The Price of Quota-based Diversity in Assignment Problems

Nawal Benabbou, Mithun Chakraborty, Xuan-Vinh Ho, Jakub Sliwinski, Yair Zick

The Price of Quota-based Diversity in Assignment Problems[*]

NAWAL BENABBOU[†], Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, France

MITHUN CHAKRABORTY[†], University of Michigan, Ann Arbor, USA

XUAN-VINH HO[†], Micron Technology, Singapore

JAKUB SLIWINSKI[†], ETH Zürich, Switzerland

YAIR ZICK[†], University of Massachusetts, Amherst, USA

In this paper, we introduce and analyze an extension to the matching problem on a weighted bipartite graph (i.e. the assignment problem): Assignment with Type Constraints. Here, the two parts of the graph are each partitioned into subsets, called types and blocks respectively; we seek a matching with the largest sum of weights under the constraint that there is a pre-specified cap on the number of vertices matched in every type-block pair. Our primary motivation stems from the large-scale public housing program run by the state of Singapore, accounting for over 70% of its residential real estate. To promote ethnic diversity within its housing projects, Singapore imposes ethnicity quotas: the population is divided into ethnicity-based groups and each new housing development into blocks of flats such that each group must not own more than a certain percentage of flats in a block. However, other domains use similar hard capacity constraints to maintain diversity: these include matching prospective students to schools or medical residents to hospitals. Limiting agents' choices for ensuring diversity in this manner naturally entails some welfare loss. One of our goals is to study the tradeoff between diversity and (utilitarian) social welfare in such settings. We first show that, while the classic assignment program is polynomial-time computable, adding diversity constraints makes the problem computationally intractable; however, we identify a $\frac{1}{2}$-approximation algorithm, as well as reasonable assumptions on the structure of utilities (or weights) which permit poly-time algorithms. Next, we provide two upper bounds on the *price of diversity* — a measure of the loss in welfare incurred by imposing diversity constraints — as functions of natural problem parameters. We conclude the paper with simulations based on publicly available data from two diversity-constrained allocation problems — Singapore Public Housing and Chicago School Choice — which shed light on how the constrained maximization as well as lottery-based variants perform in practice.

CCS Concepts: • **Theory of computation → Problems, reductions and completeness**; • **Computing methodologies → Artificial intelligence**;

Additional Key Words and Phrases: Assignment Problem; Diversity Constraints; Price of Diversity

---

Authors' addresses: Nawal Benabbou, Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, LIP6 F-75005, Paris, France, nawal.benabbou@lip6.fr; Mithun Chakraborty, University of Michigan, Ann Arbor, 2260 Hayward Street, Ann Arbor, MI 48109, USA, dcsmc@umich.edu; Xuan-Vinh Ho, Micron Technology, Singapore, hxvinh.hcmus@gmail.com; Jakub Sliwinski, ETH Zürich, Rämistrasse 101, 8092, Zurich, Switzerland, sljakub@ethz.ch; Yair Zick, University of Massachusetts, Amherst, 140 Governors Drive, Amherst, MA 01002, USA, yairzick@cs.umass.edu.

---

1 INTRODUCTION

Consider a mechanism that allocates a set of goods to agents; agents have utilities over items, and we are interested in finding a *socially optimal allocation*. This setting (known in the literature as the *assignment problem*) is often used to model real-world problems such as allocating public housing, assigning slots in schools, or courses to students. But it is also often the case in these contexts that one wishes to maintain a *diverse* allocation: it would be undesirable (from the mechanism designer's perspective) to have certain apartment blocks that predominantly consist of a specific ethnic group, or to have a public school serving students from a specific district. In both cases, agents have different *types*, and goods are partitioned into *blocks*; our goal is to ensure that each block of goods is allocated to a diverse population of agent types. A diverse allocation of goods is desirable for many reasons (especially in the case of government-funded public goods). First and foremost, it avoids the inadvertent creation of segregated communities; secondly, by ensuring equal access to a public resource, one avoids the risk of discriminatory funding: for example, systematically underfunding schools that serve certain segments of the population, or investing in parks and public facilities in neighborhoods dominated by certain ethnic groups.

In this work, we study *quota-based* mechanisms for maintaining diversity; the initial motivation for this work stems from Singapore's public housing system.

The state of Singapore operates a unique national public housing program, offering a variety of flats for sale at subsidized rates to Singapore citizens and permanent residents. The construction of public housing projects as well as the sale of the flats in these projects on a large-scale public market is centrally managed by a government body called the Housing and Development Board (HDB),[1] a statutory board of the Ministry of National Development.[2] As per the latest reports available at the time of writing this paper, an estimated 82% of the resident population of Singapore live in HDB flats [Housing and Development Board, Singapore 2017] that constitute approximately 73% of all apartments in the country [Department of Statistics, Singapore 2017]. Since its inception in 1960, HDB has been providing a public good — affordable apartments in a small country with little real estate — but by 1989, the system began to exhibit an unforeseen side-effect: the emergence of de facto ethnic enclaves. Mr. S. Dhanabalan, then Minister for National Development, voiced the following concerns as he introduced the Ethnic Integration Policy (EIP) in parliament on 16 February 1989 [Parliament of Singapore. Parliament Debates: Official Report. 1989]:

> [P]roportionately more Chinese applied for flats in Ang Mo Kio/Hougang Zone and proportionately more Malays applied for flats in the Bedok/Tampines Zone. [. . .] Malays bought more than half (55%) of the flats in the Bedok/Tampines Zone. In Bedok new town alone, if present trends continue, the proportion of Malays will reach 30% by 1991, and will exceed 40% in 10 years' time. [. . .]
>
> There are clear signs that racial congregations are re-emerging. Although the problem has not reached

---

[1] http://www.hdb.gov.sg
[2] https://www.mnd.gov.sg/

crisis proportions, the experience in other multi-racial societies such as the United States shows that while racial groupings start slowly, once a critical point is passed, racial groupings accelerate suddenly.[3]

The EIP was officially implemented on 1 March 1989; it imposes quotas on the number of units occupied by each of the three ethnic groups: Chinese, Malay and Indian/Others. In 1989, when the percentages of the three ethnic groups (Chinese, Malay, and Indian/Others) in the population were 76.0%, 15.1%, and 8.9%, the corresponding quotas (a cap on the percentage of flats in every block that can be occupied by families of each ethnic group) were set at 87%, 25%, and 13% respectively; since 5 March 2010, the quota for Indian/Others has been revised to 15% [Deng et al. 2013; Housing and Development Board, Singapore 2010].[4]

Ethnic quotas add another layer of complexity to what is, at its foundation, a straightforward allocation problem. HDB uses a lottery mechanism to allocate new developments: all applicants who apply for a particular development pick their flats in random order (see Section 1.3 for further details). Consider an applicant i of Chinese ethnicity applying for an estate with 100 flats per block, up to 87 of which may be assigned to ethnically Chinese applicants, and at most 25 of which can be assigned to ethnically Malay applicants. Assume that i is 90th in line to select an apartment; will she get a chance to pick a flat in a block she prefers? If at least 87 Chinese applicants were allowed to choose a flat before i and all of them picked flats in this block, the Chinese ethnic quota for the block will have been filled and applicant i will no longer be eligible for the block, even if it still has vacant flats. On the other hand, suppose that i is 105th in line to select an apartment. If 40 of the applicants picked before i are Malay, then 15 of them will be rejected from i's preferred block and at least 75 of the flats will be available for non-Malay applicants. Hence, i will have a spot even if all other 64 applicants before her also received flats in the same block.[5]

As the example above shows, diversity constraints interact with the allocation mechanism in peculiar ways to affect the overall welfare of the allocation. This issue is not restricted to Singapore public housing; the following are a few examples where upper bounds similar to those in the above housing allocation problem are applied (see [Fragiadakis and Troyan 2017] and references therein for more detailed expositions). To circumvent a shortage of doctors in rural areas due to medical graduates' preference for urban residency programs, the Japanese government places a "regional cap" on the total number of residents matched within each of its 47 prefectures [Kamada and Kojima 2015] – here, the population of residency applicants is not partitioned, but hospitals within a prefecture can be thought of as forming a block of items. Many school districts in the U.S.A. take active measures for the integration of students from families with differing *soci-economic statuses* (SES) [U.S. Department of Education, Office of Elementary and Secondary Education 2017], one of which is to allot a fraction of the vacant spots in schools via lotteries with percentage caps for all SES groups, as is done in the city of Chicago, Illinois (see Section 1.4 for further details). The United States Military Academy assigns newly graduated cadets to positions in the army branches, taking cadets' preferences into account but under "artificial caps"[6] on the number of assignments per branch [Fragiadakis and Troyan 2017].

---

[3]We note that this aligns with long-known models of segregation: in his seminal paper, Schelling [1971] shows how agents of two types, who are allowed to distribute themselves over an area based on their preferences for the composition of their immediate neigborhoods, lead to the emergence of segregated enclaves, even when each individual prefers a minority of neighbors of a different type to having all neighbors of the same type as herself.

[4]In addition to these block capacities, the EIP also imposes neighborhood capacities, where each neighborhood comprises several blocks. Thus, an estate is partitioned into neighborhoods and a neighborhood into blocks with the neighborhood capacity being naturally smaller than the block capacity for each ethnic group: 84%, 22%, and 12% (increased from 10% in 2010) for Chinese, Malay, and Indian/Others respectively. Moreover, the EIP applies to both new and resale flats, e.g. a Chinese occupant of an HDB flat is free to resell it to another Chinese buyer but will not be able to resell to a Malay buyer if the Malay ethnic quota for that block is already filled. In this paper, we do not address these complications for the most part.

[5]While this example is stylized, the effects it describes are quite real: one often hears stories of young couples who arrive at the HDB office to select a flat, only to be notified that their ethnic quota has just been filled.

[6]These caps are called artificial since they are calculated by the Academy in such a way that any feasible assignment ex post satisfies the maximum and minimum quotas for each branch that are based on actual staffing needs.

The imposition of diversity constraints as above can naturally lead to a reduction in the total achievable utility/economic value by the assignment, but we must bear in mind that diversity is a social desideratum external to any such economic consideration. For purposes such as policy making and the proper functioning of diversity-inducing measures included in automated decision-making systems, it is imperative to deepen our understanding of the impact that these measures have on the underlying assignment mechanism. In this paper, we investigate this impact from both computational and economic angles.

1.1 Our Contributions

We study the interplay between diversity and utility in assignment problems; we set up a benchmark where a central planner (e.g. HDB) has access to the correct utilities (or, in general, weights) of all agents (e.g. applicant households) for all items (e.g. flats); agents are partitioned into *types* with respect to a single attribute (e.g. ethnic groups) and goods are also similarly divided into disjoint *blocks* (e.g. blocks of flats as defined by HDB); a limited number of goods in each block can be allocated to agents of each type. We call these upper limits *type-block capacities*.

These restrictions result in several interesting outcomes. While the unconstrained optimal assignment problem is well-known to be polynomial-time solvable [Kuhn 1955], we show that imposing type-block constraints makes it computationally intractable (Section 3). However, we also show that, in general, a polynomial-time $\frac{1}{2}$-approximation algorithm (Section 3.1) exists, and identify utility models for which one can find the optimal assignment with type-block constraints in polynomial time (Section 3.2). In Section 4, we study the potential welfare loss from imposing type-block constraints, which we term the *price of diversity* as in Ahmed et al. [2017], and we show that it can be bounded by natural problem parameters. Finally, we analyze the empirical price of diversity as well as the welfare loss induced by the lottery mechanism on simulated instances generated from publicly available, real-world data pertaining to public housing in Singapore and school choice in Chicago, IL, USA (Section 5).

1.2 Related work

The problem we study is an extension to the bipartite matching problem [Lovász and Plummer 2009] where each edge joins an agent to an item and is weighted with the utility the agent will receive if she is allocated that item. There is a rich body of literature on weighted bipartite matching problems (also known as assignment problems [Munkres 1957]), and polynomial-time algorithms for the unconstrained version have long been known (e.g. [Kuhn 1955]). Several generalizations and/or constrained versions have been studied, e.g. recent work by Lian et al. [2018] who allow each agent (resp. item) to be matched to multiple items (resp. agents) but within upper and lower capacities. Some previously studied variants correspond to (polynomial-time) special cases of our problem. For example, the assignment problem with subset constraints studied by Bauer [2004] can be thought of as a special case of our problem, with a single block or a single type; if all agents of each type have identical utilities for all apartments in each block, and each type-block capacity is smaller than both the corresponding type and block sizes, then our problem reduces to a special case of the polynomial-time solvable *capacitated b-matching* on a bipartite graph [Ahn and Guha 2014].

In addition to our main motivating problem of HDB housing allocation and the other documented examples [Fragiadakis and Troyan 2017] noted in the introduction, type-block constraints can naturally arise in many other settings related to assignment/allocation problems with no monetary transfers [Hylland and Zeckhauser 1979; Zhou 1990]. For example, consider the course allocation problem analyzed by Budish and Cantillon [2012]; one might require that each course has students from different departments and impose maximal quotas to ensure this. Other examples include allocating subsidized on-campus housing to students [Abdulkadiroğlu and Sönmez 1998], appointing teachers at

public schools in different regions as done by some non-profit organizations [Featherstone 2015], or assigning first year business school students to overseas programs [Featherstone 2015]. Our results apply to the seminal work on public school allocation [Abdulkadiroğlu et al. 2009; Abdulkadiroğlu and Sönmez 2003; Pathak and Sönmez 2013] and matching medical interns or residents to hospitals [Roth 1984] that does not concern itself with diversity/distributional constraints. This line of work mainly explores the interaction between individual selfish behavior and allocative efficiency (e.g. Pareto-optimality) of matching mechanisms, under either ordinal preferences or cardinal utilities, one-sided or two-sided (see, e.g. [Anshelevich et al. 2013; Bade 2016; Bhalgat et al. 2011; Bogomolnaia and Moulin 2001] and references therein); we, on the other hand, focus on the impact of type-block constraints on welfare loss, when agents' utilities are known to a central planner.

Another relevant strand of literature is that on the fair allocation of indivisible goods (see, e.g. [Barman et al. 2017; Barman and Murthy 2017; Caragiannis et al. 2016; Kurokawa et al. 2016; Procaccia and Wang 2014] and references therein): fairness is usually quantified in terms of the utilities or preferences of agents for allocated items (e.g. proportionality, envy-freeness and the maximin share guarantee) but our contribution deals with a different notion of fairness: the proportionate representation of groups in the realized allocation, with no regard to agents' utilities.

Some recent work has formally addressed diversity issues in computational social choice. Unlike our paper, Ahmed et al. [2017] treat "diversity as an objective, not a constraint" in a b-matching context (e.g. matching papers to reviewers with diverse interests): they minimize a supermodular objective function to encourage the matching of each item to agents of different types. Our diversity concept comes close to that of Bredereck et al. [2018] who also achieve diversity by imposing hard constraints on the maximization of a (submodular) objective that measures the quality of the solution; however, they work in a committee (subset) selection setting with variously structured agent labels while we solve a matching problem with both agents and items split into disjoint subsets. Lang and Skowron [2016] focus on multi-attribute *proportional representation* in committee selection where they essentially define diversity in terms of the divergence between the realized distribution of attribute values in the outcome and some target distribution, but admit no notion of solution quality in addition to diversity.

