THEORY OF COMPUTING, Volume 18 (6), 2022, pp. 1–33
www.theoryofcomputing.org

SPECIAL ISSUE: APPROX-RANDOM 2019

Max-Min Greedy Matching

Alon Eden* Uriel Feige† Michal Feldman‡

Received March 2, 2020; Revised February 17, 2021; Published April 13, 2022

Abstract. A bipartite graph $G(U,V;E)$ that admits a perfect matching is given. One player imposes a linear order π over V, the other player imposes a linear order σ over U. In the greedy matching algorithm, vertices of U arrive in order σ and each vertex is matched to the highest (under π) yet unmatched neighbor in V (or is left unmatched, if all its neighbors are already matched). The matching obtained is maximal, thus matches at least half of the vertices. The max-min greedy matching problem asks: Suppose the first (max) player reveals π, and the second (min) player responds with the worst possible σ for π. Does there exist a linear order π ensuring to match strictly more than half of the vertices? Can such a linear order be computed in polynomial time?

The main result of this paper is an affirmative answer for these questions: we show that there exists a polynomial-time algorithm to compute π for which for every σ at least $\rho > 0.51$ fraction of the vertices of V are matched. We provide additional lower and upper bounds for

A conference version of this paper appeared in the Proceedings of the 22nd International Conference on Approximation Algorithms for Combinatorial Optimization Problems, 2019 (APPROX'19) [6].

*Partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement number 337122, the Israel Science Foundation (grant number 317/17), and the Blavatnik family foundation.

†Partially supported by the Israel Science Foundation (grant No. 1388/16).

‡Partially supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement number 337122, the Israel Science Foundation (grant number 317/17), and the Blavatnik family foundation.

ACM Classification: G.2.2

AMS Classification: 68W25,68R10,91B26,05C70

Key words and phrases: Online matching, pricing mechanism, markets

 DOI: 10.4086/toc.2022.v018a006

special families of graphs, including regular and Hamiltonian graphs. Our solution solves an open problem regarding the welfare guarantees attainable by pricing in sequential markets with binary unit-demand valuations.

1 Introduction

Given a bipartite graph $G(U,V;E)$, where U and V are the sets of vertices and $E \subset U \times V$ is the set of edges, a matching $M \subset E$ is a set of edges such that every vertex is incident with at most one edge of M. For simplicity of notation, for every n we shall only consider the following class of bipartite graphs, that we shall refer to as G_n. For every $G(U,V;E) \in G_n$ it holds that $|U| = |V| = n$ and that E contains a matching of size n (and hence G has a perfect matching). All results that we will state for G_n hold without change for all bipartite graphs that have a matching of size n (and arbitrarily many vertices).

Karp, Vazirani and Vazirani [13] introduced the *online bipartite matching* problem. In this problem, a bipartite graph $G(U,V;E)$ is revealed in an online manner, where side V is known in advance, and vertices in U are revealed one after another, along with their edges to V. When a vertex $u \in U$ is revealed together with its set of neighbors $N(u) \subset V$, the algorithm must decide, immediately and irrevocably, which of the yet unmatched vertices in $N(u)$ to match to u (if any). The goal of the algorithm is to maximize the size of the matching achieved.

This setting can be viewed as a game between two players: a maximizing player who wishes the resulting matching to be as large as possible, and a minimizing player who wishes the matching to be as small as possible. First, the minimizing player chooses $G(U,V;E)$ in private (without the maximizing player seeing E), subject to $G \in G_n$. Thereafter, the structure of G is revealed to the maximizing player in n steps, where at step j (for $1 \leq j \leq n$) the set $N(u_j) \subset V$ of vertices adjacent to u_j is revealed. At every step j, upon seeing $N(u_j)$ (and based on all edges previously seen and all previous matching decisions made), the maximizing player needs to irrevocably either match u_j to a currently unmatched vertex in $N(u_j)$, or leave u_j unmatched.

There has been much recent interest in the online bipartite matching problem and variations and generalizations of it, as such models have applications to allocation problems in certain economic settings, in which buyers (vertices of U) arrive online and are interested in purchasing various items (vertices of V). A prominent example of such an application is online advertising; for more details, see for example the survey by Mehta [17]. The new problems are both theoretically elegant and practically relevant.

Max-min greedy matching We study a setting related to online bipartite matching, that we call *max-min greedy matching*. Our setting is also a game between a maximizing player and a minimizing player. The bipartite graph $G(U,V;E) \in G_n$ is given upfront. Upon seeing G the maximizing player chooses a linear order π over V. Upon seeing G and π, the minimizing player chooses a linear order σ over U. The combination of G, π and σ defines a unique matching $M_G[\sigma,\pi]$ that we refer to as the greedy matching. It is the matching produced by the greedy matching algorithm in which vertices of U arrive in order σ and each vertex $u \in U$ is matched to the highest (under π) yet unmatched $v \in N(u)$ (or left unmatched, if all of $N(u)$ has already been matched).

The matching $M_G[\sigma,\pi]$ has several additional equivalent definitions. For example, $M_G[\sigma,\pi]$ is the matching produced by the greedy matching algorithm in which vertices of V arrive in order π and each vertex $v \in V$ is matched to the highest (earliest in arrival order under σ) yet unmatched $u \in N(v)$ (or left unmatched, if all of $N(v)$ has already been matched). Also, $M_G[\sigma, \pi]$ is the unique stable matching in G (in the sense of [10]), if the preference order of every vertex $u \in U$ over its neighbors is consistent with π, and the preference order of every vertex $v \in V$ over its neighbors is consistent with σ. More details are given in Section 8.

Let

$$\rho[G] = \frac{1}{n} \max_{\pi} \min_{\sigma} [|M_G[\sigma, \pi]|],$$

and let

$$\rho = \min_{G \in G_n} [\rho[G]].$$

It is easy to see that $\rho \geq 1/2$. In fact, to ensure a matching of size $n/2$, the max player need not work hard. Since every greedy matching is a *maximal* matching, for *every* linear order π the obtained matching is of size at least $n/2$. The question we study here is whether the max player can ensure a matching of size strictly greater than $n/2$; that is, whether ρ is bounded away from $1/2$.

For an upper bound on ρ, it was observed by Cohen-Addad et al. [4] that $\rho \leq 2/3$. To show this, they observe that in the 6-cycle graph, depicted in Figure 1, no linear order π can guarantee to match more than two vertices in the worst case. Indeed, suppose (without loss of generality) that $\pi = (v_3, v_2, v_1)$. For $\sigma = (u_1, u_3, u_2)$, u_1 is matched to v_2, u_3 is matched to v_3, and u_2 is left unmatched, resulting in a matching of size 2.

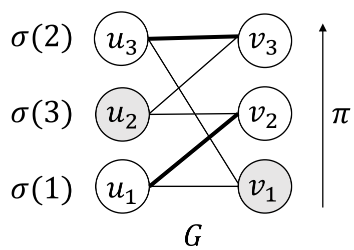

Figure 1: For every linear order π there exists a linear order σ that matches only 2 of the 3 vertices. Thick edges are in the matching; gray vertices are unmatched.

1.1 Our results

Our main result resolves the open problem in the affirmative.

Main result [Theorem 2.1]: *For all graphs, $\rho \geq 1/2 + 1/86 > 0.51$. Moreover, there is a polynomial-time algorithm that given $G(U,V;E)$ produces a linear order π over V satisfying the above bound.*

The significance of this result is that $1/2$ is not the optimal answer. We believe that further improvements are possible. In fact, for Hamiltonian graphs we show that $\rho \geq 5/9$ (see Section 6).

The proof method is quite involved; it is natural to ask whether simpler approaches may work. In what follows we specify three natural attempts that all fail.

