by matchings in the compatibility graph.

To achieve individual fairness, we construct a probability distribution $\{x_m\}_{m \in M}$ over M, such that if we sample according to x, every element $a \in A$ has equal probability p to be in the sampled subset. Our aim is to maximize this probability, which we call p_U (for *uniform* probability). For any $a \in A$, let $M_a \subseteq M$ denote the collection of subsets from M that contain the a. We model the problem of maximizing p_U with the following linear program (LP).

$$
\begin{aligned}
p_U = \quad & \text{maximize} \quad & & p & & \qquad \text{(1a)}\\
& \text{subject to} \quad & & \sum_{m \in M_a} x_m = p & & \forall a \in A, \qquad \text{(1b)}\\
& & & \sum_{m \in M} x_m = 1, & & \qquad \text{(1c)}\\
& & & x_m \geq 0 & & \forall m \in M, \qquad \text{(1d)}\\
& & & p \in \mathbb{R}. & & \qquad \text{(1e)}
\end{aligned}
$$

Constraints (1b) ensure that each element $a \in A$ is selected with uniform probability, while Constraint (1c) and Constraints (1d) ensure that x is a probability distribution over M.

We also consider a variant of this problem, where our notion of fairness is relaxed to maximizing the minimum probability p_R that a ground set element is selected. This is also referred to as *Rawlsian justice*, based on fairness principles introduced by Rawls [4], and can be modeled with an LP that is very similar to (1), where Constraints (1b) are replaced with $\sum_{m \in M_a} x_m \geq p$

for all $a \in A$. For some applications, Rawlsian justice may be more applicable than uniform fairness as a fairness measure, or vice versa, depending on the objective of the decision maker.

2.1. Relation to fractional hypergraph theory

The fairness measures p_U and p_R are related to concepts developed in fractional hypergraph theory. We refer to Scheinerman and Ullman [5], and Füredi [6] for an overview of fractional graph and hypergraph theory. The main idea is to see the set system (A, M) as a *hypergraph* $\mathcal{H}$ with vertex set A and hyperedges M. The *fractional partitioning number* $k_f^{=}(\mathcal{H})$ is the minimum weight of a fractional partition: an assignment of non-negative weights $w \in \mathbb{R}_+^M$ to the hyperedges such that for every vertex the sum of weights of the incident hyperedges is exactly 1. When the weights sum to *at least* 1 for every vertex, we call the assignment of weights a *fractional cover* and consequently the minimum weight of a fractional cover is the *fractional covering number* $k_f^{\geq}(\mathcal{H})$ of the hypergraph.

One can show that determining p_U can then be alternatively formulated as determining the fractional partitioning number and similarly that measure p_R can be formulated as determining the fractional covering number, as summarized in the following result.

Lemma 1. *Let $\mathcal{H} = (A, M)$ be a hypergraph. Then it holds that $p_R = \frac{1}{k_f^{\geq}(\mathcal{H})}$. Furthermore, $p_U = \frac{1}{k_f^{=}(\mathcal{H})}$ if $k_f^{=}(\mathcal{H})$ is finite, and $p_U = 0$ otherwise.*

In fact, when we apply this to the kidney exchange example, one can show that this implies the following, rather high, lower bound on p_U when it is positive.

Corollary 1. *For fair matching for vertices, if $p_U > 0$, then $p_U \geq \frac{2}{3}$.*

2.2. Computational complexity

A fair probability distribution with respect to p_U and p_R can be found by solving the LP (1) for p_U and p_R, respectively. Linear programs are solved routinely in practice by the simplex algorithm [7], and algorithms such as the ellipsoid method [8] and interior point methods [9] have a provable polynomial running time. Note, however, that this does not imply that the LPs can be solved in $\mathcal{O}(\text{poly}(|A|))$ time, as the running time of the ellipsoid and interior point methods depend polynomially on the number of variables $|M|$, which can depend exponentially on $|A|$. We make use of a more advanced technique for solving LPs: column generation. The idea of column generation is to select a small subset $M' \subseteq M$ of variables and solve the LP restricted to this subset of variables. Afterward, it is checked whether the solution is optimal w.r.t. the entire set of variables M. If this is the case, the algorithm stops and returns an optimal solution. Otherwise, it adds a variable from $M \setminus M'$ to M' that can improve the objective and the method starts anew. Using column generation, we can observe that p_U and p_R can be found in polynomial time whenever the corresponding optimization problem over M can be solved in polynomial time, based on the seminal result by Grötschel, Lovász, and Schrijver [10].