In a recent paper, Immorlica et al. [2017] study the efficiency of lottery mechanisms such as the ones used by HDB to allocate apartments; however, their work does not account for block ethnicity constraints; as we show both theoretically and empirically, these type-block constraints can have a significant effect on allocative efficiency.

1.3 The Singapore Public Housing Allocation System

A few facts about HDB public housing, a dominant force in Singapore, are in order. New HDB flats are purchased directly from the government, which offers them at a heavily subsidized rate. New apartments are typically released at quarterly sales launches; these normally consist of plans for several estates at various locations around Singapore, an estate consisting of four or five blocks (each apartment block has approximately 100 apartments) sharing some communal facilities (e.g. a playground, a food court, a few shops etc.). Estates take between 3 to 5 years to complete, during which HDB publicly advertises calls to ballot for an apartment in the new estate. A household (say, a newly married couple looking for a new house) would normally ballot for a few estates (balloting is cheap: only S$10 per application [Housing and Development Board, Singapore 2015]). HDB allocates apartments using a lottery: all applicants to a certain estate choose their flat in some random order; they are only allowed to select an apartment in a block such that their ethnic quota is not reached.

The lottery mechanism actually employed by HDB has further necessary complications: HDB has elaborate eligibility criteria[7] as well as privilege and priority schemes[8] that take into account sales launch types, flat types, and relevant attributes of the applicants, e.g. first-timers and low-income families usually have improved chances of being balloted for a flat; moreover, the same estate may have several balloting rounds in order to ensure that all apartments are allocated by the time of completion. However, the focus of this work is on the welfare effects of using ethnic quotas rather than the intricacies of the HDB lottery mechanism. Hence, we use a simplified version of the HDB lottery mechanism where applicants are selected one by one uniformly at random from the remaining pool and assigned the available flat which they value the most, respecting ethnic quotas (see Section 5).

We must mention the existing literature on the documentation of Singapore's residential desegregation policies [Chua 1991; Deng et al. 2013; Phang and Kim 2013] and the empirical evaluation of their impact on various socioeconomic factors [Sim et al. 2003; Wong 2014]; to the best of our knowledge, ours is the first formal approach towards this problem.

1.4   Public School Choice in Chicago, U.S.A.

Many school districts across the U.S.A. employ a variety of strategies for promoting student diversity [Kahlenberg 2016; U.S. Department of Education, Office of Elementary and Secondary Education 2017], e.g. controlled choice systems wherein parents are allowed to apply for options beyond their neighborhood schools, thereby counteracting underlying residential segregation. Following restrictions placed on the explicit use of race in defining diversity goals in school choice by the U.S. Supreme Court in 2007, it has been common to use some indicator of the *socio-economic status* (SES) of a family in integration efforts. The system in Chicago, IL, is a notable example.

Chicago Public Schools (CPS) is one of the largest school districts in the U.S.A.[9], overseeing more than 600 schools of various types: neighborhood schools, selective schools, magnet schools, and charter schools.[10] The application and selection processes for these schools [Chicago Public Schools 2017] may involve a number of computerized lotteries with no diversity component, e.g. sibling lottery, proximity lottery, school staff preference lottery; however, a significant number of entry-level seats in magnet and selective enrollment schools are filled by lotteries based on a tier system. We briefly describe its operation as follows. A composite SES score is computed for each of the census tracts that Chicago is divided into, based on six factors (median family income, adult education level, home-ownership rate, single-parent family rate, rate of English-speaking, and neighborhood school performance), and each tract is placed in one of four tiers based on its score. The maximum and minimum scores defining a tier are set in such a way that (roughly) a quarter of school-aged children end up in each tier, with Tier 1 having the lowest scores. The tier of a child is determined by the residential address furnished by the parents. Of the seats in each school earmarked for a *citywide SES lottery* or *general lottery*, an equal number is allocated to each tier. There is an upper limit on the number of schools that a child can apply to, and each applicant is entered into a lottery for each school they apply to, for their own tier (thus, there is a lottery per school per tier); an applicant, who comes up in the lottery and accepts the offer from the school under consideration, is removed from all lotteries. If the size of the applicant pool from a tier to a school falls short of the number of its allocated seats for that tier at any stage, "the unfilled seats will be divided evenly and redistributed across the remaining tier(s) as the process continues" [Chicago Public Schools 2017].

For an empirical study of the impact of Chicago's diversity-promoting measures on integration and student outcomes, the interested reader is referred to [Quick 2016] and citations therein.

---

[7]http://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/hdb-flat
[8]http://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/eligibility/priority-schemes
[9]http://www.cps.edu/About_CPS/At-a-glance/Pages/Stats_and_facts.aspx
[10]http://cpstiers.opencityapps.org/about.html

2 PRELIMINARIES

We first describe a formal model for the allocation problem with diversity quotas. Throughout the paper, given $s \in \mathbb{N}$, we denote the set $\{1, 2, \ldots, s\}$ by $[s]$.

Definition 2.1 (AssignTC). An instance of the Assignment with Type Constraints (AssignTC) problem is given by:

(i) a set N of n agents partitioned into k types $N_1, \ldots, N_k$,

(ii) a set M of m items/goods partitioned into l blocks $M_1, \ldots, M_l$,

(iii) a utility $u(i, j) \in \mathbb{R}_+$ for each agent $i \in N$ and each item $j \in M$,

(iv) a capacity $\lambda_{pq} \in \mathbb{N}$ for all $(p, q) \in [k] \times [l]$, indicating the upper bound on the number of agents of type N_p allowed in the block M_q. ♦

Without loss of generality, we assume that the inequality $\lambda_{pq} \leq |M_q|$ holds for all type-block pairs $(p, q) \in [k] \times [l]$, since it is not possible to assign more than $|M_q|$ agents of type N_p to block M_q by definition. In general, agents types could be based on any criterion such as gender, profession, or geographical location. We consider the idealized scenario where we have a central planner who has access to the utilities of each agent for all items, and determines an assignment that maximizes social welfare under type-block constraints.

A few words about the type-block capacities are in order. Note that our analysis is agnostic to how these capacities are determined and just treats the vector $\{\lambda_{pq}\}_{p \in [k], q \in [l]}$ as a problem input.[11] Moreover, neither do we assume inequalities of the form $\lambda_{pq} \leq |N_p|$ nor is there any positive lower bound on the number of assignments for any type-block pair: this is in keeping with the actual HDB housing problem where λ_{pq}'s are fixed by policy (as percentages of block size) even before observing the applicant pool so that capacities larger than the size of an ethnic group are possible. Adding lower bounds *a priori* may render the problem infeasible if there not enough applicants of a certain type.[12]

An assignment of items to agents can be represented by a $(0, 1)$-matrix $X = (x_{ij})_{n \times m}$ where $x_{ij} = 1$ if and only if item j is assigned to agent i; a feasible solution is an assignment in which each item is allocated to at most one agent, and each agent receives at most one item, respecting the type-block capacities defined in (iv). We define the objective value (or total utility) as the utilitarian social welfare, i.e. the sum of the utilities of all agents in an assignment $u(X) \triangleq \sum_{i \in N} \sum_{j \in M} x_{ij} u(i, j)$. Clearly, this optimization problem can be formulated as the following integer linear program:

$$\max \quad \sum_{i \in N} \sum_{j \in M} x_{ij} u(i, j) \tag{1}$$

$$s.t. \quad \sum_{i \in N_p} \sum_{j \in M_q} x_{ij} \leq \lambda_{pq} \qquad \forall p \in [k], \forall q \in [l] \tag{2}$$

$$\sum_{j \in M} x_{ij} \leq 1 \qquad \forall i \in N \tag{3}$$

$$\sum_{i \in N} x_{ij} \leq 1 \qquad \forall j \in M \tag{4}$$

$$x_{ij} \in \{0, 1\} \qquad \forall i \in N, \forall j \in M \tag{5}$$

---

[11] The Singapore EIP percentage caps consider various factors such as "[t]he racial composition of the population[,] [. . .] the rate at which new households are being formed in each one of the racial groups and the present composition of applications" [Parliament of Singapore. Parliament Debates: Official Report. 1989], but these aspects of the problem are beyond the scope of the present work.

[12] Fragiadakis and Troyan [2017] show that, for some assignment problems with actual floor and ceiling constraints for each type-block pair, where the agent population is known beforehand and there is a guarantee that no agent remains unassigned, it is possible to reformulate the problem constraints in terms of "artificial caps" (modifying block sizes as well as type-block ceilings) and no floors: our analysis applies to these problems in this modified form.

where constraints (3-5) jointly ensure that X is a matching of items to agents, and inequalities (2) embody our type-block constraints.

Finally, an instance of the decision version of AssignTC consists of parameters (i) to (iv) in Definition 2.1, as well as a positive value U: it is a 'yes'-instance iff there exists a feasible assignment, satisfying constraints (2-5), whose objective value is at least U.

3 THE COMPLEXITY OF THE ASSIGNMENT PROBLEM WITH TYPE CONSTRAINTS

The hardness and approximation results in this section are based on a deep connection between AssignTC and the known NP-complete problem, Bounded Color Matching problem [Garey and Johnson 1979], defined as follows:

Definition 3.1 (BCMatching). An instance of the Bounded Color Matching (BCMatching) problem is given by (i) a bipartite graph $G = (A \cup B, E)$, where the set of edges E is partitioned into r subsets $E_1, \ldots, E_r$ representing the r different edge colors, (ii) a capacity $w_t \in \mathbb{N}$ for each color $t \in [r]$, (iii) a profit $\pi_e \in \mathbb{Q}_+$ for each edge $e \in E$, and (iv) a positive integer P called the threshold. It is a 'yes'-instance iff there exists a matching (i.e. a collection of pairwise non-adjacent edges) $E' \subseteq E$ such that the sum of the profits of all edges in the matching is at least P, and there are at most w_t edges of color t in it, i.e. $\sum_{e \in E'} \pi_e \geq P$ and $|E' \cap E_t| \leq w_t$ for all $t \in [r]$. ♦

The following lemma, together with its proof, establishes that AssignTC is, in fact, a special case of BCMatching.

Lemma 3.2. *There exists an S-reduction [Crescenzi 1997; Crescenzi et al. 1991] from the AssignTC problem to the BCMatching problem.*

Proof. Given an instance of the AssignTC problem, we construct a graph as follows. We define a node corresponding to each agent in N and also each item in M; we draw an edge from agent-node i to item-node j and make the associated profit equal to the utility $u(i, j)$ for all $i \in N, j \in M$. This gives us a complete bipartite graph with bipartition (N, M). We also give all edges joining agents of one type to items in one block the same color, one unique color for each agent-item pair. Thus, there are kl colors indexed lexicographically by pairs $(p, q) \in [k] \times [l]$. We set the capacity for color (p, q) at λ_{pq}. This produces, in $O(mn)$ time, an instance of BCMatching (Definition 3.1). The size of this instance is obviously polynomial in that of the original. By construction, there is a one-to-one correspondence between the sets of feasible solutions of the original and reduced instances with each corresponding pair having the same objective value (sum of edge-profits/utilities). Hence, the optimal values of the instances are also equal. □

This result will be useful in Section 3.1 for obtaining approximation guarantees. But it does not settle the question of the hardness of AssignTC since a special case of an NP-complete problem may be tractable. We could consider AssignTC and BCMatching to be identical if there existed a trivial reduction from BCMatching to AssignTC in the following sense. For a BCMatching instance, define an agent for each node in one part of the graph and an item for each node in the other; make the utility of an agent-item pair equal to the profit of the edge joining the corresponding pair of nodes; define types of agents and blocks of items such that each type-block pair corresponds to a unique edge-color. Unfortunately, the last part may not be possible depending on how the edges are colored, as illustrated by the following example.

Example 3.3. Consider the instance of the BCMatching problem defined on the graph in Figure 1. There are three colors: (a_1, b_1) and (a_1, b_2) are blue, (a_2, b_1) is red, and (a_2, b_2) is gray. The color-capacities, the edge-profits, and the threshold are arbitrary, hence omitted.
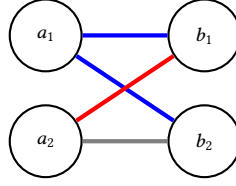
Fig. 1. An instance of BCMatching that does not trivially reduce to AssignTC.

Let the sets of agents and items be $N = \{a_1, a_2\}$ and $M = \{b_1, b_2\}$ respectively. Let us now try to define types and blocks that are consistent with edge-colors. Recall that, to achieve the desired reduction, no two edges joining a_1 to items in different blocks can have the same color. Since edges (a_1, b_1) and (a_1, b_2) have the same color, b_1 and b_2 must be in the same block. This also implies that all edges joining b_1 and b_2 to a_2 must be of the same color, regardless of whether a_1 and a_2 belong to the same type. However, (a_2, b_1) and (a_2, b_2) are of different colors – a contradiction. ∎

However, we will now show that there does exist a non-trivial polynomial-time reduction in the desired direction where edges of the graph map to agents (the colors corresponding to types) and items are defined in a more complicated way based on the nodes (the two parts of the bipartition corresponding to blocks)!

THEOREM 3.4. *The AssignTC problem is NP-complete.*

We will prove this by describing a polynomial-time reduction from BCMatching to the decision problem we introduced in Section 2.

PROOF. That the problem is in NP is immediate: given an assignment, one can verify in poly-time that it satisfies the problem constraints and compute total social welfare. Given an instance $\langle G; \vec{w}; \vec{\pi}; P \rangle$ of BCMatching, we construct an instance of the AssignTC problem as follows (see Example 3.5 for an illustration). Each edge $e \in E$ is an agent, whose type is its color. Items in our construction are partitioned into two blocks: M_1 and M_2. The items in block M_1 correspond to the vertices in B: there is one item j_b for each node $b \in B$. For every $a \in A$, we add $\deg(a) - 1$ items $j_a^1, \ldots, j_a^{\deg(a)-1}$ to M_2, for a total of $|E| - |A|$ items. Thus, there is a total of $m = |B| + |E| - |A|$ items. Block M_1 accepts at most w_p agents of type N_p, whereas block M_2 has unlimited type-block capacity; in other words, $\lambda_{p1} = w_p$ and $\lambda_{p2} = \min\{|N_p|, |M_2|\}$ for all $p \in [k]$. Given $e = (a, b)$, we define the utility function of agent e as follows:

$$u(e, j) = \begin{cases} \pi_e & \text{if } j = j_b, \\ \Phi & \text{if } j = j_a^s \text{ for some } s \in [\deg(a) - 1], \\ 0 & \text{otherwise.} \end{cases}$$

Here, Φ is an arbitrarily large constant, e.g. $\Phi = 1 + \sum_{e \in E} \pi_e$. Finally, let $U = P + \Phi(|E| - |A|)$; that is, our derived AssignTC instance is a "yes" instance iff there is some assignment of items to agents such that the social welfare exceeds U.

We begin by showing that if the original BCMatching instance is a 'yes' instance, then so is our constructed AssignTC instance. Let $E' \subseteq E$ be a valid matching whose value is at least P; let us construct an assignment X of items to agents via E' as follows. Observe some node $a \in A$; if $(a, b) \in E'$ then we assign the item $j_b \in M_1$ to the agent (a, b); the remaining $\deg(a) - 1$ agents of the form (a, b'), with $b' \in B$, are arbitrarily assigned to the items $j_a^1, \ldots, j_a^{\deg(a)-1} \in M_2$. If E' contains no edges adjacent to a, then we arbitrarily choose $\deg(a) - 1$ edges adjacent to

a and assign the corresponding agents to the items $j_a^1, \ldots, j_a^{\deg(a)-1}$. We now show that this indeed results in a valid assignment satisfying the type-block constraints.