Failed attempt 1: random linear order. A first attempt would be to check whether a random linear order π obtains the desired result (in expectation)[1]. The performance of a random linear order is interesting for an additional reason: it is the performance in scenarios where the graph structure is unknown to the designer. Unfortunately, there exists a bipartite graph G, even one where all vertices have high degree, for which a random linear order matches no more than a $1/2+o(1)$ fraction of the vertices (see Section 4).

In contrast, we show that in the case of Hamiltonian graphs a random linear order guarantees a competitive ratio strictly greater than $1/2$ (Section 4). A similar proof approach applies to regular graphs as well.

Results for a random linear order [Theorem 6.2]: *There is some constant $\rho_0^{\mathrm{H}} > 1/2$ such that for every Hamiltonian graph $G \in G_n$, regardless of n, a random linear order π results in $\rho \geq \rho_0^{\mathrm{H}}$. Similarly, there is some constant $\rho_0^{\mathrm{R}} > 1/2$ such that for every d-regular graph G, regardless of d, n, a random linear order π results in $\rho \geq \rho_0^{\mathrm{R}}$.*

Failed attempt 2: iterative upgrading. A second attempt would be to iteratively "upgrade" unmatched vertices, with the hope that the iterative process will reach a state where many vertices will be matched. That is, in every iteration consider the worst order σ for the current linear order π and move all unmatched vertices (in the matching produced by (π,σ)) to be ranked higher in π. This algorithm is similar to the *k-pass Category Advice* algorithm of [1], but with the difference that in [1] σ remains unchanged throughout the k iterations. In [1] it was shown that in their setting, the fraction of matched vertices approaches $2/(1+\sqrt{5}) \simeq 0.619$ as k grows. In contrast, in Section 7 we show that in our setting this process can go on for $\log n$ iterations before reaching a linear order that matches more than a half of the vertices. This fact gives some indication that establishing a proof using this operator might be difficult.

Failed attempt 3: degree-based ranking. A third attempt would be to give preference to vertices with lower degrees, as they would have fewer opportunities to be matched to incoming vertices of U. Consider a graph with multiple copies of the subgraph $(u_1,v_1),(u_2,v_2),(u_1,v_2)$ along with two additional vertices u_a,u_b (and their partners v_a,v_b). If we connect all vertices of type v_1 to u_a and u_b, we get that their degree is 3, while the degree of vertices of type v_2 is 2. If π is chosen according to the degree, vertices of type v_2 will be ranked higher than vertices of type v_1. In this case, if σ orders the vertices of type u_1 first, they will be matched to vertices of type v_2, leaving the vertices of type v_1 unmatched. The resulting matching will therefore be of size $(1/2+o(1))n$.

Why is this model interesting mathematically? The setting of max-min greedy matching is easy to state. The deceptively simple problem of getting a ratio bounded away from one half turns out to be quite difficult.

[1] Unless stated otherwise, whenever way say "random linear order," we refer to choosing a linear order uniformly at random.

As discussed above, several natural approaches fail to achieve this. The problem remained open for quite some time, despite attempts to solve it. Indeed, the solution that we find is not simple; it involves taking the best of four algorithms. However, these algorithms are not unrelated. They all share a unifying theme that involves a clean combinatorial property, referred to as a maximal path cover (see Section 2). This theme enabled us to break the barrier of half, but interesting problems remain open, such as whether the bound of 2/3 can be achieved. We hope that the progress made in this paper will motivate and enable further improvement in this interesting problem.

1.2 Additional results

We further establish lower and upper bounds for regular graphs.

Results for regular graphs [Corollary 3.3 and Theorem 3.5]: *For d-regulars bipartite graphs, $\rho \geq 5/9 - O(1/\sqrt{d})$. On the other hand, for every integer $d \geq 1$, there is a regular graph G_d of even degree $2d$ such that $\rho(G_d) \leq 8/9$.*

An additional natural problem is to find the best linear order π, given a graph G. We suspect that this is a difficult computational problem. However, the special case of determining whether there is a *perfect* π (a linear order on V that for every linear order σ leads to a perfect matching) does have a polynomial-time algorithm. (The proof appears in Section 5.)

Theorem 1.1. *There is a polynomial-time algorithm that given a graph $G \in G_n$ determines whether G has a* perfect *π, and if so, computes a perfect π.*

1.3 Application to resource allocation and pricing

Various problems related to online bipartite matching are closely related to problems that attract attention in the algorithms community. Gaining better understanding of the max-min greedy matching problem sheds light on more general problems, some of which are still open. In what follows we elaborate on an application to a pricing problem.

Feldman et al. [9] study the design of pricing mechanisms for allocation of items in markets. The basic setting is a matching market, where v_{ij} is the value agent i obtains from getting item j, and every agent can receive at most a single item. The seller assigns prices to items, and agents arrive in an adversarial order (after observing the prices), each purchasing an (arbitrary) item that maximizes their utility (defined as value minus price). It is shown that, given a weighted bipartite graph (with agents on one side, items of the other side, and weight v_{ij} for the edge between agent i and item j), one can set item prices that guarantee at least half of the optimal welfare for any arrival order. The last result holds in much more general settings, namely settings where buyers have submodular valuations over bundles of items [2], and even in a Bayesain settings, where the seller knows only the (product) distribution from which agent valuations are drawn, but not their realizations.

In the Bayesian setting no item prices can guarantee better than half of the optimal welfare in the worst case. A natural question is whether this ratio can be improved in scenarios where the designer

[2] A valuation is said to be submodular if for every two sets S,T, $v(S)+v(T) \geq v(S \cup T)+v(S \cap T)$,.

knows the realized values of the buyers from the outset [3]. Concretely, do there exist item prices that guarantee strictly more than half the optimal welfare, for any arrival order σ? Not only has this question been open for general combinatorial auctions with submodular valuations, it has been open even for unit-demand buyers, and even if all individual values are in $\{0,1\}$ (henceforth referred to as binary unit demand valuations). In the latter setting, pricing is equivalent to imposing a linear order over the items, hence the max-min greedy matching is a precise formulation for the pricing problem in binary unit-demand settings.

An equivalent scenario is one where in each step the "items player" offers an item, and the "buyers player," upon seeing the item, allocates the item to one of the buyers who want the item (if there are any), and that buyer leaves. The items player is non-adaptive (plays blindfolded, without seeing which buyers remain[4]). The size of the matching that can be guaranteed by the items player is equivalent to the max-min greedy matching problem.

Yet an additional equivalent formulation of the problem is one where the linear order π is imposed over the buyers rather than over the items. The buyers then arrive in the order of π, each taking an arbitrary item they want. One can verify that the size of the matching that can be guaranteed by an ordering over the buyers is equivalent to the max-min greedy matching problem.

1.4 Relation to prior work

Relation to online bipartite matching The max-min greedy Matching problem is a nonstandard version of online problems. In the standard online matching problem [13], the algorithm designer has control over the matching algorithm, but has no control over the arrival order of clients (vertices). Our setting can model a situation in which the designer (the maximizing player) has full control over the arrival order of clients (it knows which "items" in U each "client" in V wants, and it chooses π based on this knowledge), but no control over the matching algorithm (the minimizing player can choose the worst possible match in every step, effectively resulting in a linear order σ).

Karp et al. [13] introduced the *Ranking* algorithm which has a $1-1/\mathrm{e}$ competitive ratio in the online bipartite matching setting [13, 11]. Translating this algorithm to our max-min greedy setting, it amounts to simply selecting π at random, and then the minimizing player selects σ after seeing π. We show that there are bipartite graphs $G \in G_n$ for which with high probability over the random choice of π, there is a choice of σ resulting in $M_G[\sigma,\pi] \leq 1/2+o(1)$. Karp et al. [13] also showed that no algorithm for online bipartite matching has a competitive ratio better than $1-1/\mathrm{e}+o(1)$. This was shown by exhibiting a distribution over "difficult" graphs. Each graph in the support of this distribution has a unique perfect matching, and consequently (see Proposition 2.2), there is a linear order π in the max-min greedy setting that ensures that all vertices are matched (regardless of σ). Hence neither the lower bounds nor the upper bounds known for the online matching model give useful bounds in the max-min greedy model.