Theorem 2. *Let A be a finite set and let M be a collection of subsets of A such that $\varnothing \in M$. If, for every $c \in \mathbb{Q}^A$, a set $m \in M$ that minimizes $\sum_{a \in m} c_a$ can be found in time polynomial in $|A|$, then p_U and p_R can be found in $\mathcal{O}(poly(|A|))$ time.*

Examples where minimizing $\sum_{a \in m} c_a$ can be done in polynomial time are matching, shortest paths, spanning trees, and many other combinatorial problems.

3. Independence systems

The framework that determines p_U and p_R (Section 2) can be applied to arbitrary set systems (A, M). The system (A, M) is an *independence system* when $\varnothing \in M$ and for every $m \in M$, we have $m' \in M$ for all $m' \subseteq m$. Independence systems are also known as *simplicial complexes* or *downward-closed set systems*. Many independence systems are known; we mention here: edge sets that form matchings, edge sets that form forests, vertex sets that are independent sets, and matroids. While in general there is no relation between p_U and p_R other than $p_U \leq p_R$, one of our main contributions is that for independence systems that these two measures coincide.

Theorem 3. *For any independence system (A, M), $p_U = p_R$.*

4. Group fairness

Other than arbitrariness in selecting a solution, unfairness can also be introduced through systemic bias in the definition of the model or the description of the problem. For the kidney exchange example, patients with certain blood types are more likely to be matched, simply because they are compatible with a larger group of donors. To achieve this so-called *group fairness* among different types of patients, one can ensure representation of protected groups in the solution. Not much existing work combines individual and group fairness, while we show that both approaches can be combined in a single framework.

Our problem input now not only consists of a set system (A, M), but also of a collection of *groups* $\mathcal{G} = \{G_1, \ldots, G_k\}$, with $G_i \subseteq A$ subsets of the ground set A, for all $i \in [k]$. We assume that the groups are pairwise disjoint; when considering groups that arise from a single sensitive attribute, this is in many cases a natural assumption. We impose group fairness constraints on the possible solutions, i.e., we require *ex-post* group fairness. We thus restrict the set M to only solutions that satisfy the group fairness constraints, $M_{\text{gf}} \subseteq M$. Although it is possible to also handle group fairness in an ex-ante setting as opposed to ex-post, group fairness constraints are often hard conditions imposed on solutions, and in some cases required by regulations or law.

We consider two types of group fairness constraints. *Absolute* constraints enforce that the number of individuals from a group is between a given lower and upper bound. *Relative* constraints bound the ratio between the number of individuals for a pair of groups, e.g., to impose that a solution must contain the same number of individuals from each group. For the fair matching for vertices problem in the kidney exchange example, we obtain the following.

Theorem 4. *Let $G = (V, E)$ be a graph with a constant number of pairwise disjoint groups of vertices. Let M_{gf} denote the collection of vertex-subsets covered by matchings in G, restricted to absolute and/or relative group fairness constraints. Then determining p_U and p_R for M_{gf} can be done in polynomial time.*

Acknowledgments

This research is supported by NWO Gravitation Project NETWORKS, Grant Number 024.002.003.

References

[1] C. Hojny, F. Spieksma, S. Wessel, Fairness in graph-theoretical optimization problems, 2023. doi:10.48550/arXiv.2311.15953.

[2] G. Farnadi, W. St-Arnaud, B. Babaki, M. Carvalho, Individual fairness in kidney exchange programs, Proceedings of the AAAI Conference on Artificial Intelligence 35 (2021) 11496–11505. doi:10.1609/aaai.v35i13.17369.

[3] B. Flanigan, P. Gölz, A. Gupta, B. Hennig, A. D. Procaccia, Fair algorithms for selecting citizens' assemblies, Nature 596 (2021) 548–552. doi:10.1038/s41586-021-03788-6.

[4] J. Rawls, A Theory of Justice, Belknap Press, 1971.

[5] E. R. Scheinerman, D. H. Ullman, Fractional graph theory: a rational approach to the theory of graphs, Wiley-Inerscience Series in Discrete Mathematics and Optimization, Wiley, New York, NY, USA, 1997.

[6] Z. Füredi, Matchings and covers in hypergraphs, Graphs and Combinatorics 4 (1988) 115–206. doi:10.1007/BF01864160.

[7] G. B. Dantzig, Maximization of a linear function of variables subject to linear inequalities, Activity analysis of production and allocation 13 (1951) 339–347.

[8] L. G. Khachiyan, A polynomial algorithm in linear programming, in: Soviet Mathematics Doklady, volume 20, 1979, pp. 191–194.

[9] N. Karmarkar, A new polynomial-time algorithm for linear programming, in: Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing, STOC '84, Association for Computing Machinery, New York, NY, USA, 1984, p. 302–311. doi:10.1145/800057.808695.

[10] M. Grötschel, L. Lovász, A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, Combinatorica 1 (1981) 169–197. doi:10.1007/BF02579273.