First, by construction, every agent (a, b) is assigned at most one item. Moreover, since E' is a matching, every item $j_b \in M_1$ is assigned to at most one agent of the form (a, b); hence, every item in M_2 is assigned to at most one agent.

Let $E'_p = E_p \cap E'$ be the edges of color p in E'. Since the matching E' satisfies the capacity constraints of the BCMatching instance, we have $|E'_p| \le w_p$ for all $p \in [k]$; in particular, the number of items in M_1 assigned to agents of type p is no more than $w_p = \lambda_{p1}$. Thus, the type-block constraints for M_1 are satisfied. On the other hand, the type-block constraints for M_2 are trivially satisfied. We conclude that our constructed assignment is indeed valid, and satisfies the type-block constraints.

Finally, we want to show that total social welfare exceeds U the prescribed bound. Let us fix a node $a \in A$. By our construction, if the edge $e = (a, b)$ is in the matching E', then agent e is assigned the item j_b for a utility of π_e. Thus the total welfare of agents in E' equals $\sum_{e \in E'} \pi_e$, which is at least P by choice of E'. In addition, for every $a \in A$, there are exactly $\deg(a) - 1$ agents assigned to items in M_2 for a total utility of $\Phi(\deg(a) - 1)$. Summing over all $a \in A$, we have that the total utility derived by agents in $E \setminus E'$ is

$$\sum_{a \in A} \Phi(\deg(a) - 1) = \Phi\left(\sum_{a \in A} \deg(a) - \sum_{a \in A} 1\right) = \Phi(|E| - |A|).$$

Putting it all together, we have that the total utility obtained by our assignment is at least $P + \Phi(|E| - |A|) = U$.

Next, we assume that our constructed AssignTC instance is a 'yes' instance, and show that the original BCMatching instance must also be a 'yes' instance. Let X be a constrained assignment whose social welfare is at least $U = P + \Phi(|E| - |A|)$. Let E' be the set of edges corresponding to agents (a, b) assigned to items in M_1; we show that E' is a valid matching whose value is at least P. First, for any $b \in B$, X must assign the item j_b to at most one agent $e \in E'$. Next, since Φ is greater than the total utility obtainable from assigning all items in M_1, it must be the case that X assigns all items $j_a^1, \ldots, j_a^{\deg(a)-1}$ to $\deg(a) - 1$ agents of the form (a, b), with $b \in B$, for every node $a \in A$; thus, there can be one edge in E' that is incident on a for every $a \in A$. Next, since X satisfies the type-block constraints, we know that for every $p \in [k]$, there are at most $\lambda_{p1} = w_p$ agents from E_p that are assigned items in M_1; thus, E' satisfies the capacity constraints. Finally, the utility extracted from the agents assigned to items in M_2 is exactly $\Phi(|E| - |A|)$; the total utility of the matching X is at least $U = P + \Phi(|E| - |A|)$, thus E' has a total profit of at least P in the original BCMatching instance, and we are done. □

Example 3.5. In Figure 2, the graph $G = (A \cup B, E_1 \cup E_2)$, with $A = \{a_1, a_2\}$, $B = \{b_1, b_2, b_3\}$, $E_1 = \{(a_1, b_1), (a_2, b_2)\}$ and $E_2 = \{(a_1, b_2), (a_2, b_1), (a_2, b_3)\}$, is an instance of the BCMatching problem; edge labels are profits. The associated instance of the AssignTC problem is defined by $N = N_1 \cup N_2$ and $M = M_1 \cup M_2$, where $N_1 = \{(a_1, b_1), (a_2, b_2)\}$, $N_2 = \{(a_1, b_2), (a_2, b_1), (a_2, b_3)\}$, $M_1 = \{j_{b_1}, j_{b_2}, j_{b_3}\}$ and $M_2 = \{j_{a_1}^1, j_{a_2}^1, j_{a_2}^2\}$; the utility of an agent for an item is equal to 0 if there is no edge between them, to Φ if the edge is dashed, and to the edge label otherwise. ∎

3.1 Polynomial-Time Constant-Factor Approximation

Having established that the AssignTC problem is computationally intractable in general, we next present an efficient constant-factor approximation algorithm, based on a known polynomial-time approximation algorithm for the BCMatching problem.

THEOREM 3.6. *There exists a polynomial-time $\frac{1}{2}$-approximation algorithm for the AssignTC problem.*
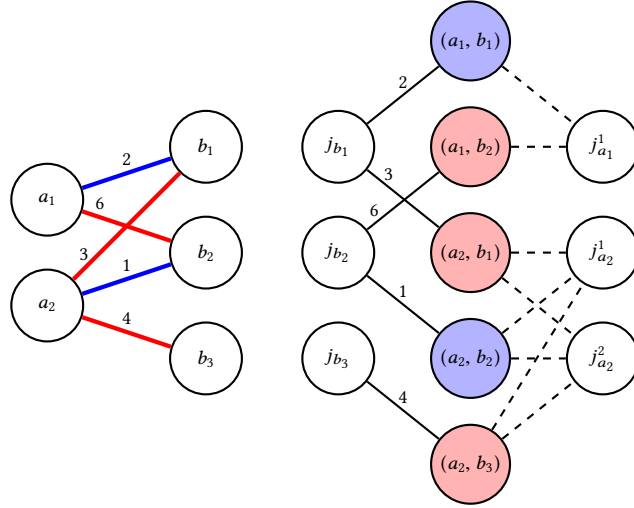
Fig. 2. A reduction from BCMATCHING to ASSIGNTC.

PROOF. Note that S-reduction is an approximation-preserving reduction [Crescenzi 1997; Orponen and Mannila 1987]. Thus, given an instance of the ASSIGNTC problem, we transform it into the corresponding BCMATCHING instance in accordance with Lemma 3.2, and then apply the polynomial-time $\frac{1}{2}$-approximation algorithm introduced by Stamoulis [2014] for BCMATCHING on general weighted graphs. □

Theorem 3.6 does not prove that $\frac{1}{2}$ is the best approximation-ratio possible for the ASSIGNTC problem. It is left for future work to investigate whether a better polynomial-time approximation algorithm exists.

3.2 Uniformity Breeds Simplicity: Polynomial-Time Special Cases

Our results thus far make no assumptions on agent-item utilities; as we now show, the ASSIGNTC problem admits a polynomial-time algorithm under some assumptions on the utility model.

Definition 3.7 (Type-uniformity and Block-uniformity). A utility model u is called *type-uniform* if all agents of the same type have the same utility for each item, i.e. for all $p \in [k]$ and for all $j \in M$, there exists $U_{pj} \in \mathbb{R}_+$ such that $u(i, j) = U_{pj}$ for all $i \in N_p$. A utility model u is called *block-uniform* if all items in the same block offer the same utility to every agent; that is, for all $q \in [l]$ and for all $i \in N$, there exists $U_{iq} \in \mathbb{R}_+$ such that $u(i, j) = U_{iq}$ for all $j \in M_q$. ♦

In the context of the HDB allocation problem, type uniformity implies that Singaporeans of the same ethnicity share the same preferences over apartments (perhaps due to cultural or socioeconomic factors). Cases that deal with uniform goods satisfy the block-uniformity assumption: e.g. students applying for spots in public schools or job applicants applying for multiple (identical) positions; in the HDB domain, block-uniformity captures purely location-based preferences, i.e. a tenant does not care which apartment she gets as long as it is in a specific block close to her workplace, family, or favorite public space.

THEOREM 3.8. *The ASSIGNTC problem can be solved in poly(n, m) time under either a type-uniform or a block-uniform utility model.*

We prove the result for a type-uniform utility model; the result for block-uniform utilities can be similarly derived. We propose a polynomial time algorithm based on the *Minimum-Cost Flow* problem which is known to be solvable in polynomial time. Recall that a flow network is a directed graph $G = (V, E)$ with a source node $s \in V$ and a sink node $t \in V$, where each arc $(a, b) \in E$ has a cost $\gamma(a, b) \in \mathbb{R}$ and a capacity $\psi(a, b) > 0$ representing the maximum amount that can flow on the arc; for convenience, we set $\gamma(a, b) = 0$ and $\psi(a, b) = 0$ for all $a, b \in V$ such that $(a, b) \notin E$. Let us denote by Γ and Ψ the matrices of costs and capacities respectively defined by $\Gamma = (\gamma(a, b))_{|V| \times |V|}$ and $\Psi = (\psi(a, b))_{|V| \times |V|}$. A flow in the network is a function $f : V \times V \to \mathbb{R}_+$ satisfying:

(i) $f(a, b) \le \psi(a, b)$ for all $a, b \in V$ (capacity constraints),
(ii) $f(a, b) = -f(b, a)$ for all $a, b \in V$ (skew symmetry), and
(iii) $\sum_{b \in V} f(a, b) = 0$ for all $a \in V \setminus \{s, t\}$ (flow conservation).

The value $v(f)$ of a flow f is defined by $v(f) = \sum_{a \in V} f(s, a) = \sum_{a \in V} f(a, t)$ and its cost is given by $\gamma(f) = \sum_{(a,b) \in E} f(a, b)\gamma(a, b)$. The optimization problem can be formulated as follows. Given a value F, find a flow f that minimizes the cost $\gamma(f)$ subject to $v(f) = F$. This optimization problem that takes as input the graph $G = (V, E)$, the matrices Γ and Ψ, and the value F, will be denoted by MinCostFlow hereafter; given an instance $\langle G; \Gamma; \Psi; F \rangle$ of the MinCostFlow problem, we let $\gamma(G, \Gamma, \Psi, F)$ be the cost of the optimal flow for that instance.

Given an instance $\mathcal{I}$ of AssignTC, we construct a flow network $G_{\mathcal{I}}(V, E)$ and matrices $\Gamma_{\mathcal{I}}$ and $\Psi_{\mathcal{I}}$ as follows (see Figure 3 for an illustration). The node set V is partitioned into layers: $V = \{s\} \cup A \cup B \cup C \cup \{t\}$. A is the *agent type layer*: there is one node $a_p \in A$ for all agent types $N_p, p \in [k]$. B is the *type-block layer*: it has a node $b_{pq} \in B$ for every type-block pair $(p, q) \in [k] \times [l]$. Finally, C is the *item layer*: there is one node $c_j \in C$ for all items $j \in M$. The arcs in E are as follows: for every $a_p \in A$, there is an arc from s to a_p whose capacity $\psi(s, a_p)$ is $|N_p|$. Fixing $p \in [k]$, there is an arc from $a_p \in A$ to every $b_{pq} \in B$, where the capacity of (a_p, b_{pq}) is the quota for type N_p in block M_q (i.e. $\psi(a_p, b_{pq}) = \lambda_{pq}$). Finally, given $q \in [l]$, there is an arc from b_{pq} to c_j iff $j \in M_q$; in that case, we have $\psi(b_{pq}, c_j) = 1$. The costs associated with arcs from B to C (i.e. arcs of the form (b_{pq}, c_j) where $j \in M_q$) are $-U_{pj}$; recall that U_{pj} is the utility that every agent of type N_p assigns to item j. All other arc costs are set to 0. We begin by proving a few technical lemmas on the above network.

Given a positive integer F, there exists an optimal flow that is integer-valued since $\langle G_{\mathcal{I}}; \Gamma_{\mathcal{I}}; \Psi_{\mathcal{I}}; F \rangle$ is integer-valued as well. Let f^* be an integer-valued optimal flow, taken over all possible values of F; that is:

$$f^* \in \operatorname*{argmin}_{F \in [n]} \gamma(G_{\mathcal{I}}, \Gamma_{\mathcal{I}}, \Psi_{\mathcal{I}}, F) \tag{6}$$

Finding the flow f^* involves solving n instances of MinCostFlow by definition; thus, one can find f^* in polynomial time. Given f^* as defined in (6), let $X^* = (x_{ij}^*)_{n \times m}$ be defined as follows: for every item $j \in M_q$, if $f^*(b_{pq}, c_j) = 1$ for some $p \in [k]$, then we choose an arbitrary unassigned agent $i \in N_p$ and set $x_{ij}^* = 1$.

LEMMA 3.9. *X^* is a feasible solution of the AssignTC instance $\mathcal{I}$.*

PROOF. First, we assign at most one item to every agent by construction; next, let us show that each item $j \in M_q$ is assigned to at most one agent. Since f^* is a flow, we have $\sum_{p=1}^{k} f^*(b_{pq}, c_j) = f^*(c_j, t)$ due to flow conservation; note that the capacity of the arc (c_j, t) is 1, thus at most one arc (b_{pq}, c_j) has $f^*(b_{pq}, c_j) = 1$. Finally, since item j is assigned to an agent in N_p iff $f^*(b_{pq}, c_j) = 1$, we conclude that item j is assigned to at most one of the agents in N.

Next, let us prove that assignment X^* satisfies the type-block constraints; in other words, we need to show that:

$$\sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* \leq \lambda_{pq}, \forall p \in [k], \forall q \in [l] \tag{7}$$

Since f^* is a flow, we have $f^*(a_p, b_{pq}) = \sum_{j \in M_q} f^*(b_{pq}, c_j)$ for every type-block pair $(p, q) \in [k] \times [l]$ due to flow conservation; moreover, we have $f^*(a_p, b_{pq}) \leq \psi(b_{pq}, c_j) = \lambda_{pq}$ by construction. As a consequence, we necessarily have $\sum_{j \in M_q} f^*(b_{pq}, c_j) \leq \lambda_{pq}$ for all $p \in [k]$. Since an item $j \in M_q$ is matched with some agent $i \in N_p$ if and only if we have $f^*(b_{pq}, c_j) = 1$, we conclude that (7) indeed holds. $\qquad\square$

Now, let us establish a relation between the cost of f^* and the utility of the feasible assignment X^*.

LEMMA 3.10. *The cost of the flow f^* satisfies $\gamma(f^*) = -u(X^*)$.*

PROOF. By construction, the cost of f^* can only be induced by arcs from nodes in B to nodes in C, where the cost of all arcs of the form (b_{pq}, c_j), with $j \in M_q$, is equal to $-U_{pj}$ (the negative of the uniform utility derived from item j by members of N_p). In other words, the cost of f^* can be written as follows:

$$\gamma(f^*) = -\sum_{p=1}^{k} \sum_{q=1}^{l} \sum_{j \in M_q} f^*(b_{pq}, c_j) U_{pj}$$

As previously argued, we have that $f^*(b_{pq}, c_j) \in \{0, 1\}$ for all arcs (b_{pq}, c_j); moreover, $f^*(b_{pq}, c_j) = 1$ iff item j is assigned to some agent in N_p. Therefore, we obtain:

$$\gamma(f^*) = -\sum_{p=1}^{k} \sum_{i \in N_p} \sum_{j \in M} x_{ij}^* U_{pj} = -\sum_{i \in N} \sum_{j \in M} x_{ij}^* u(i, j) = -u(X^*)$$

where the second equality holds since all agents in N_p have the same utility by assumption. $\qquad\square$

Finally, we show that for every feasible solution to the AssignTC instance $\mathcal{I}$, there exists a flow with a matching cost.