There are additional known results for online bipartite matching. For d-regular graphs, Cohen and Wajc [3] present a randomized algorithm that obtains $1-O(\sqrt{\log d}/\sqrt{d})$ in expectation, and a lower

[3]The full information assumption is sensible in repeated markets or in markets where the stakes are high and the designer may invest in learning the demand in the market before setting prices.

[4]We note that when the items player is adaptive (chooses the next item based on what happened in the past), the items player can ensure a perfect matching. This is done as follows: in each step, find a minimal tight set of items, and offer an arbitrary item from that set. Here, a set of items is tight if the number of buyers that want items in the set is equal to the size of the set.

bound of $1-O(1/\sqrt{d})$. This is in contrast to our Theorem 3.5 that shows that ρ is bounded away from 1 even when d is arbitrarily large. For general bipartite graphs, under random (rather than adversarial) arrival order, the deterministic greedy algorithm gives $1-1/e$ and no deterministic algorithm can obtain more than $3/4$ [11]. *Ranking* (which is a randomized algorithm) obtains at least 0.696 of the optimal matching [15] and at most 0.727 [12]. No randomized algorithm can obtain more than 0.823 [16].

Relation to pricing mechanisms Our work is also related to the recent body of literature on pricing mechanisms. Motivated by the fact that in real-life situations one is often willing to trade optimality for simplicity, the study of simple mechanisms has gained a lot of interest in the literature on algorithmic mechanism design. One of the simplest forms of mechanisms is that of posted price mechanisms, where prices are associated with items and agents buy their most preferred bundles as they arrive. Pricing mechanisms have many advantages: they are simple, straightforward, and allow for asynchronous arrival and departure of buyers. Various forms of posted price mechanisms for welfare maximization have been proposed for various combinatorial settings [9, 5, 14, 7]. These mechanisms are divided along several axes, such as item vs. bundle pricing, static vs. dynamic pricing, and anonymous vs. personalized pricing. For any market with submodular valuations, one can obtain $1/2$ of the optimal welfare by static, anonymous item prices [9]. Until the present paper, no results better than $1/2$ were known even for markets with unit-demand valuations with $\{0,1\}$ individual values. For a market with m identical items, there exists a pricing scheme that obtains at least $5/7-1/m$ of the optimal welfare for submodular valuations [7].

2 Proof of main result

The graph $G(U,V;E)$ with $|U|=|V|=n$ has a perfect matching M in which $u_i \in U$ is matched with $v_i \in V$ for every $1 \le i \le n$. For a given i, we refer to u_i and v_i as *partners* of each other. Given a set $S \subset V$, the set of neighbors of S is denoted as $N(S)$ (where necessarily $N(S) \subset U$). In this section we prove our main result.

Theorem 2.1. *Given a bipartite graph $G(U,V;E)$ with a perfect matching $\{(u_i,v_i)\}$, there exists a linear order π that guarantees that the greedy matching will be of size at least $\frac{22}{43}n$, regardless of σ. Moreover, there is a polynomial-time algorithm that chooses π with such a guarantee.*

Our proof approach is as follows. We shall first associate with G an auxiliary directed graph that we refer to as the *spoiling graph* $H(V,D)$. This notion by itself is not new—similar notions appear in previous related work. The new aspect related to the spoiling graph and the key to our approach is a notion of a *maximal path cover*. Given a maximal path cover of the spoiling graph (which, as we show in Proposition 2.3, can be found in polynomial time), we partition the set V of vertices into four classes, depending on their roles in the maximal path cover. The classes are V_1 (singleton vertices), S (start vertices of paths), T (end vertices of paths), and I (intermediate vertices of paths). By considering several carefully chosen orders among the classes of vertices, and also of vertices within the classes, we obtain four possible candidate linear orders for π, denoted π_1,π_2,π_3,π_4. We show that for every bipartite graph with a perfect matching, at least one of these linear orders, if used as π, guarantees that the greedy matching will be of size at least $\frac{22}{43}n$, for every σ. Put in other words, if for each of $\{\pi_1,\pi_2,\pi_3,\pi_4\}$ there is a linear order over U for which the greedy matching is smaller than $\frac{22}{43}n$, this would imply (using properties listed in Lemma 2.4) that the path cover giving rise to these linear orders was not maximal. Our lemma statements provide tools to identify the correct π in polynomial time using properties of the classes V_1, S, T, and I described above.

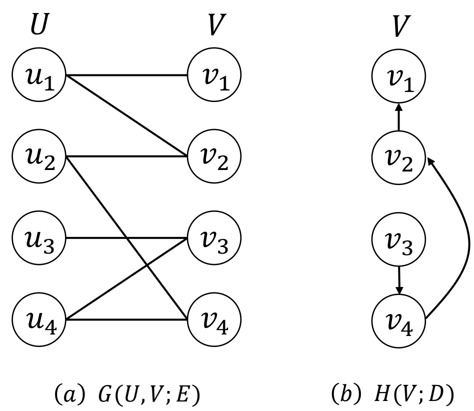

Figure 2: (a) A graph $G(U,V;E)$; (b) The corresponding spoiling graph of G, $H(V,D)$.

We now proceed to define the spoiling graph. Given $G(U,V;E)$, consider a directed graph $H(V,D)$ whose vertices are the set V, and whose set D of directed edges (arcs) is defined as follows: $(v_i, v_j) \in D$ iff $(u_j, v_i) \in E$ (see Figure 2 for an illustration). We refer to $H(V,D)$ as the *spoiling graph* for G, because arc $(v_i, v_j) \in D$ allows for the possibility that edge $(u_j, v_i) \in E$ is chosen into a matching M' in G, spoiling for v_j the possibility (offered by M) of being matched to u_j. Note that this spoiling effect may materialize in a (σ, π) matching only if v_i is ranked higher than v_j in π. Hence the spoiling graph conveys information that may be relevant to the choice of π.

As an example of the information that can be derived from the spoiling graph, consider the following proposition (whose proof can be also obtained as a special case of a result given in [4] and [19] for the more general case of Gross Substitutes valuations).

Proposition 2.2. *If G has a unique perfect matching, then $\rho(G) = 1$.*

Proof. Let $u_i \in U$ and $v_i \in V$ be partners in the unique perfect matching M of G. We claim that the spoiling graph H of G is a *directed acyclic graph* (DAG). Suppose toward a contradiction that H contains a simple directed cycle $v_{i_1}, v_{i_2}, \ldots, v_{i_\ell}, v_{i_1}$. This directed cycle corresponds to the cycle $u_{i_1}, v_{i_1}, u_{i_2}, v_{i_2}, \ldots, u_{i_\ell}, v_{i_\ell}, u_{i_1}$ in G. But removing the edges (u_{i_j}, v_{i_j}), $1 \leq j \leq \ell$ from M and adding the edges $(u_{i_j}, v_{i_{j+1}})$ to M (where $\ell + 1 = 1$) yields a different perfect matching, contradicting the uniqueness of M.