LEMMA 3.11. *Let X be a feasible assignment for the AssignTC instance $\mathcal{I}$; there exists some feasible flow f such that $\gamma(f) = -u(X)$. Moreover, we have $v(f) = |\{i \in N : \sum_{j \in M} x_{ij} = 1\}|$.*

PROOF. Given a feasible assignment $X = (x_{ij})_{n \times m}$, we define $f : V \times V \rightarrow \mathbb{R}_+$ as follows:

$$\begin{cases} f(s, a_p) = \sum_{i \in N_p} \sum_{j \in M} x_{ij} & \forall a_p \in A \\ f(a_p, b_{pq}) = \sum_{i \in N_p} \sum_{j \in M_q} x_{ij} & \forall (a_p, b_{pq}) \in E \\ f(b_{pq}, c_j) = \sum_{i \in N_p} x_{ij} & \forall (b_{pq}, c_j) \in E \\ f(c_j, t) = \sum_{i \in N} x_{ij} & \forall c_j \in C \\ f(a, b) = -f(b, a) & \forall (a, b) \in E \\ f(a, b) = 0 & \forall (a, b) \notin E \end{cases}$$

The function f is indeed a flow: f trivially satisfies the skew symmetry condition by construction; next, we show that f satisfies flow conservation. For all $a_p \in A$, the incoming flow to node a_p from node s is $f(s, a_p) = \sum_{i \in N_p} \sum_{j \in M} x_{ij}$, and the outgoing flow to every b_{pq} is $\sum_{q=1}^{l} f(a_p, b_{pq}) = \sum_{i \in N_p} \sum_{j \in M} x_{ij}$ since M is partitioned into $M_1, \ldots, M_l$; hence flow is conserved. For a node $b_{pq} \in B$, the incoming flow equals $f(a_p, b_{pq}) = \sum_{i \in N_p} \sum_{j \in M_q} x_{ij}$ and an amount of
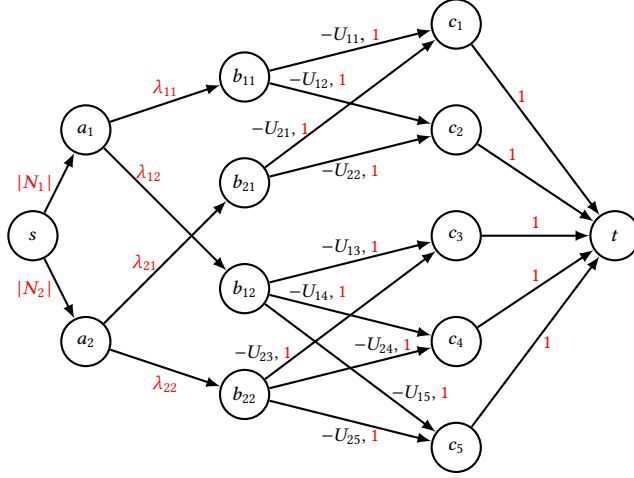
Fig. 3. Network flow constructed for the proof of Theorem 3.8; in this case, we have 2 types and 2 blocks: $M_1 = \{1, 2\}$ and $M_2 = \{3, 4, 5\}$. Arc capacities are given in red. All arcs have a cost of 0, except those between $b_{pq} \in B$ and $c_j \in C$ whose cost equals $-U_{pj}$.

$f(b_{pq}, c_j) = \sum_{i \in N_p} x_{ij}$ flows to every node c_j such that $j \in M_q$, thus flow is conserved. For a node $c_j \in C$ such that $j \in M_q$, its incoming flow equals $f(b_{pq}, c_j) = \sum_{i \in N_p} x_{ij}$ from every b_{pq}, for a total flow of $\sum_{p=1}^{k} \sum_{i \in N_p} x_{ij}$, which equals its outgoing flow to t. To conclude, f satisfies flow conservation.

Now let us prove that f satisfies the capacity constraints (i.e. $f(a, b) \leq \psi(a, b)$ for all arcs $(a, b) \in E$). For all $(s, a_p) \in E$, we have $f(s, a_p) = \sum_{i \in N_p} \sum_{j \in M} x_{ij} \leq |N_p| = \psi(s, a_p)$ since every agent $i \in N_p$ is matched with at most one item. For all $(a_p, b_{pq}) \in E$, we have $f(a_p, b_{pq}) = \sum_{i \in N_p} \sum_{j \in M_l} x_{ij} \leq \lambda_{pq} = \psi(a_p, b_{pq})$ since X satisfies the type-block constraints. For all arcs $(b_{pq}, c_j) \in E$, we have $f(b_{pq}, c_j) = \sum_{i \in N_p} x_{ij} \leq 1 = \psi(b_{pq}, c_j)$ since item j is matched with at most one of the agents in N_p. For all $(c_j, t) \in E$, we have $f(c_j, t) = \sum_{i \in N} x_{ij} \leq 1 = \psi(c_j, t)$ since item j is matched with at most one of the agents in N. Hence, f satisfies the capacity constraints and is a valid flow. Note that we have:

$$v(f) = \sum_{a \in V} f(s, a) = \sum_{p=1}^{k} f(s, a_p) = \sum_{p=1}^{k} \sum_{i \in N_p} \sum_{j \in M} x_{ij} = \sum_{i \in N} \sum_{j \in M} x_{ij}$$

Then, since X is a feasible assignment of the AssignTC instance $\mathcal{I}$, we conclude that we have $v(f) = |\{i \in N : \sum_{j \in M} x_{ij} = 1\}|$. We just need to prove that we have $\gamma(f) = -u(X)$, and we are done. By definition of the flow network, only arcs of the form (b_{pq}, c_j) contribute to the cost $\gamma(f)$ and we have $\gamma(b_{pq}, c_j) = -U_{pj}$; therefore, $\gamma(f) = -\sum_{(b_{pq}, c_j) \in E} f(b_{pq}, c_j) U_{pj}$. Since $f(b_{pq}, c_j) = \sum_{i \in N_p} x_{ij}$ (by definition of f) and $u(i, j) = U_{pj}$ for all agents $i \in N_p$ (by hypothesis), we finally obtain $\gamma(f) = -\sum_{j \in M} \sum_{p=1}^{k} \sum_{i \in N_p} x_{ij} u(i, j) = -\sum_{j \in M} \sum_{i \in N} x_{ij} u(i, j) = -u(X)$.   □

We are now ready to prove Theorem 3.8.

Proof of Theorem 3.8. We begin by observing the flow f^* as defined in (6), and the assignment X^* derived from it. First, according to Lemma 3.9, X^* is a feasible assignment of the AssignTC instance $\mathcal{I}$. Moreover, we have $u(X^*) = -\gamma(f^*)$ according to Lemma 3.10. Finally, for any feasible assignment X of the AssignTC instance $\mathcal{I}$, there exists a flow f such that $\gamma(f) = -u(X)$; furthermore, since $v(f) = |\{i \in N : \sum_{j \in M} x_{ij} = 1\}| \in [n]$, flow f is a feasible

solution of the MinCostFlow instance $\langle G_{\mathcal{I}}; \Gamma_{\mathcal{I}}; \Psi_{\mathcal{I}}; F \rangle$ for some $F \in [n]$. Therefore, we have:

$$u(X) = -\gamma(f) \le -\gamma(G_{\mathcal{I}}, \Gamma_{\mathcal{I}}, \Psi_{\mathcal{I}}, v(f)) \le -\gamma(f^*) = u(X^*)$$

Thus, X^* is an optimal solution of the AssignTC instance $\mathcal{I}$; since X^* can be computed in poly-time (Proposition 6), we are done.   □

4   THE PRICE OF DIVERSITY

We now turn to the allocative efficiency of the constrained assignment. As before, an instance of the AssignTC problem is given by a set of n agents N partitioned into types $N_1, \ldots, N_k$, a set of m items M partitioned into $M_1, \ldots, M_l$, a list of capacity values $(\lambda_{pq})_{k \times l}$, and agent utilities for items given by $u = (u(i, j))_{n \times m}$. We denote the set of all assignments X of items to agents satisfying only the matching constraints (3-5) of Section 2 by $\mathcal{X}$, and that of all assignments additionally satisfying the type-block constraints (2) by $\mathcal{X}_C$; the corresponding optimal social welfares for any given utility matrix $(u(i, j))_{n \times m}$ are:

$$OPT(u) \triangleq \max_{X \in \mathcal{X}} u(X); \; OPT_C(u) \triangleq \max_{X \in \mathcal{X}_C} u(X).$$

Clearly, $OPT_C(u) \le OPT(u)$ since $\mathcal{X}_C \subseteq \mathcal{X}$; we define the following natural measure of this welfare loss that lies in $[1, \infty]$:

Definition 4.1. For any instance of the AssignTC problem, we define the *Price of Diversity* as follows, along the lines of Ahmed et al. [2017] and Bredereck et al. [2018]:   ♦

$$PoD(u) \triangleq \frac{OPT(u)}{OPT_C(u)}.$$

The main result of this section is to establish an upper bound on $PoD(u)$ that is independent of the utility model. Denote the ratio of a type-block capacity to the size of the corresponding block by:

$$\alpha_{pq} \triangleq \frac{\lambda_{pq}}{|M_q|}.$$

THEOREM 4.2.   *For any instance of AssignTC, we have:*

$$PoD(u) \le \frac{1}{\min_{(p,q) \in [k] \times [l]} \alpha_{pq}}$$

and the above upper bound is tight.

In general, the bound in Theorem 4.2 can grow linearly in m; in the following family of problem instances where the capacities λ_{pq} are fixed constants, the *PoD* can be indefinitely large.

Example 4.3. Consider any instance of the AssignTC problem with $l = k$ and $|N_p| = |M_p| = \mu \; \forall p \in [k]$ so that $n = m = k\mu$; the utilities are:

$$u(i, j) = \begin{cases} 1 & \text{if } i \in N_p \text{ and } j \in M_p \quad \forall p \in [k], \\ 0 & \text{otherwise.} \end{cases}$$

Evidently, any complete matching of items in M_p to agents in $N_p \; \forall p \in [k]$ is an optimal solution for the unconstrained version of the problem, hence $OPT(u) = k\mu$. But if the capacities are $\lambda_{pq} = 1 \; \forall (p, q) \in [k] \times [k]$ then only one agent per group can receive an item for which she has non-zero utility, hence $OPT_C(u) = k$. Thus, $PoD = k\mu/k = \mu$.   ■

However, type-block capacities are determined by a central planner in our model; a natural way of setting them is to fix the proportional capacities or quotas α_{pq} in advance, and then compute $\lambda_{pq} = \alpha_{pq} \times |M_q|$ when block sizes become available: by committing to a fixed minimum type-block quota α^* (i.e. $\alpha_{pq} \geq \alpha^*$ for all $(p, q) \in [k] \times [l]$), the planner can ensure a $PoD(u)$ of at most $1/\alpha^*$, regardless of the problem size and utility function. Higher values of α^* reduce the upper bound on $PoD(u)$ but also increase the capacity of a block for every ethnicity, potentially affecting the diversity objective adversely: it thus functions as a tunable tradeoff parameter between ethnic integration and worst-case welfare loss. In fact, in the Singapore allocation problem, the Ethnic Integration Policy fixes a universal percentage cap for each of the three ethnicities in all blocks; these percentages are set slightly higher than the actual respective population proportions: the current block quotas α_{pq} are 0.87 for Chinese, 0.25 for Malays and 0.15 for Indian/Others [Deng et al. 2013]. Hence, for the Singapore housing system, we have $\min_{(p, q) \in [k] \times [l]} \alpha_{pq} = 0.15$ which is achieved for Indian/Others and any block, so that from Theorem 4.2,

$$PoD(u) \leq \frac{1}{0.15} \approx 6.67.$$

This bound makes no assumptions on agent utilities; in other words, it holds under *any utility model*.[13]

The proof relies on the following lemma. Given an assignment $X \in \mathcal{X}$, let $u_p(X)$ denote the total utility of agents in N_p under X:

$$u_p(X) \triangleq \sum_{i \in N_p} \sum_{j \in M} x_{ij} u(i, j) = \sum_{q \in [l]} \sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i, j). \tag{8}$$

LEMMA 4.4. *For any instance of AssignTC and any optimal unconstrained assignment $X^* \in \mathcal{X}$, we have:*

$$PoD(u) \leq \frac{u(X^*)}{\sum_{p \in [k]} u_p(X^*) \min_{q \in [l]} \alpha_{pq}}.$$

PROOF. Based on the optimal assignment X^*, we can construct an assignment $X \in \mathcal{X}_C$ satisfying the type-block constraints, by carefully 'revoking' the smallest-utility items in M_q from agents in N_p for every (p, q)-pair that violates the corresponding type-block constraint. In other words, let n_{pq} denote the number of items in M_q assigned to agents in N_p under X^*. If $n_{pq} \leq \lambda_{pq}$, we leave that type-block pair untouched, so that $\sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i, j) = \sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i, j)$. If $n_{pq} > \lambda_{pq}$, we order these n_{pq} agents according to their utilities for the items they are assigned and retain only the top λ_{pq} agents in that order (breaking ties lexicographically), setting $x_{ij} = 0$ for the remaining agents. This change increases the average utility of assignments for this type-block pair:

$$\frac{\sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i, j)}{\lambda_{pq}} \geq \frac{\sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i, j)}{n_{pq}}.$$

(To see why this is true, consider a sequence $z_1 \geq z_2 \geq \ldots$, where $z_i \geq 0$ and $z_i \geq z_{i+1}$ $\forall i = 1, 2, \ldots$, and two positive integers $\nu > \mu \geq 1$. Clearly, $(\nu - \mu) \sum_{i=1}^{\mu} z_i \geq (\nu - \mu)\mu z_\mu \geq \mu \sum_{i=\mu+1}^{\nu} z_i$ since $z_i \geq z_\mu$ $\forall i = 1, 2, \ldots, \mu$ and $z_i \leq z_\mu$ $\forall i = \mu + 1, \mu + 2, \ldots, \nu$. Rearranging and simplifying, we get $\frac{1}{\mu} \sum_{i=1}^{\mu} z_i \geq \frac{1}{\nu} \sum_{i=1}^{\nu} z_i$.)

---

[13]In practice, the *effective* value of each fractional capacity $\lambda_{pq}/|M_q|$ might be smaller than the corresponding pre-specified fraction α_{pq}. Since each λ_{pq} must be an integer, we need to set $\lambda_{pq} = \lfloor \alpha_{pq} \times |M_q| \rfloor$ $\forall (p, q) \in [k] \times [l]$ to respect all capacity constraints. Hence, for a given instance of AssignTC, the effective upper bound on $PoD(u)$ is given by $1/\min_{p,q} \frac{\lfloor \alpha_{pq} \times |M_q| \rfloor}{|M_q|}$, which depends on the $|M_q|$-values and may be higher than $1/\min_{p,q} \alpha_{pq}$. For example, if we have a uniform block size of 10, the actual numerical capacities for Chinese, Malay, and Indian/Others based on EIP quotas become 8, 2, and 1 respectively, so that the effective PoD-bound is 10. However, the effective bound is still independent of utility values as well as the population size of agents of any type; moreover, larger block sizes reduce the discrepancy between the effective bound and the theoretical bound $1/\min_{p,q} \alpha_{pq}$ provided by Theorem 4.2. For example, for the values of α_{pq} and $|M_q|$ used in our experiments in Section 5 (see Figure 5), the minimum effective capacity of any block for any type is 14.42%, hence the effective upper bound on $PoD(u)$ is 6.93. Similar considerations apply to Theorem 4.6.