Since H is a DAG, we can topologically sort its vertices and choose a linear order π such that earlier vertices in the topological order have a lower rank in π. This ensures that for every directed edge (v,w) in H, the partner of vertex v will never prefer w over v. Thus, every vertex chooses its partner in M upon arrival. $\square$

We now proceed to define the notion of a maximal path cover. A *directed path* P (whose length is denoted by $|P|$) in H is a sequence of $|P|$ vertices (say, $v_1, \ldots, v_{|P|}$) such that $(v_i, v_{i+1}) \in D$ for all $1 \le i \le |P| - 1$. A single vertex is a path of length 1. A *path cover* of H is a collection of vertex-disjoint directed paths that covers all vertices in V. We consider the following operations that can transform a given path cover to a different one (see Figure 3 for an illustration):

1. *Path merging*: Two paths can be merged into one longer path if $H(V,D)$ has an arc from the end of one path to the start of the other path.

2. *Path unbalancing*: Consider any two paths P and P' with $1 < |P| \le |P'|$, let v_s and v_t denote the first and last vertices of P, and let v'_s and v'_t denote the first and last vertices of P'. If $(v_s, v'_s) \in D$ we may remove v_s from P and append it at the beginning of P'. Likewise, if $(v'_t, v_t) \in D$ we may remove v_t from P and append it at the end of P'.

3. *Rotation*: if there is a path P (say, $v_s, \ldots, v_t$) such that $(v_t, v_s) \in D$, we may add the arc (v_t, v_s) to P (obtaining a cycle), and then remove any other single arc from the resulting cycle to get a path P'. Observe that P' and P have the same vertex set, but they differ in their starting vertex along the cycle $v_s, \ldots, v_t, v_s$.

A path cover is *maximal* if no path merging operation and no path unbalancing can be applied to it, and furthermore, this continues to hold even after performing any single rotation operation.

Proposition 2.3. *Given a bipartite graph $G(U,V;E)$ with a perfect matching $\{(u_i, v_i)\}$, a maximal path cover in the associated spoiling graph $H(V,D)$ can be found in $O(n^3)$ time.*

Proof. Start with the trivial path cover in which each vertex of V forms a path of length 1, and perform arbitrary path merging and path unbalancing operations (some of which are preceded by a single rotation operation) until no longer possible. The process must end within $O(n^2)$ operations, because each path merging and each path unbalancing operation increases the sum of squares of the lengths of the paths, and the sum of squares of the lengths is at most n^2. Each of these operations can be performed in $O(n)$ time. ☐

Given a maximal path cover of H (where p denotes the number of paths in the path cover), sort the paths in order of increasing lengths, breaking ties arbitrarily. Hence $1 \le |P_1| \le |P_2| \le \ldots \le |P_p|$. We consider the following classes of vertices of V:

1. *Singleton vertices* V_1. These are the vertices that belong to paths of length 1 in the given maximal path cover. Let $k = |V_1|$ denote the number of singleton vertices. Observe that $|P_k| = 1$ and $|P_{k+1}| > 1$.

2. *Other vertices* $V_2 = V \setminus V_1$. We partition V_2 into three subclasses of vertices:

   (a) *Start vertices* S. These are the starting vertices of those paths that have length larger than 1. The start vertex of path j, for $k < j \le p$, is denoted by s_j.

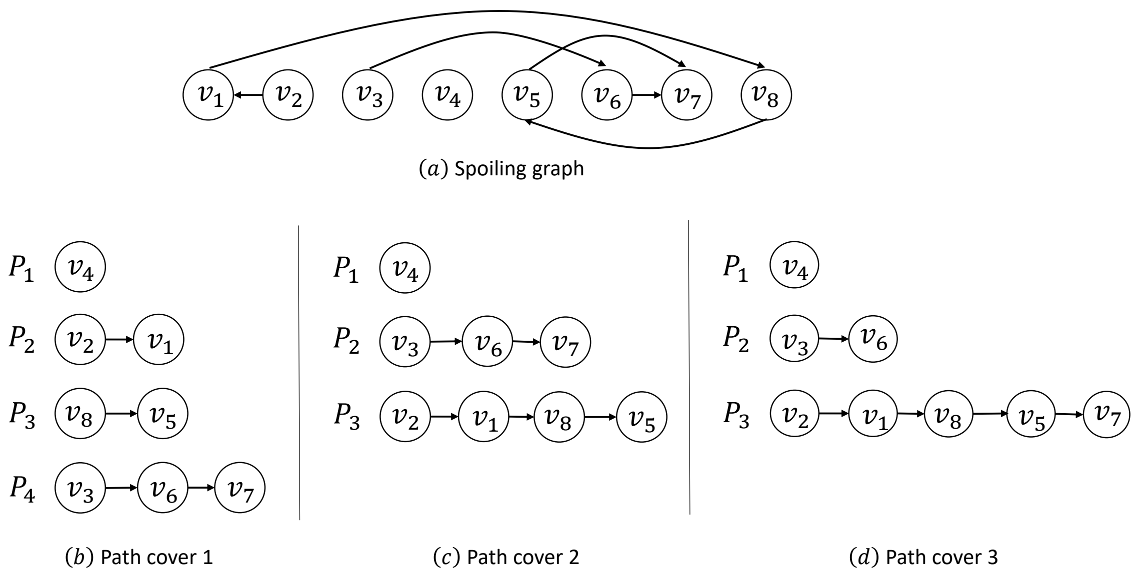

Figure 3: (a) A spoiling graph H; (b) Path cover 1 — a path cover of H; (c) Path cover 2, obtained from Path cover 1 by *merging* paths P_2, P_3; (d) Path cover 3, obtained from Path cover 2 by *unbalancing* paths P_2, P_3. One can verify that Path cover 3 is *maximal*. In Path cover 3, $V_1 = \{v_4\}$, $S = \{v_3, v_2\}$, $T = \{v_6, v_7\}$, $I = \{v_1, v_8, v_5\}$. To illustrate the notion of *rotation*, suppose H had contained two additional edges $(v_7, v_2), (v_1, v_3)$, then Path cover 3 would not have been maximal, as one could *rotate* P_3 to get $P_3' = v_8, v_5, v_7, v_2, v_1$, then merge it with P_2.

(b) *End vertices T*. These are the end vertices of those paths that have length larger than 1. The end vertex of path j, for $k < j \le p$, is denoted by t_j.

(c) *Intermediate vertices $I = V_2 \setminus (S \cup T)$*.

Lemma 2.4. *The classes of vertices listed above have the following properties.*

1. *There are no arcs in H between vertices of V_1. Hence no vertex of V_1 can be a spoiler vertex for another vertex in V_1.*

2. *There are no arcs in H from vertices of V_1 to vertices in S. Hence no vertex of V_1 can be a spoiler vertex for a vertex in S.*

3. *There are no arcs in H from vertices of T to vertices in V_1. Hence no vertex of T can be a spoiler vertex for a vertex in V_1.*

4. *For $i \neq j$, there are no arcs in H from any vertex $t_i \in T$ to any vertex $s_j \in S$. Hence no vertex of T can be a spoiler vertex for a vertex in S, unless they both belong to the same path in the given maximal path cover.*

5. *$(s_i, s_j) \notin D$ for any $s_i, s_j \in S$ with $i < j$. Hence s_i cannot be a spoiler vertex for s_j if $i < j$.*

6. *$(t_j, t_i) \notin D$ for any $t_i, t_j \in S$ with $i < j$. Hence t_j cannot be a spoiler vertex for t_i if $i < j$.*

7. *If for some $s_j \in S$ and $t_j \in T$ (where s_j and t_j are start and end vertices of the same path P_j) we have $(t_j, s_j) \in D$, then there are no arcs from $(T \setminus \{t_j\}) \cup V_1$ to any of the vertices of P_j, and likewise, no arcs from $s_{k+1}, \ldots, s_{j-1}$ to any of the vertices of P_j.*

Proof. All properties follow from the maximality of the path cover. Properties 1,2,3 and 4 hold because otherwise one could perform a path merging operation. Properties 5 and 6 hold because otherwise one could perform a path unbalancing operation. Property 7 holds because otherwise one could perform a rotation operation for path P_j, followed either by a path merging operation (if there is an arc from $(T \setminus \{t_j\}) \cup V_1$ to any of the vertices of P_j) or a path unbalancing operation (if there is an arc from $s_{k+1}, \ldots, s_{j-1}$ to any of the vertices of P_j). $\square$

We now introduce additional notation. Considering only the arcs in D leading from V_2 to V_1, we let M_{21} denote the maximum matching among these arcs. In our analysis we shall consider three parameters:

1. ε_1: its value is such that $k = (1/2 - \varepsilon_1)n$ (recall that $k = |V_1|$ is the number of singleton paths in the maximal path cover). Observe that ε_1 might be negative.

2. ε_2: its value is such that $p = k + \varepsilon_2 n = (1/2 - \varepsilon_1 + \varepsilon_2)n$ (recall that p is the total number of paths in the maximal path cover). Necessarily, $\varepsilon_2 \geq 0$.

3. ε_3: its value is such that $|M_{21}| = (1/2 - \varepsilon_3)n$. Necessarily, $\varepsilon_3 \geq 0$.

Given the above classes of vertices, we consider four possible candidate linear orders for π (denoted $\pi_1, \pi_2, \pi_3, \pi_4$, see below for details). Given some linear order π, we shall use the notation $\rho(\pi)$ to denote the fraction of vertices guaranteed to be matched under π. This fraction will be expressed as a function of the parameters ε_1, ε_2 and ε_3, and we will show that regardless of the value of these parameters, there must be some π with $\rho(\pi) \geq 22/43$.

The following four lemmas present the four candidate linear orders for π along with their corresponding guarantees. Whenever unspecified, the order within a set of vertices can be arbitrary; e. g., for two sets of vertices X, Y, $\pi = X, Y$ means that the set X precedes Y and the order within X, as well as the order within Y, is arbitrary.

Lemma 2.5. *For G and $\pi_1 = V_2, V_1$,*

$$\rho(\pi_1) \geq \frac{1}{n}\left(|V_1| + \frac{|V_2|}{2} - \frac{|M_{21}|}{2}\right) = \frac{1}{2} - \frac{\varepsilon_1}{2} + \frac{\varepsilon_3}{2}.$$

Proof. Let σ be an arbitrary linear order over U. Let m denote the number of vertices in V_2 that are matched to vertices in U_1, where U_1 is the set of partners of V_1. Then $m \le |M_{21}|$. Of the $|V_2| - m$ vertices of V_2 not matched to vertices in U_1, at least half are matched (because for every unmatched vertex from this set, its partner must be matched to a different vertex from this set). In addition, all those vertices of V_1 whose partner is not matched to V_2 are matched, because of property 1 of Lemma 2.4. Hence the total number of vertices matched is at least

$$m + \frac{|V_2| - m}{2} + |V_1| - m \ge |V_1| + \frac{|V_2|}{2} - \frac{|M_{21}|}{2}, \tag{2.1}$$

as desired. □

Lemma 2.6. *For G and $\pi_2 = V_1, V_2$,*

$$\rho(\pi_2) \ge \frac{2}{3} - \frac{1}{3}(\varepsilon_1 + \varepsilon_3).$$

Proof. Let σ be an arbitrary linear order over U. All vertices in V_1 are matched because of property 1 of Lemma 2.4. As to the vertices in V_2, observe that $|N(V_2)| \ge |V_2| + |M_{21}|$, as the set $N(V_2)$ includes the $|V_2|$ partners of V_2, plus at least $|M_{21}|$ additional neighbors in U_1 (due to the matching M_{21}). Moreover, if x vertices are removed from V_2, the number of remaining neighbors is at least $|V_2| + |M_{21}| - 2x$, because each vertex of V_2 contributed at most two neighbors to the lower bound that we provided on the number of neighbors.

Let x denote the number of vertices in V_2 matched under (π_2, σ). Then the size of the matching is $|V_1| + x$, the number of unmatched vertices in V_2 is $|V_2| - x$, and they have at least $|V_2| + |M_{21}| - 2x$ neighbors which have to be matched. Since the number of matched vertices on each side is the same, we have that $|V_1| + x \ge |V_2| + |M_{21}| - 2x$.

We get that

$$3x \ \ge\ |V_2| + |M_{21}| - |V_1| \ = \ n\left(\frac{1}{2} + \varepsilon_1\right) + n\left(\frac{1}{2} - \varepsilon_3\right) - n\left(\frac{1}{2} - \varepsilon_1\right) \tag{2.2}$$

$$= \left(\frac{1}{2} + 2\varepsilon_1 - \varepsilon_3\right) n. \tag{2.3}$$

Therefore, the size of the matching is at least

$$|V_1| + x \ge \left(\frac{1}{2} - \varepsilon_1\right) n + \left(\frac{1}{6} + \frac{2\varepsilon_1}{3} - \frac{\varepsilon_3}{3}\right) n = \left(\frac{2}{3} - \frac{1}{3}(\varepsilon_1 + \varepsilon_3)\right) n. \quad \square \tag{2.4}$$

Lemma 2.7. *For G and $\pi_3 = t_p, \ldots, t_{k+1}, V_1, s_{k+1}, \ldots, s_p, I$,*

$$\rho(\pi_3) \ge \frac{2p - k}{n} = \frac{1}{2} - \varepsilon_1 + 2\varepsilon_2.$$

Proof. In π_3, we refer to the vertices of $T \cup V_1 \cup S$ as the *prefix* of π_3, and to the vertices of I as the *suffix*. Lemma 2.4 implies that within the prefix, the only arcs of H that go from an earlier vertex to a later vertex are of the form (t_j, s_j) (for a path P_j that can undergo a rotation). We claim that regardless of σ, all the prefix will be matched. As the length of this prefix is $2p-k$, this will prove the lemma.

It remains to prove the claim. Suppose first that in the above prefix there are no arcs of H that go from an earlier vertex to a later vertex. Then earlier vertices in this prefix cannot be spoiling vertices for later vertices. Hence indeed, regardless of σ, all the prefix will be matched.

Suppose now that in the prefix of π_3 there are arcs of H that go from an earlier vertex to a later vertex. As noted above, such an arc would be of the form (t_j, s_j). We need to show that even if t_j acts as a spoiling vertex for s_j under π_3 and σ, the vertex s_j will still be matched. Consider the path P_j, and let us rename its vertices as $x_1, \ldots, x_\ell$ (where previously we used $s_j = x_1$ and $t_j = x_\ell$). We wish to show the x_1 would be matched even if x_ℓ is matched to the partner of x_1. The path P_j implies that the partner of x_2 is a neighbor of x_1 in G. Hence x_1 will be matched if no vertex preceding $x_1 = s_j$ in π_3 is matched to the partner of x_2. By property 7 of Lemma 2.4, there is no arc in H from any of the vertices $T \cup V_1 \cap \{s_{k+1}, \ldots, s_{j-1}\} \setminus \{t_j\}$ to x_2, and consequently none of them can be matched to the partner of x_2. As to $t_j = x_\ell$, it might be a neighbor of the partner of x_2 (in fact, it could be that $\ell = 2$), but we already assumed that t_j is matched to the partner of x_1, and hence it is not matched to the partner of x_2. Hence no vertex preceding $x_1 = s_j$ in π_3 is matched to the partner of x_2, and hence s_j will be matched. □

Let V_e (V_o, respectively) denote those vertices of $S \cup I$ that are at even (odd, respectively) distance from the beginning of their respective path. Observe that $S \subset V_e$.