Further, since $n_{pq} \leq |M_q|$, the above inequality implies that

$$\frac{\sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i,j)}{\lambda_{pq}} \geq \frac{\sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i,j)}{|M_q|}$$

$$\implies \sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i,j) \geq \frac{\lambda_{pq}}{|M_q|} \sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i,j)$$

$$= \alpha_{pq} \sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i,j), \quad \text{since } \alpha_{pq} = \frac{\lambda_{pq}}{|M_q|}.$$

Thus, for every $p \in [k]$ and every $q \in [l]$, we have

$$\sum_{i \in N_p} \sum_{j \in M_q} x_{ij} u(i,j) \geq \left(\min_{q \in [l]} \alpha_{pq} \right) \sum_{i \in N_p} \sum_{j \in M_q} x_{ij}^* u(i,j), \quad \text{since } \min_{q \in [l]} \alpha_{pq} \leq \alpha_{pq} \leq 1.$$

Summing over blocks, we obtain from Equation (8):

$$u_p(X) \geq u_p(X^*) \min_{q \in [l]} \alpha_{pq}, \; \forall p \in [k].$$

By definition, $u(X^*) = OPT(u)$. Moreover, since $X \in \mathcal{X}_C$, we have $u(X) \leq OPT_C(u)$. Hence, by Definition 4.1,

$$PoD(u) \leq \frac{u(X^*)}{u(X)} \leq \frac{u(X^*)}{\sum_{p \in [k]} u_p(X^*) \min_{q \in [l]} \alpha_{pq}}.$$

$\square$

We can now complete the proof of the theorem.

PROOF OF THEOREM 4.2. Since we have $\min_{(p,q) \in [k] \times [l]} \alpha_{pq} \leq \min_{q \in [l]} \alpha_{p'q}$ for all $p' \in [k]$, Lemma 4.4 implies that:

$$PoD(u) \leq \frac{u(X^*)}{\sum_{p \in [k]} u_p(X^*) \min_{(p,q) \in [k] \times [l]} \alpha_{pq}} = \frac{1}{\min_{(p,q) \in [k] \times [l]} \alpha_{pq}}.$$

Depending on the utility matrix u, this upper bound can be tight whenever $|N_{p_0}| \geq |M_{q_0}|$ for some type-block pair (p_0, q_0) in the set $\text{argmin}_{(p,q) \in [k] \times [l]} \alpha_{pq}$. We identify an agent utility matrix for which the bound holds with equality:

$$u(i,j) = \begin{cases} 1 & \text{if } i \in N_{p_0} \text{ and } j \in M_{q_0}, \\ 0 & \text{otherwise.} \end{cases}$$

The optimal assignment without type-block constraints fully allocates the items in block M_{q_0} to agents in N_{p_0} for a total utility of $|M_{q_0}|$; furthermore, we know that any optimal constrained assignment allocates exactly $\lambda_{p_0 q_0}$ items in M_{q_0} to agents in N_{p_0} for a total utility of $\lambda_{p_0 q_0}$. Since $\lambda_{p_0 q_0} = \alpha_{p_0 q_0} \times |M_{q_0}|$, we have:

$$PoD(u) = \frac{|M_{q_0}|}{\alpha_{p_0 q_0} \times |M_{q_0}|} = \frac{1}{\alpha_{p_0 q_0}} = \frac{1}{\min_{(p,q) \in [k] \times [l]} \alpha_{pq}}.$$

$\square$

4.1 The Impact of Disparity among Types

Theorem 4.2 offers a worst-case tight bound on the price of diversity, making no assumptions on agent utilities. However, its proof suggests that this upper bound is attained when social welfare is solely extracted from a single agent type and

a single block. Intuitively, we can obtain a better bound on the price of diversity if a less 'disparate' optimal assignment exists. To formalize this notion, we introduce a new parameter:

Definition 4.5. For an optimal unconstrained assignment $X^* \in \mathcal{X}$, denote by $\beta_p(X^*)$ the ratio of the average utility of agents in N_p to the average utility of all agents under X^*. The *inter-type disparity parameter* $\beta(X^*)$ is defined as: ♦

$$\beta(X^*) \triangleq \min_{p \in [k]} \beta_p(X^*) = \min_{p \in [k]} \frac{u_p(X^*)/|N_p|}{u(X^*)/n}.$$

Notice that $\beta(X^*) \in (0, 1]$ can be computed in polynomial time and is fully independent of the type-block capacities. The closer $\beta(X^*)$ is to 1, the lower the disparity between average agents of different types under X^*.

THEOREM 4.6. *For any AssignTC instance and any unconstrained optimal assignment $X^* \in \mathcal{X}$, we have:*

$$PoD(u) \leq \frac{1/\beta(X^*)}{\sum_{p \in [k]} v_p \min_{q \in [l]} \alpha_{pq}},$$

where $v_p = \frac{|N_p|}{n}$ is the proportion of type p in the agent population, for every $p \in [k]$.

PROOF. By definition of $\beta(X^*)$, for every $p \in [k]$, we have:

$$u_p(X^*) \geq \beta(X^*) \frac{|N_p|}{n} u(X^*) = \beta(X^*) v_p u(X^*).$$

Substituting this in Lemma 4.4, we obtain the desired bound.   □

Let us now apply the result to the Singapore public housing domain; we use the ethnic proportions reported in the 2010 census report [Department of Statistics, Singapore 2010] to obtain $|N_1|/n = 0.741$ (Chinese), $|N_2|/n = 0.134$ (Malay), and $|N_3|/n = 0.125$ (Indian/Others). Using the same block quotas α_{pq} as before, we have:

$$PoD(u) \leq \frac{1/\beta(X^*)}{0.87 \times 0.741 + 0.25 \times 0.134 + 0.15 \times 0.125} \approx \frac{1.43}{\beta(X^*)}.$$

In other words, if inter-type disparity is low, i.e. $\beta(X^*)$ is close to 1, the *PoD* may be significantly lower than the bound provided by Theorem 4.2.[14]

The two upper bounds provided by Theorems 4.2 and 4.6 are incomparable due to the dependence of the latter on the parameter $\beta(X^*)$. We will elaborate on this point with the help of two examples (4.7 and 4.8) in each of which every agent has a positive utility for every item (unlike Example 4.3).

In each example, we have only one block of items $M = [m]$ and two types $N_1 = \{a_1, a_2, \ldots, a_m\}$ and $N_2 = \{b_1, b_2, \ldots, b_m\}$ for an arbitrary positive integer $m > 1$. Hence the proportion of each type in the population is $\frac{1}{2}$. Let the proportional capacities of the single block for the two types be α_1 and α_2 respectively, such that $\alpha_1 + \alpha_2 \geq 1$ and $\alpha_1 m$ and $\alpha_2 m$ are both integers. Thus, for each of these examples, Theorem 4.2 puts the upper bound on the price of diversity at $\frac{1}{\min\{\alpha_1, \alpha_2\}}$. In the first example, the bound based on the inter-type disparity parameter turns out to be useless/uninformative.

---

[14]If there are multiple optimal assignments, they may have different values of the inter-type disparity parameter; if that is the case, we should choose the largest of these values for computing the upper bound provided by Theorem 4.6 since the inequality holds for *any* unconstrained optimal assignment.

Example 4.7. For some $\varepsilon \ll 1 - \frac{1}{m}$, let the utilities be

$$u(a_r, j) = \begin{cases} 1 - \varepsilon & \text{if } j = r, \\ \frac{\varepsilon}{m-1} & \text{otherwise} \end{cases} \qquad \forall r \in [m];$$

$$u(b_r, j) = \frac{1}{m} \qquad \forall j \in M, \forall r \in [m].$$

Evidently, the unique unconstrained optimal assignment is to match item j with agent a_j for every $j \in [m]$. Hence, per-agent average utilities of N_1 and N_2 are $1 - \varepsilon$ and 0 respectively, making the inter-type disparity parameter zero. Thus, Theorem 4.6 does not place any finite upper bound on $PoD(u)$.

However, a little thought reveals that, in a constrained optimal allocation, $\alpha_1 m$ items are allocated to N_1 such that each of these items j is assigned to the agent a_j, and the remaining $(1 - \alpha_1)m$ items are allocated to N_2 and arbitrarily assigned to one agent each. Since $(1 - \alpha_1)m \leq \alpha_2 m$, all type-block capacities are satisfied. Thus,

$$PoD(u) = \frac{m(1 - \varepsilon)}{\alpha_1 m(1 - \varepsilon) + (1 - \alpha_1)m \cdot \frac{1}{m}} = \frac{1}{\alpha_1 + \psi \cdot (1 - \alpha_1)},$$

where $\psi = \frac{1}{m(1-\varepsilon)} \in (0, 1)$. Since $\alpha_1 < 1$, the denominator exceeds α_1, so that $PoD(u) < \frac{1}{\alpha_1}$. If $\alpha_1 > \alpha_2$, then $\alpha_1 + \psi \cdot (1 - \alpha_1) = \psi + (1 - \psi)\alpha_1 > \psi + (1 - \psi)\alpha_2 = \alpha_2 + \psi \cdot (1 - \alpha_2) > \alpha_2$; hence $PoD(u) < \frac{1}{\alpha_2}$. In any case, $PoD(u) \leq \max\{\frac{1}{\alpha_1}, \frac{1}{\alpha_2}\} = \frac{1}{\min\{\alpha_1, \alpha_2\}}$, i.e. the realized price of diversity respects the Theorem 4.2 bound. ∎

In the next example, the parametrized bound is more informative than the other.

Example 4.8. We will further assume that m is even. Let $u(a_r, j) = u(b_r, j) = \frac{1}{m}$ for every $r \in [m], j \in [m]$. There is an unconstrained optimal assignment which also achieves $\beta(X^*) = 1$: assign $\frac{m}{2}$ items arbitrarily to $\frac{m}{2}$ agents of each type, giving $OPT(u) = 1$. Thus, the Theorem 4.6 bound $\frac{1}{(\alpha_1 + \alpha_2)/2} \leq \frac{1}{\min\{\alpha_1, \alpha_2\}}$, the Theorem 4.2 bound. Since $\alpha_1 + \alpha_2 \geq 1$, there are at least two assignments, respecting capacity constraints, whose social welfare is $OPT(u) = 1$: either $\alpha_1 m$ to N_1 and $(1 - \alpha_1)m$ to N_2, or $(1 - \alpha_2)m$ to N_1 and $\alpha_2 m$ to N_2. In either case, the constrained optimum is 1; hence $PoD(u)$ has its ideal value of 1. ∎

Finally, combining Theorems 4.2 and 4.6, we obtain the following upper bound on the price of diversity of any instance of AssignTC:

$$PoD(u) \leq \min \left\{ \frac{1}{\min_{(p,q) \in [k] \times [l]} \alpha_{pq}}, \frac{1/\beta(X^*)}{\sum_{p \in [k]} v_p \min_{q \in [l]} \alpha_{pq}} \right\}. \tag{9}$$

Thus, if we plot the $PoD(u)$ against the disparity parameter $\beta(X^*)$, the point corresponding to any AssignTC instance with block quotas and ethnic proportions as in Singapore must lie in the shaded region of Figure 4.

5 EXPERIMENTAL ANALYSIS

In this section, we simulate instances of the AssignTC problem using recent, publicly available Singaporean demographic and housing allocation statistics and Chicago public school admission data. We compare the welfare of three assignment mechanisms: the optimal unconstrained mechanism, the optimal constrained mechanism, and the lottery-based mechanism (see Section 5.1 below). Both the unconstrained and constrained social welfare maximizations are solved using the Gurobi Optimizer.[15] We refer the reader to https://git.io/fNhhm for full implementation details.

---

[15] Each iteration of the constrained optimization problem took about 30 minutes while that of the unconstrained optimization and the lottery mechanism took less than 1 minute of running time on a standard laptop (Intel i7-7600U Processor, 2.8Ghz, 8GB RAM, 256 GB SSD).
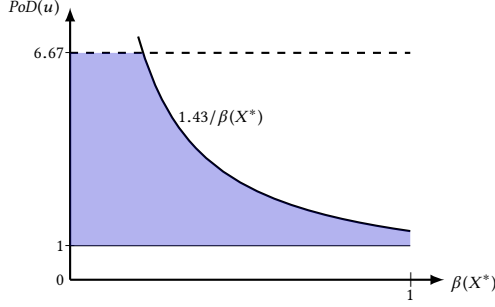
Fig. 4. *PoD* vs disparity parameter for the HDB problem for ethnic proportions $|N_1|/n = 0.741$ (Chinese), $|N_2|/n = 0.134$ (Malay), and $|N_3|/n = 0.125$ (Indian/Others), and corresponding quotas $\alpha_{1q} = 0.87$, $\alpha_{2q} = 0.25$ and $\alpha_{3q} = 0.15$ for every block M_q.

5.1 The Lottery Mechanism

Sections 2 and 3 study an *optimal* mechanism for assigning goods to agents under diversity constraints. To the best of our knowledge, this mechanism is not used for allocating goods in practice; rather, lotteries are used to allocate items in both real-world instances that inspire this work. The mechanisms randomly order agents, and let each agent pick their favorite item in turn, *while respecting predetermined quotas*. In this section, we formulate a simple one-shot lottery-based mechanism that captures the aspect of the problems described in Sections 1.3 and 1.4 that we are most interested in: the impact of type-block constraints (as defined by inequalities (2)) on assignment by lottery. Algorithm 1 is not the actual mechanism used in Singapore public housing or Chicago school choice (see the discussion in the respective sections). It is easy to see that the algorithm takes poly(mn) time to run.

Algorithm 1: Lottery Mechanism for Assignment with Type-Block Constraints

---

Input: Agents N grouped into types $N_1, N_2, \ldots, N_k$; items M split into blocks $M_1, M_2, \ldots, M_l$; type-block capacities $\lambda_{pq}\ \forall(p, q) \in [k] \times [l]$; utility matrix $(u(i, j))_{n \times m}$.

Initialize: Allocation matrix $X = (x_{ij})_{n \times m} \leftarrow (0)_{n \times m}$; remaining agents $N_{\text{rem}} \leftarrow N$; unassigned items $M_{\text{rem}} \leftarrow M$.

for $t \in \{1, 2, \ldots, n\}$ **do**

> Draw agent uniformly at random: $i_t \sim \mathbb{U}(N_{\text{rem}})$.
>
> Find type of i_t: $p_t \leftarrow p \in [k]$ s.t. $i_t \in N_p$.
>
> Find blocks that have not hit capacity for type p_t: $Q_t \leftarrow \{q \in [l] : \sum_{i \in N_{p_t}} \sum_{j \in M_q} x_{ij} < \lambda_{p_t q}\}$.
>
> Find available items: $M_t \leftarrow M_{\text{rem}} \cap \left(\cup_{q \in Q_t} M_q\right)$.
>
> **if** $M_t \neq \emptyset$ **then**
>
> > Assign to i_t available item for which she has highest utility, breaking ties lexicographically:
> >
> > $j_t \leftarrow \arg\max_{j \in M_t} u(i_t, j)$.
> >
> > $x_{i_t j_t} \leftarrow 1$.
> >
> > $M_{\text{rem}} \leftarrow M_{\text{rem}} \backslash \{j_t\}$.
>
> **end**
>
> $N_{\text{rem}} \leftarrow N_{\text{rem}} \backslash \{i_t\}$.

end

return X.