Lemma 2.8. *For G and $\pi_4 = t_p, \ldots, t_{k+1}, V_1, V_o, V_e$,*

$$\rho(\pi_4) \geq \frac{5}{9} - \frac{p}{9n} = \frac{1}{2} + \frac{\varepsilon_1}{9} - \frac{\varepsilon_2}{9}.$$

Proof. Observe that $|V_e| \geq |V_o|$, because in every path (of length above 1) the vertices alternate in entering V_e and V_o, and start with V_e. Observe also that every vertex $v \in V_e$ contributes two distinct neighbors to $N(V_e)$: the partner of v, and the partner of the vertex that follows v on its path (note that the vertex that follows v is not in V_e). Likewise, every vertex $v \in V_o$ contributes two distinct neighbors to $N(V_o)$.

Regardless of σ, all p vertices of T and V_1 are matched, as in Lemma 2.7. For a given σ, let n_o be the number of vertices matched in V_o and let n_e be the number of vertices matched in V_e. Then, $|V_o| - n_o$, the number of unmatched vertices in V_o, satisfies $2(|V_o| - n_o) \leq p + n_o$, because the neighbors of the unmatched vertices in V_o need to be matched to earlier vertices in $T \cup V_1 \cup V_o$. Likewise, $|V_e| - n_e$, the number of unmatched vertices in V_e, satisfies $2(|V_e| - n_e) \leq p + n_o + n_e$. Adding two times the first inequality and three times the second we get that $4|V_o| + 6|V_e| - 4n_o - 6n_e \leq 5p + 5n_o + 3n_e$. Using $|V_o| + |V_e| = n - p$ and $|V_e| \geq |V_o|$, we can replace $4|V_o| + 6|V_e|$ by $5(n-p)$, implying that $9(p + n_o + n_e) \geq 5n - p$, as desired. □

We can now prove Theorem 2.1.

Proof. Observe that $\rho(G) \geq \max_{i\in[1,4]}[\rho(\pi_i)]$.

$$\text{By Lemma 2.5 we have:} \qquad \rho(\pi_1) \geq \frac{1}{2} - \frac{\varepsilon_1}{2} + \frac{\varepsilon_3}{2}. \tag{2.5}$$

$$\text{By Lemma 2.6 we have:} \qquad \rho(\pi_2) \geq \frac{2}{3} - \frac{1}{3}(\varepsilon_1 + \varepsilon_3). \tag{2.6}$$

$$\text{By Lemma 2.7 we have:} \qquad \rho(\pi_3) \geq \frac{1}{2} - \varepsilon_1 + 2\varepsilon_2. \tag{2.7}$$

$$\text{By Lemma 2.8 we have:} \qquad \rho(\pi_4) \geq \frac{1}{2} + \frac{\varepsilon_1}{9} - \frac{\varepsilon_2}{9}. \tag{2.8}$$

Taking a weighted average of the lower bounds provided by the four lemmas, with weights

$$\frac{2}{43}, \frac{3}{43}, \frac{2}{43}, \frac{36}{43},$$

respectively, results in a weighted average of $22/43$. Hence regardless, of the values of ε_1, ε_2 and ε_3, at least one of the lemmas gives $\rho(G) \geq 22/43$. For

$$\varepsilon_1 = \frac{19}{86}, \varepsilon_2 = \frac{10}{86}, \varepsilon_3 = \frac{21}{86},$$

none of the lemmas implies a bound better than

$$\frac{1}{2} + \frac{1}{86} = \frac{22}{43}.$$

The above analysis leads to a polynomial-time algorithm for finding π that ensures $\rho(G) \geq 22/43$. A maximal path cover of $H(V,D)$ can be found in polynomial time by Proposition 2.3. Thereafter, the sets V_1, S, T, V_e and V_o can easily be determined, and likewise, the values of ε_1 and ε_2 can be easily computed. A maximum matching M_{21} (from V_2 to V_1 in H) can be computed in polynomial time using any standard algorithm for maximum bipartite matching. Thereafter, ε_3 can be easily computed. Given the values $\varepsilon_1, \varepsilon_2$ and ε_3, one can determine which of π_1, π_2, π_3 or π_4 provides a higher lower bound on ρ, and use that linear order as π. □

3 Regular graphs

In this section we consider the case where $G(U,V;E)$ is a d-regular bipartite graph with $2n$ vertices. Given that such graphs have d edge-disjoint perfect matchings, one can hope to achieve higher values for ρ for these graphs.

3.1 Positive result

The following known fact (see for example [18]) establishes a lower bound on ρ, as a function of d. A proof is provided for completeness.

Proposition 3.1. *For every d-regular graph $G \in G_n$, the inequality $\rho[G] \geq d/(2d-1)$ holds.*

Proof. Since the greedy algorithm produces a maximal matching, it suffices to show that every maximal matching in a d-regular graph has size at least $\frac{d}{2d-1}n$. To see this, let $S \subset U$ and $T \subset V$ be the sets of unmatched nodes in an arbitrary maximal matching, and suppose $|S| = |T| = (1-\alpha)n$. The nodes in S,T must form an independent set. Consider the size of the edge set connecting S and $V \setminus T$. On the one hand, this size equals $(1-\alpha)nd$ (since all edges from S go to $V \setminus T$); on the other hand, this size is at most $\alpha n(d-1)$ (since at least one edge from each node in $V \setminus T$ goes to $U \setminus S$). Thus, $(1-\alpha)nd \leq \alpha n(d-1)$, implying that $\alpha \geq d/(2d-1)$. Hence we have that $|M_G[\sigma,\pi]| \geq \frac{d}{2d-1}n$, for every π. □

Remark 3.2. For every d there exists a d-regular graph with a perfect matching that admits a maximal matching of size $\frac{d}{2d-1}n$. Suppose that $n = 2d-1$, and consider a d-regular graph where $|S| = |T| = d-1$ for some $S \subset U, T \subset V$, every node in $U \setminus S$ is connected to a single, different node in $V \setminus T$, and to all $d-1$ nodes in T, and every node in $V \setminus T$ is connected to a single, different node in $U \setminus S$, and to all $d-1$ nodes in S. The perfect matching between $U \setminus S$ and $V \setminus T$ is a maximal matching of size $\frac{d}{2d-1}n$.

The lower bound of Proposition 3.1 approaches $1/2$ from above as d grows. The following theorem shows that there exists a linear order π that ensures that the fraction of matched vertices approaches $5/9$. This is a direct corollary from Lemma 2.8 and a theorem in [8].

Corollary 3.3. *For d-regulars bipartite graphs,*

$$\rho \geq \frac{5}{9} - O\left(\frac{1}{\sqrt{d}}\right).$$

Proof. Theorem 3 in [8] shows that every n-vertex d-regular graph has a path cover (referred to as a *linear forest*) with $p = O(n/\sqrt{d})$ paths. This means that there exists a maximal path cover with $O(n/\sqrt{d})$ paths, as the operations used to reach a maximal path cover (path merging, path unbalancing and rotation) can only decrease the number of paths. The desired lower bound on $\rho(G)$ follows by Lemma 2.8. □

Remark 3.4.

1. For small d, the bound of $\rho \geq d/(2d-1)$ which holds for every maximal matching is stronger than the bound in Corollary 3.3.
2. The proof of Corollary 3.3 extends to graphs that are nearly d-regular, by using Theorem 5 from [8].
3. For d-regular graphs, conjectures mentioned in [8] combined with our proof approach suggest that $\rho \geq 5/9 - O(1/d)$.

3.2 Negative result

The following example shows that even in a regular graph with arbitrarily high degree, there may be no linear order π that ensures to match more than a fraction $8/9$ of the vertices.