We know that this lottery mechanism with quotas cannot produce better welfare than the optimal constrained mechanism (AssignTC); one of the objectives of our experiments is to find out how much worse the lottery performs for various utility models. We define the *price of* (one instance of) *the diverse lottery* as follows.

Definition 5.1. Let a denote an arbitary run of Algorithm 1 and X_a the unique matching of items to agents induced by the run a. Then, the *price of the diverse lottery instance a* under utilities $u = (u(i, j))_{n \times m}$ is given by: ♦

$$PoDL(u, a) \triangleq \frac{OPT(u)}{u(X_a)}.$$

We will estimate and report the expected value of $PoDL(u, a)$ (the expectation being over all possible permutations of agents induced by the uniform random sampling without replacement in Algorithm 1) alongside the realized value of $PoD(u)$ (the corresponding performance measure for the optimal constrained mechanism) for the same set of parameter values (agent-item utilities and type-block capacities): see Sections 5.2 and 5.3. Note that the realized $PoD(u)$ is a *lower bound* on $PoDL(u, a)$ for any run a of Algorithm 1 for the same problem instance; but the upper bounds from Theorems 4.2 and 4.6 do not apply to $PoDL(u, a)$ at all.

5.2 The Singapore Public Housing Allocation Problem

Data Collection. In order to create realistic instances of the AssignTC problem within the Singaporean context, we collected data on the location and number of flats of recent HDB housing development projects advertised over the second and third quarters of 2017.[16] Each of these developments corresponds to a block in our setup, for a total of $m = 1350$ flats partitioned into $l = 9$ blocks (a detailed map is given in Figure 5). Moreover, each flat in any of these blocks belongs to one of several pre-specified categories, viz. 2-room flexi, 3-room, 4-room, and 5-room; our data set includes lower and upper bounds, $LB(t, q)$ and $UB(t, q)$ respectively, on the monthly cost (loan) for a flat of category t in block M_q for every t and q. We consider two applicant pools whose ethnic composition follows the 2010 Singapore census report [Department of Statistics, Singapore 2010]: there are $n = m = 1350$ applicants in the first pool with $|N_1| = 1000$ ($\approx 74.1\%$ Chinese), $|N_2| = 180$ ($\approx 13.4\%$ Malay), and $|N_3| = 170$ ($\approx 12.5\%$ Indian/Others); the second pool has $n = 3000$ applicants with $|N_1| = 2223$, $|N_2| = 402$, and $|N_3| = 375$. From the 2010 Singapore census report, we also collected the average salary $S(p)$ of each ethnicity group $p \in [k]$, given in Singapore dollars: $S(1) = 7,326$, $S(2) = 4,575$ and $S(3) = 7,664$.[17] From publicly available data[18] on Singapore's Master Plan 2014,[19] we collected the locations of the geographic centers of the 55 planning areas that Singapore is divided into; we also obtained the population sizes of the three ethnicity groups under consideration in each planning area from the General Household Survey 2015 data available from the Department of Statistics, Singapore.[20] Finally, we use a uniform block capacity using the latest HDB block quotas [Deng et al. 2013]: for every block M_q, we have $\alpha_{1q} = 0.87$, $\alpha_{2q} = 0.25$ and $\alpha_{3q} = 0.15$.

Utility Models. All parameters used to generate AssignTC instances in our simulations are based on real data, except for agent utilities over apartments. Conducting large-scale surveys that elicit user preferences over apartments is beyond the scope of this work; thus, we base our agent utility models on simulated utilities. We examine four utility models, each characterized by a parameter whose value does not come from the data: *distance-based* ($Dist(\sigma^2)$), *type-based* ($Type(\sigma^2)$), *project approval-based* ($Project(\rho)$), and *price-based* ($Price(\sigma^2)$).

---

[16]http://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/bto-sbf

[17]We found no public data on applicant pools for public housing allocation in Singapore. We wanted to test the performance of the constrained optimization approach (mainly) for representative values of n in the interesting domain $n \geq m$. We chose $n = m$ (for which it is possible to achieve a perfect matching) and $n = \lceil 2m \times 10^{-3} \rceil \times 10^3$; we observed a surprisingly small difference in the realized $PoD(u)$ for these two values of n and did not repeat our expensive experiments for higher values of n. Moreover, in both our motivating real-world problems, it is unlikely that the number of agents is orders of magnitude higher than the number of items. Similar reasoning applies to our choices of n in Section 5.3.

[18]https://data.gov.sg/dataset/master-plan-2014-planning-area-boundary-web

[19]https://www.ura.gov.sg/Corporate/Planning/Master-Plan/

[20]https://www.singstat.gov.sg/publications/ghs/ghs2015content > Statistical Tables > Basic Demographic Characteristics

- M_1: Sky Vista (128 flats)
- M_2: West Scape (162 flats)
- M_3: Rivervale Shores (156 flats)
- M_4: Marsiling Grove (249 flats)
- M_5: Woodlands Spring (108 flats)
- M_6: Forest Spring (94 flats)
- M_7: Woodleigh Hillside (104 flats)
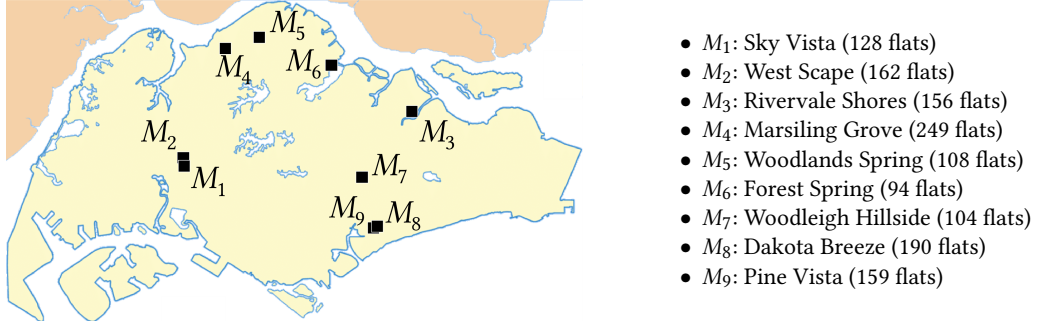- M_8: Dakota Breeze (190 flats)
- M_9: Pine Vista (159 flats)

Fig. 5. Block locations and number of flats.

- In the distance-based utility model, each agent $i \in N$ has a preferred geographic location $\vec{a}_i \in \mathbb{R}^2$ (chosen uniformly at random within the physical landmass of Singapore)[21] that she would like to live as close as possible to (say, the location of her parents' apartment, workplace, or preferred school). For every block M_q, we generate the utility that agent i derives from apartment $j \in M_q$ by first drawing a sample from the normal distribution $\mathcal{N}(1/d(\vec{a}_i, loc(M_q)), \sigma^2)$, where $loc(M_q) \in \mathbb{R}^2$ is the geographical location of block M_q and $d(\cdot, \cdot)$ represents Euclidean distance, and then renormalizing to make the sum of utilities of each agent for all apartments in M equal to one.

- In the type-based utility model, we assume that all agents of the same type (i.e. ethnic group) have the same preferred location (i.e. $\forall p \in [k], \forall i, i' \in N_p, \vec{a}_i = \vec{a}_{i'}$). The rest of the model description follows the above distance-based model.

- In the project approval-based utility model, we construct, for each type, a categorical distribution over the 55 planning areas of Singapore, the probability of each area being proportional to the fraction of the sub-population of that type living in that area; for each agent i, we sample a preferred planning area from the above distribution corresponding to i's type; if a project M_q is within a radius ρ of the geographic center of agent i's preferred planning area, then i approves of the project, i.e. $u(i, j) = 1 \; \forall j \in M_q$, else i disapproves of the project, i.e. $u(i, j) = 0 \; \forall j \in M_q$.[22]

- In the price-based utility model, each agent $i \in N_p$ has a salary s_i that is generated according to the normal distribution $\mathcal{N}(S(p), \sigma^2)$. Each flat $j \in M_q$ of category t has a monthly cost p_j that is chosen uniformly in $[LB(t, q), UB(t, q)]$. We assume that agent i is willing to pay one-third of her monthly salary on mortgage installments.[23] The utility that agent i derives from flat j is then defined by

$$u(i, j) = \frac{1/(p_j - \frac{s_i}{3})^2}{\widehat{U}_i},$$

---

[21]More precisely, we considered the largest rectangle that fits within the main island of Singapore, constructed a two-dimensional Cartesian coordinate system with one degree of latitude/longitude (each roughly equal to 111 km at the location of Singapore), and then picked x- and y-coordinates independently and uniformly within this rectangle.

[22]This is a specific instance of the *approval model* (or *dichotomous preferences* over items) where an agent either wants or does not want an item but does not distinguish among the items she wants. This also corresponds to an unweighted bipartite matching setting (between agents and items) where there is an edge between an agent and item if and only if the agent approves/wants the item. In such a situation, the utilities of an agent are usually not normalized over items (see, e.g. [Bogomolnaia and Moulin 2004]).

[23]The choice of the one-third fraction is inspired by the "3-3-5 rule" for deciding whether one can afford a flat given one's income (https://www.areyouready.sg/YourInfoHub/Pages/News-How-to-use-the-3-3-5-rule-to-consider-if-you-can-afford-your-new-home.aspx), endorsed by the Central Provident Fund Board of Singapore (https://www.cpf.gov.sg/members).

where $\widehat{U}_i \triangleq \sum_{z \in M} 1/(p_z - \frac{s_i}{3})^2$ is the normalization factor. The rationale for the utility formula is that a much higher cost relative to the budget makes the flat unaffordable while a much lower cost indicates unsatisfactory quality, making the agent unhappy in both scenarios.

Evaluation. For each of our treatments (Figures 6-8); we report the price of diversity $PoD(u)$ as per Definition 4.1 (hatched bar); the theoretical upper bound on $PoD(u)$ as per Theorem 4.6 (dark gray bar); and the price of the diverse lottery $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations (light gray bar).[24]

First, we want to compare the distance-based utility model $Dist(\sigma^2)$ and the type-based model $Type(\sigma^2)$ in order to estimate the welfare loss due to imposing ethnicity constraints. To do so, we vary both σ^2 in $\{1, 5, 10\}$ and n in $\{1350, 3000\}$; the results reported in Figures 6 are on average performance over 100 randomly generated instances. Our first observation is that, in all our experiments, the $Dist(\sigma^2)$ exhibits virtually no reduction in welfare due to the imposition of type-block constraints (see the hatched bars in the charts on the left). This is because utilities in $Dist(\sigma^2)$ are independent of ethnicities, resulting in a very low value for the inter-type disparity parameter (see the dark gray bars) — in fact, for any utility model where utilities are independent of ethnicities, the value of the disparity parameter should intuitively be close to 1 with a high probability. For utilities generated based on the $Type(\sigma^2)$ model, the disparity parameter is somewhat higher (utilities do strongly depend on ethnicities), resulting in a higher $PoD(u)$. Despite making no attempt to optimize social welfare under type-block constraints, the HDB lottery mechanism does surprisingly well when the number of agents equals the number of apartments (Figure 6 (a)), extracting at least 84% of the optimal unconstrained welfare under the $Dist(\sigma^2)$ utility model, and at least 79% of the social welfare under the $Type(\sigma^2)$ model. However, the welfare loss induced by the lottery mechanism is negatively impacted by the number of agents (Figure 6 (b)); for instance, it only extracts 65% of the optimal unconstrained welfare under $Dist(1)$ with $n = 3000$ and, in fact, the lottery-induced welfare loss for this treatment even exceeds the theoretical upper bound on the price of diversity.

Let us now turn to the project approval-based utility model $Project(\rho)$. To compute distances in km, we make use of the fact that one degree of latitude or longitude at the location of Singapore corresponds to roughly 111 km; we vary the radius ρ in $\{5, 7.5, 10\}$ (in km). The results averaged over 100 runs are provided in Figure 7. In all instances, not only is the price of diversity almost one but the lottery-induced welfare is also nearly as good, achieving at least 87% of the unconstrained optimum for 1350 agents and practically 100% for 3000 agents; the disparity parameter is also consistently close to its ideal value of 1, keeping the upper bound at around 1.45 regardless of the radius. Thus, this can be considered an example of a utility model for which the lottery mechanism virtually implements a constrained optimal allocation for a wide range of model parameters.

Finally, we study the price-based utility model $Price(\sigma^2)$, varying σ^2 in $\{0, 10, 50\}$; the results obtained by averaging over 100 runs are given in Figure 8. While the price of diversity is practically equal to one in all instances, the welfare loss observed with the lottery mechanism drastically increases with σ^2 (recall that agents from the same ethnicity group have identical preferences when $\sigma^2 = 0$): for instance, for 1350 agents, it extracts 98% of the optimal unconstrained welfare under $Price(0)$ while it only extracts 35% of this value under $Price(50)$. These numerical tests show that utility models exist for which the lottery mechanism may perform poorly compared to the optimal constrained allocation mechanism, even in allocation problems with a very low price of diversity.

---

[24]The error bars in Figures 6-8 and 10 represent one (estimated) standard error of the mean on either side; see e.g. Clymo [2019] for recommendations on how many significant digits of the mean to report based on its standard error, which we have tried to follow.
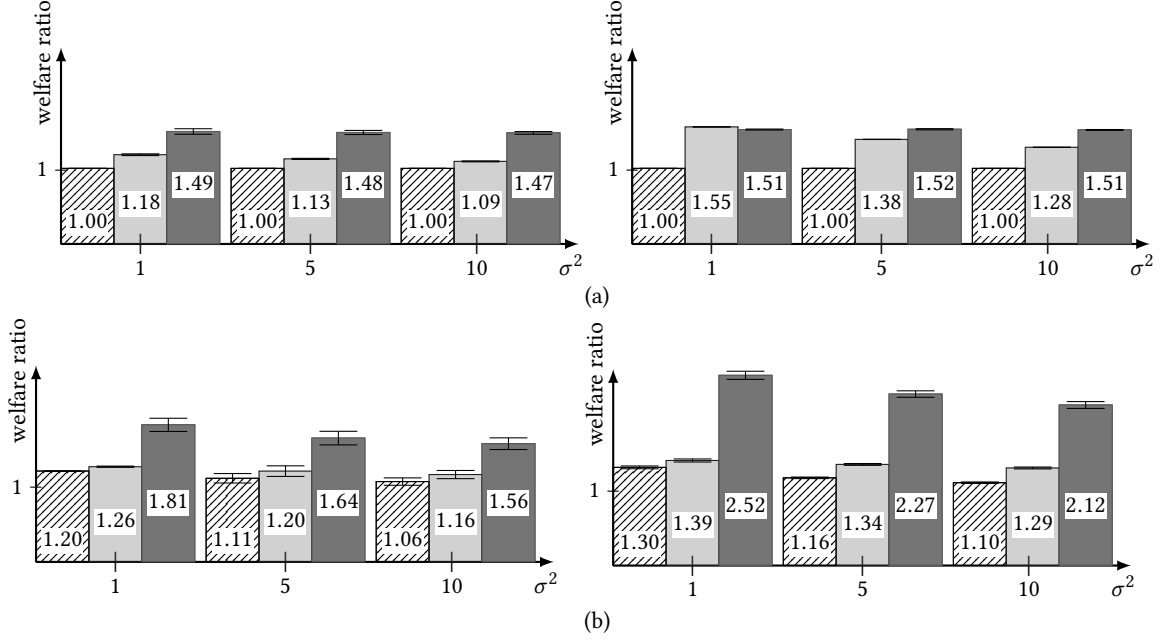
(a)