Theorem 3.5. *For every integers $d,t \geq 1$, there is a regular bipartite graph $G_{d,t}$ of even degree $2d$ and $n = 3dt$ vertices on each side such that $\rho(G_{d,t}) \leq 8/9$.*

Proof. Consider a regular bipartite graph $G(U,V;E)$ with even degree $2d$, and $3d$ vertices on each side. To define the edge set, let $U = U_1 \cup U_2 \cup U_3$ with each U_i of cardinality d, and similarly $V = V_1 \cup V_2 \cup V_3$ with each V_i of cardinality d. For every $i \neq j$, we have a complete bipartite graph between U_i and V_j, and for every i, there are no edges between U_i and V_i.

Let π be an arbitrary linear order over V, let S be the first $2d$ vertices in π, and let T be the last d vertices. Let i be such that $|V_i \cap T|$ is largest (breaking ties arbitrarily). Without loss of generality we may assume that $i = 3$, and then $|V_3 \cap T| \geq d/3$. Hall's condition implies that there is a perfect matching between $U_1 \cup U_2$ and S (and more generally, between $U_1 \cup U_2$ and any $2d$ vertices from V). Hence one can choose a linear order σ over U whose first $2d$ vertices are $U_1 \cup U_2$ that will match the vertices of S one by one. Thereafter, the vertices of $T \cap V_3$ will remain unmatched.

To get the graph $G_{d,t}$ claimed in the theorem, take t disjoint copies of $G(U,V;E)$ above. □

4 Random linear order

In this section we consider scenarios in which the maximizing player is unaware of the graph structure. In such scenarios, the best she can do is impose a random linear order over the vertices in V.

We first show that there exists a graph $G \in G_n$ for which a random linear order does not match significantly more than half of the vertices.

Proposition 4.1. *There exists a bipartite graph $G(U,V;E) \in G_n$ such that a random linear order gets $\rho(G) = 1/2 + o(1)$ almost surely.*

Proof. Consider the graph $G(U,V;E)$, where $U = (U_1,U_2)$, $V = (V_1,V_2)$, and each of U_1,U_2,V_1,V_2 is of size $n/2$. The set of edges constitutes a perfect matching between U_1 and V_1, a perfect matching between U_2 and V_2, and a bi-clique between U_1 and V_2. Let π be a random linear order. Rename the vertices in V_1 such that v_{1j} is the j-th highest vertex from V_1 in the linear order π, and u_{1j} is v_{1j}'s match in the perfect matching between U_1 and V_1. Consider an arrival order σ in which vertex u_{1j} is the j-th vertex (and U_2 arrives after U_1 in an arbitrary order). All vertices of U_1 will be matched (as they arrive first, there are only $n/2$ of them, and each of them has more than $n/2$ neighbors in V_2). Vertices in V_1 can only be matched with vertices of U_1.

Let k denote the number of vertices in V_1 that are matched, and let $v_{1\ell}$ be the highest index (i. e., lowest ranked) vertex in V_1 to be matched (that is, all vertices v_{1j} with $j > \ell$ are unmatched). Necessarily, $v_{1\ell}$ is matched with $u_{1\ell}$. This can happen only if every vertex from V_2 that precedes $v_{1\ell}$ according to π is matched at the time that $u_{1\ell}$ arrives. As $k-1$ vertices of V_1 are matched before $u_{1\ell}$ arrives, there can be at most $\ell - k$ vertices from V_2 that precede $v_{1\ell}$ in the linear order π. Hence, for k vertices from V_1 to be matched, the linear order π should be such that for some ℓ in the range $k \leq \ell \leq n/2$, in the prefix of π of length $2\ell - k$, there are at least ℓ vertices from V_1 and at most $\ell - k$ vertices from V_2. Such an ℓ is said to be *good*.

We claim that the probability that a particular ℓ is good is smaller than the probability that in $2\ell - k$ random tosses of an unbiased coin, at most $\ell - k$ tosses come up heads. To see this, note that the prefix of π is sampled from V without repetition; the random coin version is equivalent to sampling from V with repetition, where in every sample, regardless of past samples, there is probability $1/2$ of sampling a vertex from V_1 (coin coming up tails), and probability $1/2$ of sampling a vertex from V_2 (coin coming up heads). In both sampling methods, the expectation of the number of vertices from V_2 in the prefix is $\ell-(k/2)$. The probability of deviating by more than $k/2$ from the expectation is larger when sampling with repetitions than when sampling without repetition.

Now, by standard bounds on binomial coefficients (or alternatively, by standard Bernstein bounds[5] this probability is at most

$$e^{-k^2/2(2\ell-k)} \le e^{-k^2/4\ell}. \tag{4.1}$$

A union bound over all values of ℓ (in the range $k \le \ell \le n/2$) shows that if $k \ge 2\sqrt{n\log n}$, it is highly unlikely that a good ℓ exists. Consequently, with overwhelming probability, the size of the matching is $(1/2+o(1))n$, whereas $OPT = n$. □

In the above example, if the vertices of V with degree 1 are placed in the prefix of π, then the matching obtained is optimal. This might suggest that prioritizing low-degree vertices in π (and randomizing within sets of vertices of comparable degrees) leads to good performance. However, the example above can be transformed into one where all vertices in V have the same degree, and moreover, each vertex has a degree of $\Omega(\sqrt{n})$. To see this, consider a graph where vertices are partitioned into sets of perfect matchings of size $\sqrt{n}$, $\{(U_{11},V_{11}),\ldots,(U_{1\sqrt{n}},V_{1\sqrt{n}}),(U_{21},V_{21}),\ldots,(U_{2\sqrt{n}},V_{2\sqrt{n}})\}$. Each V_{1i} is also connected in a bi-clique to U_{2i}, and in addition, there are sets U',V' of size $\sqrt{n}$ each connected to the vertices of the other side to balance out the degrees. A similar argument shows that in this graph, a random linear order performs badly as well.

In contrast to the last examples, in some classes of graphs, a random linear order guarantees to match a fraction of the vertices that is bounded away from a half. This is the case, for example, in Hamiltonian graphs. The formal statement and proof are deferred to Section 6.

5 Finding a perfect π

A linear order π over V is said to be *perfect* if for every linear order σ over U, $|M_G[\sigma,\pi]| = n$. A vertex $v \in V$ is *good* with respect to a set of vertices $S \subset V$ if there is no perfect matching between $N(v)$ and S. Given a linear order π over V, let $\overline{\pi}(v)$ be the set of vertices preceding v in π.

Observation 5.1. *π is perfect if and only if every vertex $v \in V$ is good with respect to $\overline{\pi}(v)$.*

Proof. Suppose there exists a vertex $v \in V$ that is not good. Then, consider a linear order σ where the vertices in $N(v)$ are placed first, in an order corresponding to the rank (according to π) of their partners in the perfect matching between $N(v)$ and $\overline{\pi}(v)$. In such a σ, v will not be matched. Now suppose that all vertices in V are good. Then, for every σ, for every $v \in V$, there exists a vertex $u \in U$ that is not matched to a vertex in $\overline{\pi}(v)$; therefore v will surely be matched. □

We present an algorithm that finds a perfect π if one exists, and claims that no such π exists otherwise.

Checking whether v_i is good can be done in polynomial time by running a maximum matching algorithm on $N(v_i)$ and S_i.

[5]These concentration inequalities, commonly misattributed to Chernoff, were published by Sergei N. Bernstein in 1924 and 1937. (See the Wikipedia entry "Bernstein inequalities (probability theory).")

An algorithm for finding a perfect π:

1. Let $S_1 = V$.
2. For $i = 1, \ldots, n$:
   (a) Find an arbitrary good vertex $v_i \in S_i$ with respect to $S_i \setminus \{v_i\}$, and place it in rank $n-i+1$ in π.
   (b) Set $S_{i+1} = S_i \setminus \{v_i\}$.