(b)

Fig. 6. Averaged welfare ratios obtained for (a) $Dist(\sigma^2)$ and (b) $Type(\sigma^2)$ with $n = m = 1350$ (left) and $n = 3000$ (right), $m = 1350$ in our simulated instances of the Singapore public housing allocation problem. Horizontal axis shows values of utility model parameter σ^2 (variance); vertical axis shows ratio (or upper bound on ratio) of optimal unconstrained welfare to welfare induced by some constrained mechanism — hatched bar: $PoD(u)$ as per Definition 4.1, light gray bar: upper bound from Theorem 4.6, dark gray bar: $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations.
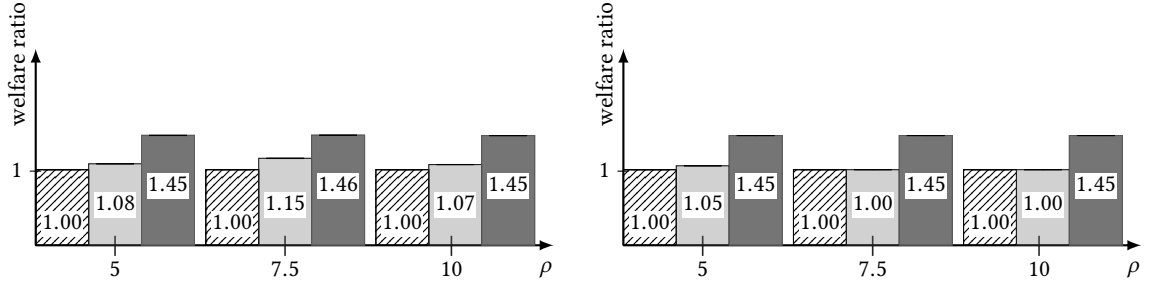

Fig. 7. Averaged welfare ratios obtained for $Project(\rho)$ with $n = m = 1350$ (left) and $n = 3000$, $m = 1350$ (right) in our simulated instances of the Singapore public housing allocation problem. Horizontal axis shows values of the utility model parameter ρ (radius); vertical axis shows ratio (or upper bound on ratio) of optimal unconstrained welfare to welfare induced by some constrained mechanism — hatched bar: $PoD(u)$ as per Definition 4.1, light gray bar: upper bound from Theorem 4.6, dark gray bar: $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations.

5.3 Chicago Public School Admissions

Data Collection. From the Chicago Public Schools website,[25] we collected data on the locations of magnet schools in Chicago; we focus on these schools because they use a lottery mechanism to select students. We also collected the

---
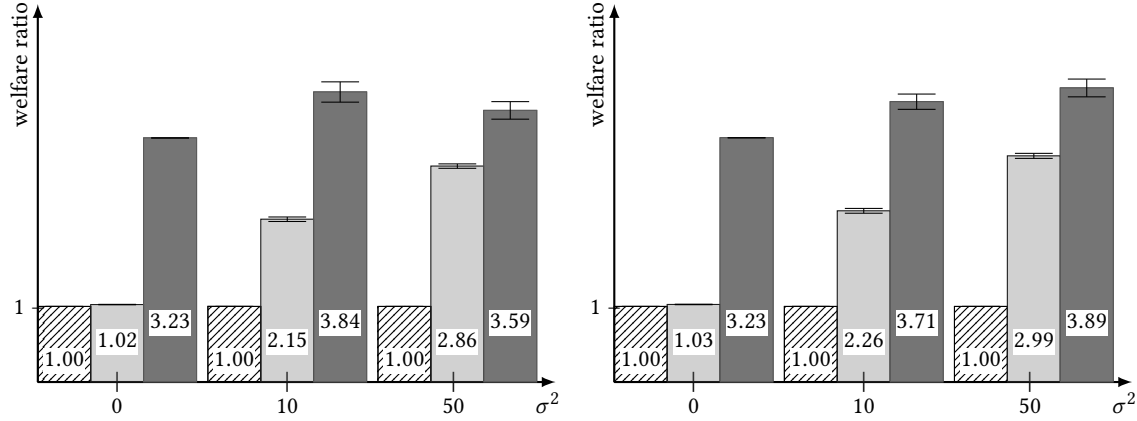
[25]http://cps.edu/Pages/home.aspx

Fig. 8. Averaged welfare ratios obtained for $Price(\sigma^2)$ with $n = m = 1350$ (left) and $n = 3000$, $m = 1350$ (right) in our simulated instances of the Singapore public housing allocation problem. Horizontal axis shows values of utility model parameter σ^2 (variance); vertical axis shows ratio (or upper bound on ratio) of optimal unconstrained welfare to welfare induced by some constrained mechanism — hatched bar: $PoD(u)$ as per Definition 4.1, light gray bar: upper bound from Theorem 4.6, dark gray bar: $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations.

total number of students enrolled in these schools in 2018, and divided this number by 9 to obtain the approximate number of students that can be accepted to the first grade (there are nine grades in total). This leads us to instances with $l = 37$ blocks (schools) and $m = 2261$ items in total (available spots). In this school admission problem, students are partitioned into $k = 4$ types, viz. Tiers 1, 2, 3 and 4, depending on their residence (see Figure 9). In our experiments, we consider a pool of $n \in \{2261, 5000\}$ students whose type composition follows the real-world proportion: we have $|N_1| = 613$ ($\approx 27.1\%$: Tier 1), $|N_2| = 622$ ($\approx 27.5\%$: Tier 2), $|N_3| = 533$ ($\approx 23.6\%$: Tier 3) and $|N_4| = 493$ ($\approx 21.8\%$: Tier 4) for $n = 2261$ and we have $|N_1| = 1355$, $|N_2| = 1375$, $|N_3| = 1180$ and $|N_4| = 1090$ for $n = 5000$. We use equal fractional quotas across schools and tiers, i.e. our type-block capacities are $\lambda_{pq} = 0.25|M_q| \ \forall (p, q) \in [k] \times [l]$. Recall also that we have one-shot lotteries (Section 5.1) in our simulations.

Utility model. Since we do not have access to students' utilities for schools, we simulate them as follows. We use the distance-based utility model $Dist(\sigma^2)$ introduced in Section 5.2 with the following important modifications:

- We choose the preferred location of a student uniformly at random from the collection of census tracts (polygons) belonging to her tier (see Figure 9); the position of every polygon is approximated by taking the average of the coordinates of its extreme points.
- We reset each student's utility to 0 for any school ranked 21st or lower in the preference ordering induced by her utilities computed as in the distance-based model of Section 5.2, and then renormalize the utilities; this is because students are allowed to apply to at most 20 schools.

Evaluation. In our experiments, we vary both σ^2 in $\{0, 10, 50\}$ and n in $\{2261, 5000\}$, and for each setting, we compute the following quantities: the price of diversity $PoD(u)$ as per Definition 4.1 (hatched bar); the theoretical upper bound on $PoD(u)$ as per Theorem 4.6 (dark gray bar); and the price of the diverse lottery $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations (light gray bar). Results obtained by averaging over 100 runs are given in Figure 10. First, we observe that both the price of diversity and the loss of the lottery mechanism decrease as σ^2 increases; this
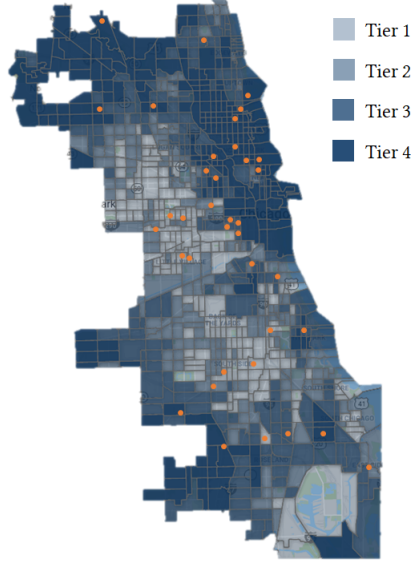
Fig. 9. Map of Chicago showing the tier statuses of census tracts (http://cpstiers.opencityapps.org/) and locations of magnet schools (orange dots) based on http://cps.edu/ScriptLibrary/Map-SchoolLocator/index.html.

is unsurprising since utilities are less type-dependent as σ^2 grows. However, we observe that the loss of the lottery mechanism is quite high in all instances. Moreover, just as in the Singapore public housing allocation problem, the outcome of the lottery mechanism is negatively impacted by the number of students; therefore, we can conclude that this mechanism seems to be better suited to problems with an equal number of agents and items.

6 DISCUSSION AND FUTURE WORK

Our work constitutes a first step towards a better understanding of the effect of hard diversity constraints on social welfare. We can summarize our results as follows. One of our main contributions is to offer computational insights into the ASSIGNTC problem: a general hardness result, sufficient conditions for tractability, and a $\frac{1}{2}$-approximation algorithm. We derive two upper bounds on the price of diversity defined as the ratio of the optimal welfare achievable with and without type-block constraints: the first is in terms of block quotas only, independent of the utility model, hence under the planner's control; the second is parametrized in terms of inter-type disparity, which shows that when the disparity is low, the welfare loss is much closer to its ideal value of 1 than the first bound would suggest. We analyze our model's behavior in simulations based on two real-world data sets.

We will conclude with further remarks on the above results, a few straightforward implications, and directions for further research.

6.1 Further questions on complexity and approximation

Some questions that remain open are:

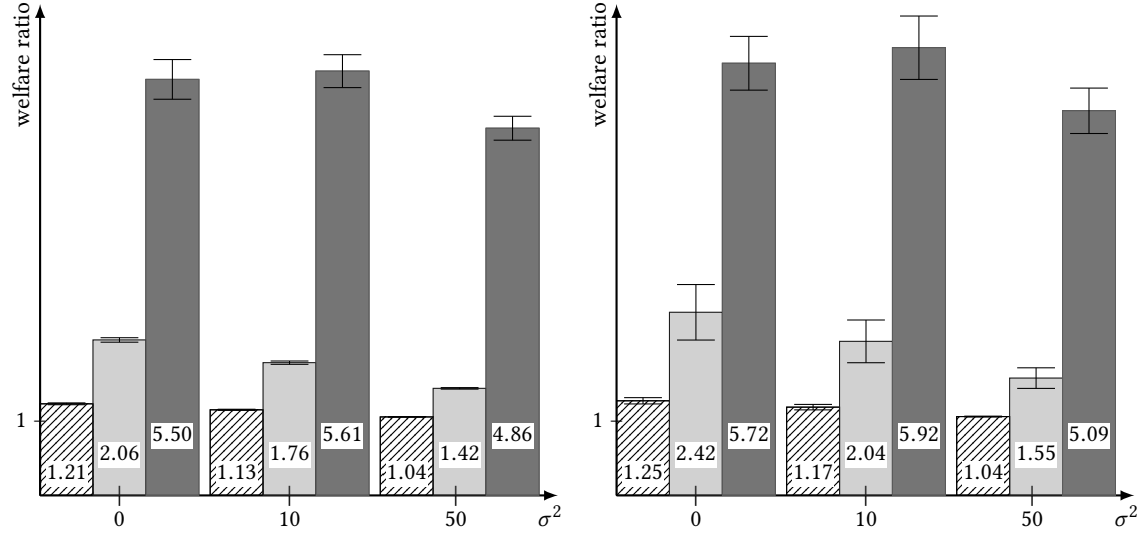- Is a better approximation to ASSIGNTC possible?

Fig. 10. Averaged welfare ratios obtained for $Dist(\sigma^2)$ with $n = m = 2261$ (left) and $n = 5000$, $m = 2261$ (right) in our simulated instances of the Chicago public school admission problem. Horizontal axis shows values of utility model parameter σ^2 (variance); vertical axis shows ratio (or upper bound on ratio) of optimal unconstrained welfare to welfare induced by some constrained mechanism — hatched bar: $PoD(u)$ as per Definition 4.1, light gray bar: upper bound from Theorem 4.6, dark gray bar: $PoDL(u, \cdot)$ as per Definition 5.1, averaged over 100 agent permutations.

- Does there exist a more natural/elegant polynomial-time approximation algorithm for AssignTC that does not involve reduction to BCMatching?
- Are there any other non-trivial special cases of AssignTC that admit polynomial-time solutions?

However, since we have shown that AssignTC is a special (still NP-complete) case of the BCMatching problem, some easy generalizations of our results follow from the literature on BCMatching. For example, there is a polynomial-time approximation scheme (PTAS) for BCMatching if one allows $(1 + \varepsilon)$-violations of the color constraints [Grandoni et al. 2009]; this immediately implies a PTAS for AssignTC where one allows $(1 + \varepsilon)$-violations of the type-block constraints. Intuitively, this means that we can efficiently achieve near-optimal constrained optimal welfare if we are willing to *soften* our hard constraints within well-defined limits.

6.2 *Soft* diversity concepts

There are alternative approaches towards diversity promotion in various branches of research that do not rely on *hard bounds* (fixed quotas):

- maximization of a carefully constructed monotone, submodular function that itself trades diversity off against "solution quality" (similar in spirit to the welfare concept we have used) in a matching/allocation/subset selection setting [Ahmed et al. 2017; Dickerson et al. 2019; Schumann et al. 2019];
- soft diversity constraints in a matching setting (e.g. school seat allocation) with ordinal preferences [Ehlers et al. 2014; Kurata et al. 2017, and references therein]: quotas are used as "flexible limits that regulate ... priorities" when assigning scarce items [Ehlers et al. 2014].

- sampling datasets to ensure an adequate representation of labeled categories (*combinatorial diversity* or *diversity indices*) and/or a broad coverage of the feature space (*geometric diversity* via *determinantal point processes*) for machine learning applications [Celis et al. 2016, and references therein].

An important future research direction would be to compare and contrast diversity quotas with these different approaches towards diverse solutions in terms of their impact on welfare, both theoretically and experimentally.

6.3 Types based on "overlapping" attributes

The types considered in our paper essentially constitute a partition over the agents with respect to one *attribute*: ethnicity in Singapore public housing and SES-based tiers in Chicago Public Schools. But in many situations, quotas/bounds are specified for multiple attributes. Consider for example, a hypothetical public housing allocation problem in a city that has the same three ethnic groups (and the respective quotas) as in Singapore; but the population is also divided into high-income and low-income groups and the city additionally requires that no block of flats should have more than 55% of its occupants from either of these two groups. Flats, in addition to be being grouped into blocks, could also be partitioned based on categories (see Section 5.2) and any number of capacity constraints could be specified combinatorially.

It is easy to see that AssignTC is a special case of this class of problems. So, the NP-hardness of AssignTC (Theorem 3.4) readily implies that of this more general class. However, it is not obvious whether this general version reduces to BCMatching, hence approximability results of AssignTC may not generalize. It is also unclear how to extend the *PoD* analysis of Section 4 – if it can be done at all – to the situation with multiple partitions each with its own set of arbitrary percentage caps. This is a challenge we wish to tackle in the future. For existing work on diversity requirements based on multiple overlapping attributes in a subset selection setting (not matching), please refer to Bredereck et al. [2018]; Lang and Skowron [2016].

6.4 Utility models and experiments

Simulating agent utilities is still a major challenge: ideally, one would elicit applicants' utilities directly via large-scale national surveys. The aim of our experiments was to see how much worse the diversity-constrained optimal allocation performs than our theoretical bounds (and, to some extent, how much better it is compared to its lottery version) depending on the utility model, and not to compare utility models based on the *PoD* as a performance measure. Hence, we formulated clean utility models that focus on individual factors that, we believe, bear on individuals' preferences in the problem domains considered. For example, our simulations with the $Dist(\sigma^2)$ and $Type(\sigma^2)$ models in Section 5.2 tested two 'extreme' cases: one where there is no correlation between ethnicity and utility, and one where utility is artificially correlated to ethnicity. The truth is likely somewhere in between. Ethnic groups in Singapore most likely do have some correlation between their utility models; this can be due to socioeconomic factors (there is some correlation between ethnicity and socioeconomic status), the location of cultural or religious centers, or other unknown factors. Developing and investigating refined utility models is an interesting direction for future work.

6.5 Lottery mechanisms with diversity quotas

Allocation based on constrained optimization under known utilities can serve as a benchmark for any mechanism respecting the same set of diversity constraints, including the lottery mechanism in Algorithm 1. Obviously, the relative loss of the lottery mechanism cannot be smaller than the *PoD* for the same problem instance; but, in our experiments,

the diverse lottery performs surprisingly well for several utility models (e.g. Figures 6 and 7), although it appears to be more sensitive to the utility model and its parameters than the constrained optimization benchmark (e.g. Figures 8 and 10). As yet, we do not have satisfactory explanations for many of our observations. Offering theoretical guarantees on the performance of the lottery mechanism with diversity constraints (and in some sense, complementing the analysis by Immorlica et al. [2017]) would provide better insights on our experimental results. In particular, we could look at the ratio of the expected utilitarian social welfare induced by a lottery without diversity constraints to that under diversity constraints for the same utilities.[26]

Moreover, recall that the lottery mechanism operates on the basis of partial information on the cardinal utilities to which the benchmark has access: specifically, for each agent in the random sequence, the algorithm only elicits the agent's top choice among the items in blocks for which the corresponding capacity constraint has not been satisfied, which also depends on the predecessors of the agent in the random sequence. This is even more restricted information than some recent work on matching/assignment [Anshelevich and Sekar 2016] and subset selection [Anshelevich et al. 2018, and references therein] has considered: these approaches assume access to ordinal information (each agent's ranking over alternatives) but the performance is measured in terms of the hidden cardinal utilities that generate these rankings – the concept of *distortion* [Procaccia and Rosenschein 2006] is often used in this line of work as a measure of fractional loss in (cardinal) solution quality due to ordinal information. It will be interesting to study (optimal) algorithms working with ranking data under diversity constraints, and the resulting combination of distortion and price of diversity – intuitively, the welfare loss with full rankings cannot be worse than that of our lottery (with significantly less information), which already performs close to the constrained optimum (with the true cardinal inputs) in many of our experiments.[27]

A relevant practical question that also needs attention is why the constrained optimization approach is not used in our motivating real-world problems instead of allocations based on lotteries. There could be several problem-specific reasons for this. We offer AssignTC as a benchmark for the *utilitarian* objective under one particular diversity-promoting policy (hard capacity bounds); but a central planner/allocator might have several objectives in mind at the same time. In a mechanism that relies on agents' subjective valuations of items (that need to be elicited in some form), one might also care about *individual fairness* and *discouraging strategic behavior*. In the absence of diversity constraints, a (uniform) lottery is identical to *random serial dictatorship* which satisfies many such desiderata [Abdulkadiroğlu and Sönmez 1998; Bogomolnaia and Moulin 2001]. As such, an extensive investigation of diversity constraints with properties other than the utilitarian social welfare is also imperative.

6.6 Closing remarks

Our results describe an inevitable tradeoff between diversity and social welfare; however, we emphasize that this does not constitute a moral judgment on the authors' part. Economic considerations are certainly important, but they are by no means an exclusive or even a first-order consideration. That said, understanding the impact of diversity constraints on social welfare is key if one is to justify their implementation.

---

[26]This ratio can never exceed the expected price of the diverse lottery as defined in Section 5.1 (see Definition 5.1 and the following paragraph), but is not guaranteed to be larger than the *PoD* for the same problem instance. A detailed analysis, both theoretical and experimental, of this concept is an important research topic in its own right, hence we have not reported the ratio of the expected welfare of the lottery mechanism without constraints to that with type-block constraints for our experiments in Section 5.

[27]We thank an anonymous reviewer for pointing out this possible connection to the literature on distortion.

ACKNOWLEDGMENTS

REFERENCES

Atila Abdulkadiroğlu, Parag A. Pathak, and Alvin E. Roth. 2009. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review* 99, 5 (2009), 1954–1978.

Atila Abdulkadiroğlu and Tayfun Sönmez. 1998. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica* 66, 3 (1998), 689–701.

Atila Abdulkadiroğlu and Tayfun Sönmez. 2003. School choice: A mechanism design approach. *American Economic Review* 93, 3 (2003), 729–747.

Faez Ahmed, John P Dickerson, and Mark Fuge. 2017. Diverse Weighted Bipartite b-Matching. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*. 35–41.

Kook Jin Ahn and Sudipto Guha. 2014. Near linear time schemes for uncapacitated and capacitated b-matching problems in nonbipartite graphs. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 239–258.

Elliot Anshelevich, Onkar Bhardwaj, Edith Elkind, John Postl, and Piotr Skowron. 2018. Approximating optimal social choice under metric preferences. *Artificial Intelligence* 264 (2018), 27–51.

Elliot Anshelevich, Sanmay Das, and Yonatan Naamad. 2013. Anarchy, Stability, and Utopia: Creating Better Matchings. *Autonomous Agents and Multi-Agent Systems* 26, 1 (2013), 120–140.

Elliot Anshelevich and Shreyas Sekar. 2016. Blind, greedy, and random: Algorithms for matching and clustering using only ordinal information. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. 383–389.

Sophie Bade. 2016. *Random serial dictatorship: the one and only.* Working Paper. Royal Holloway College, University of London.

Siddharth Barman, Sanath Kumar Krishnamurthy, and Rohit Vaish. 2017. Finding Fair and Efficient Allocations. *arXiv preprint arXiv:1707.04731* (2017).

Siddharth Barman and Sanath Kumar Krishna Murthy. 2017. Approximation Algorithms for Maximin Fair Division. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*. 647–664.

Ulrich Bauer. 2004. *Assignment Problem with Constraints.* Master's thesis. ETH Zürich Department of Computer Science. Supervised by Andreas Weißl.

Anand Bhalgat, Deeparnab Chakrabarty, and Sanjeev Khanna. 2011. Social welfare in one-sided matching markets without money. In *Proceedings of the 14th International Workshop and 15th International Conference on Approximation, Randomization, and Combinatorial Optimization: Algorithms and Techniques (APPROX/RANDOM)*. 87–98.

Anna Bogomolnaia and Hervé Moulin. 2001. A new solution to the random assignment problem. *Journal of Economic Theory* 100, 2 (2001), 295–328.

Anna Bogomolnaia and Hervé Moulin. 2004. Random matching under dichotomous preferences. *Econometrica* 72, 1 (2004), 257–279.

Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, and Piotr Skowron. 2018. Multiwinner Elections with Diversity Constraints. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. 933–940.

Eric Budish and Estelle Cantillon. 2012. The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *American Economic Review* 102, 5 (2012), 2237–2271.

Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. 2016. The unreasonable fairness of maximum Nash welfare. In *Proceedings of the 17th ACM Conference on Economics and Computation (EC)*. 305–322.

L. Elisa Celis, Amit Deshpande, Tarun Kathuria, and Nisheeth K. Vishnoi. 2016. How to be fair and diverse?. In *3rd Workshop on Fairness, Accountability, and Transparency in Machine Learning (FATML), arXiv preprint arXiv:1610.07183*.

Chicago Public Schools. 2017. Chicago Public Schools Policy Manual: Admissions Policy for Magnet, Selective Enrollment and Other Options For Knowledge Schools and Programs. (26 April 2017). Section 202.2. Board Report 17-0426-PO2. Available at http://policy.cps.edu/Policies.aspx.

Beng-Huat Chua. 1991. Race relations and public housing policy in Singapore. *Journal of Architectural and Planning Research* (1991), 343–354.

RS Clymo. 2019. How many of the digits in a mean of 12.3456789012 are worth reporting? *BMC research notes* 12, 1 (2019), 148.

Pierluigi Crescenzi. 1997. A short guide to approximation preserving reductions. In *Proceedings of the 12th IEEE Conference on Computational Complexity (CCC)*. 262–273.

Pierluigi Crescenzi, C. Fiorini, and Riccardo Silvestri. 1991. A note on the approximation of the MAX CLIQUE problem. *Inform. Process. Lett.* 40, 1 (1991), 1–5.

Yongheng Deng, Tien Foo Sing, and Chaoqun Ren. 2013. The story of Singapore's public housing: From a nation of home-seekers to a nation of homeowners. In *The Future of Public Housing: Ongoing Trends in the East and the West*, Jie Chen, Mark Stephens, and Yanyun Man (Eds.). Springer,

Chapter 7, 103–121.

Department of Statistics, Singapore. 2010. Singapore 2010 Census: Key Indicators of the Resident Population. (2010).

Department of Statistics, Singapore. 2017. Singapore in Figures. (2017).

John P Dickerson, Karthik Abinav Sankararaman, Aravind Srinivasan, and Pan Xu. 2019. Balancing Relevance and Diversity in Online Bipartite Matching via Submodularity. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 1877–1884.

Lars Ehlers, Isa E Hafalir, M Bumin Yenmez, and Muhammed A Yildirim. 2014. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory* 153 (2014), 648–683.

Clayton R. Featherstone. 2015. *Rank efficiency: Investigating a widespread ordinal welfare criterion.* Working Paper. The Wharton School, University of Pennsylvania.

Daniel Fragiadakis and Peter Troyan. 2017. Improving matching under hard distributional constraints. *Theoretical Economics* 12, 2 (2017), 863–908.

Michael R. Garey and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* W. H. Freeman & Co.

Fabrizio Grandoni, R. Ravi, and Mohit Singh. 2009. Iterative rounding for multi-objective optimization problems. In *Proceedings of the 17th Annual European Symposium on Algorithms (ESA)*. 95–106.

Housing and Development Board, Singapore. 2010. Policy changes to support an inclusive and cohesive home [Press release]. (05 March 2010). http://www.nas.gov.sg/archivesonline/speeches/record-details/809e76bf-115d-11e3-83d5-0050568939ad.

Housing and Development Board, Singapore. 2015. Costs and Fees. (07 October 2015). http://www.hdb.gov.sg/cs/infoweb/residential/buying-a-flat/new/finance/costs-and-fees.

Housing and Development Board, Singapore. 2017. Annual Report 2016/2017: Key Statistics. (2017). http://www10.hdb.gov.sg/ebook/AR2017/key-statistics.html.

Aanund Hylland and Richard Zeckhauser. 1979. The efficient allocation of individuals to positions. *Journal of Political Economy* 87, 2 (1979), 293–314.

Nicole Immorlica, Brendan Lucier, Joshua Mollner, and E. Glen Weyl. 2017. Raffles. (13 February 2017). Available at SSRN: https://ssrn.com/abstract=2916337 or http://dx.doi.org/10.2139/ssrn.2916337.

Richard D. Kahlenberg. 2016. School Integration in Practice: Lessons from Nine Districts. (14 October 2016). The Century Foundation. https://tcf.org/content/report/school-integration-practice-lessons-nine-districts.

Yuichiro Kamada and Fuhito Kojima. 2015. Efficient matching under distributional constraints: Theory and applications. *American Economic Review* 105, 1 (2015), 67–99.

Harold W. Kuhn. 1955. The Hungarian method for the assignment problem. *Naval Research Logistics (NRL)* 2, 1-2 (1955), 83–97.

Ryoji Kurata, Naoto Hamada, Atsushi Iwasaki, and Makoto Yokoo. 2017. Controlled school choice with soft bounds and overlapping types. *Journal of Artificial Intelligence Research* 58 (2017), 153–184.

David Kurokawa, Ariel D. Procaccia, and Junxing Wang. 2016. When can the maximin share guarantee be guaranteed?. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, Vol. 16. 523–529.

Jérôme Lang and Piotr Krzysztof Skowron. 2016. Multi-Attribute Proportional Representation.. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*. 530–536.

Jing Wu Lian, Nicholas Mattei, Renee Noble, and Toby Walsh. 2018. The Conference Paper Assignment Problem: Using Order Weighted Averages to Assign Indivisible Goods. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. To appear.

László Lovász and Michael D. Plummer. 2009. *Matching theory.* Vol. 367. American Mathematical Society.

James Munkres. 1957. Algorithms for the assignment and transportation problems. *J. Soc. Indust. Appl. Math.* 5, 1 (1957), 32–38.

Pekka Orponen and Heikki Mannila. 1987. *On approximation preserving reductions: Complete problems and robust measures.* Technical Report C-1987-28. Department of Computer Science, University of Helsinki. 22 pages.

Parliament of Singapore. Parliament Debates: Official Report. 1989. Better racial mix in HDB housing estates. (16 February 1989). Vol. 52, cols. 650-668.

Parag A. Pathak and Tayfun Sönmez. 2013. School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation. *American Economic Review* 103, 1 (2013), 80–106.

Sock Yong Phang and Kyunghwan Kim. 2013. Singapore's Housing Policies: 1960-2013. *Frontiers in Development Policy: Innovative Development Case Studies* (2013), 123–153.

Ariel D Procaccia and Jeffrey S Rosenschein. 2006. The distortion of cardinal preferences in voting. In *International Workshop on Cooperative Information Agents*. Springer, 317–331.

Ariel D. Procaccia and Junxing Wang. 2014. Fair enough: Guaranteeing approximate maximin shares. In *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*. 675–692.

Kimberly Quick. 2016. Chicago Public Schools: Ensuring Diversity in Selective Enrollment and Magnet Schools. (14 October 2016). The Century Foundation. https://tcf.org/content/report/chicago-public-schools.

Alvin E. Roth. 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy* 92, 6 (1984), 991–1016.

Thomas C Schelling. 1971. Dynamic models of segregation. *Journal of mathematical sociology* 1, 2 (1971), 143–186.

Candice Schumann, Samsara N Counts, Jeffrey S Foster, and John P Dickerson. 2019. The diverse cohort selection problem. In *Proceedings of the 18th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, 601–609.

Loo Lee Sim, Shi Ming Yu, and Sun Sheng Han. 2003. Public housing and ethnic integration in Singapore. *Habitat International* 27, 2 (2003), 293–307.

Georgios Stamoulis. 2014. Approximation algorithms for bounded color matchings via convex decompositions. In *Proceedings of the 39th International Symposium on Mathematical Foundations of Computer Science (MFCS)*. 625–636.

U.S. Department of Education, Office of Elementary and Secondary Education. 2017. Improving Outcomes for All Students: Strategies and Considerations to Increase Student Diversity. (19 January 2017). Washington, D.C. https://www.ed.gov/diversity-opportunity.

Maisy Wong. 2014. Estimating the distortionary effects of ethnic quotas in Singapore using housing transactions. *Journal of Public Economics* 115 (2014), 131–145.

Lin Zhou. 1990. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory* 52, 1 (1990), 123–135.