Lemma 5.2. *If there exists a perfect π, then the algorithm is guaranteed to find a good v_i in every iteration i.*

Proof. Consider some perfect linear order π (not necessarily the one produced by our algorithm), and the suffix $V_{i-1} = v_{i-1}, v_{i-2}, \ldots, v_1$ of vertices chosen in the first $i-1$ iterations of the algorithm. (Of course, there must be a good v_1 at the first iteration, otherwise there is no perfect π.) Let π' be the linear order that places $v_{i-1}, v_{i-2}, \ldots, v_1$ as the lowest-ranked vertices in the same order as the algorithm picked them, and places all other vertices of $V \setminus V_{i-1}$ in a higher rank than V_{i-1} according to their internal order in π.

Since every v_j, $1 \le j \le i-1$, is good with respect to $V \setminus \{v_1, \ldots, v_j\}$, clearly v_j is good with respect to $\overline{\pi'}(v_j)$ (since $\overline{\pi'}(v_j) = V \setminus \{v_1, \ldots, v_j\}$). Now consider a vertex $v \in V \setminus V_{i-1}$. This vertex is good with respect to $\overline{\pi}(v)$, and since $\overline{\pi'}(v) \subseteq \overline{\pi}(v)$, it is clear that v is good with respect to $\overline{\pi'}(v)$. It follows that π' is perfect as well.

Let v_i' be the vertex ranked in position $n-i+1$ in π'. Since π' is perfect, this vertex is good with respect to $\overline{\pi'}(v_i')$. But since $\overline{\pi'}(v_i') \cup \{v_i'\}$ is exactly the set S_i in iteration i, it is guaranteed that the algorithm can find a good v_i in this iteration. $\square$

Let π be the linear order computed by the algorithm. Since every vertex v is good with respect to $\overline{\pi}(v)$, it follows from Observation 5.1 that π is perfect, and Proposition 1.1 follows.

6 Hamiltonian bipartite graphs

In this section we establish two results about Hamiltonian graphs. First, we show that $\rho \ge 5/9$. Note that, since for the case of a Hamiltonian graph there exists a path cover using only a single path (i. e., $p = 1$), Lemma 2.8 directly implies that $\rho \ge 5/9 - 1/9n$. Theorem 6.1 improves this bound to $5/9$. Second, we show that for Hamiltonian graphs, even a random linear order π ensures a ratio that is bounded away from $1/2$. (This is in contrast to general graphs, see Section 4.)

Theorem 6.1. *For Hamiltonian graphs, $\rho \ge 5/9$.*

Proof. Consider a Hamiltonian graph G and a Hamiltonian cycle $u_1, v_1, u_2, v_2, \ldots, u_n, v_n, u_1$ in G that passes through the vertices in the order listed. Let $V_o = \{v_i \,:\, i = 2\ell+1,\ \ell \in \mathbb{N},\ i \le n\}$ be the set of odd labeled vertices of the cycle, and $V_e = V \setminus V_o$.

We first claim that if the number of vertices is even ($|V_o| = |V_e| = n/2$), then $\pi = V_e, V_o$ (and in fact, also $\pi = V_o, V_e$) ensures that $\rho \ge 5/9$. Let n_e and n_o be the number of vertices of V_e and V_o, respectively, matched in $M_G[\sigma, \pi]$ defined using $\pi = (V_e, V_o)$ and an arbitrary σ. Similarly to the proof of Lemma 2.8, it is easy to see that each vertex in V_e contributes two distinct neighbors to $N(V_e)$, and each vertex in V_o contributes two distinct neighbors to $N(V_o)$. (The difference from the proof of Lemma 2.8 is that this property also holds for $v_1 \in V_o$ and $v_n \in V_e$, and this follows because v_1 and v_n contribute u_1 to $N(V_o)$ and $N(V_e)$, respectively.) The number of unmatched vertices in V_e, namely $|V_e| - n_e$, satisfies

$$2(|V_e| - n_e) \le n_e, \tag{6.1}$$

because the neighbors of the unmatched vertices in V_e must be matched to vertices in V_e, as they precede the vertices in V_o in π. Likewise, the number of unmatched vertices in N_o, namely $|V_o| - n_o$, satisfies

$$2(|V_o| - n_o) \le n_e + n_o. \tag{6.2}$$

Adding two times the first inequality and three times the second, we get

$$4|V_e| + 6|V_o| \le 9(n_o + n_e) \;\Rightarrow\; \frac{5}{9} \cdot n \le n_o + n_e. \tag{6.3}$$

As $|V| = n$ and $|M_G[\sigma, \pi]| = n_o + n_e$, this implies that $\rho \ge 5/9$.

We now handle the case where n is odd. Lemma 2.8 ensures that $\frac{5}{9} \cdot n - \frac{1}{9}$ of the vertices are matched by π_4 when the path cover is of a single path. If $\frac{5}{9} \cdot n - \frac{1}{9}$ is not integral, then $\lceil \frac{5}{9} \cdot n - \frac{1}{9} \rceil$ is at least $\frac{5}{9} \cdot n$, thus $\rho \ge \frac{5}{9}$. Therefore, it only remains to handle the case where $\frac{5}{9} \cdot n - \frac{1}{9}$ is integral; namely where $n = 18\ell + 11$ for some integer ℓ (recall that n is odd in this case). In this case, we show that $\pi = V_e, V_o$ ensures that $|M[\sigma, \pi]| > \frac{5}{9} \cdot n$ for every σ. Since $n = 18\ell + 11$, it follows that $|V_o| = 9\ell + 6$ and $|V_e| = 9\ell + 5$. As in the case where n is even, every vertex in V_e contributes two distinct neighbors to $N(V_e)$. As for V_o, every vertex in $V_o \setminus \{v_1, v_n\}$ also contributes two distinct neighbors to $N(V_o)$, and v_1 and v_n contribute (together) to $N(V_o)$ three additional distinct vertices (since they share a vertex along the Hamiltonian cycles). Using the same reasoning as before, it follows that

$$2(|V_e| - n_e) \le n_e \Rightarrow 18\ell + 10 \le 3 \cdot n_e \Rightarrow n_e \ge 6\ell + 3\frac{1}{3}. \tag{6.4}$$

Since n_e is integral, this implies that

$$n_e \ge 6\ell + 4. \tag{6.5}$$

Again, for $|V_o| - n_o$ we have $2(|V_o| - n_o) - 1 \le n_o + n_e$. Rearranging gives us

$$n_e + 3n_o \ge 18\ell + 11. \tag{6.6}$$

Adding twice Inequality (6.5) to the last inequality yields

$$n_e + n_o \ge 10\ell + 6\frac{1}{3} > 10\ell + 6\frac{1}{9} = \frac{5}{9} \cdot |V|, \tag{6.7}$$

which implies $\rho > \frac{5}{9}$. This concludes the proof. □

Next, we show that for the case of a Hamiltonian graph, a random linear order yields a $\rho > 1/2$.

Theorem 6.2. *Consider choosing a linear order π at random. For every Hamiltonian graph G, we have* $E_\pi[\min_\sigma[|M_G[\sigma, \pi]|]] > 0.5012$.

Since the proof is quite involved, we first give a proof overview before providing a formal proof.

6.1 Proof overview

We first provide a high-level overview of our proof.

A linear order π (over V) is said to be *safe* for a set $S \subset V$ if for every linear order σ (over U) the greedy process matches at least one vertex in S (i. e., no σ leaves all vertices in S unmatched). Fix some constant ε. In order to establish that $\rho \geq (1/2+\varepsilon)$, we need to show that there exists a linear order π that is safe for every set S of size $(1/2-\varepsilon)n$. Our proof approach is the following: we show that for a linear order π chosen at random, the expected number (expectation taken over choice of π) of sets of size $(1/2-\varepsilon)n$ for which π is unsafe is smaller than 1. This implies that there exists a linear order π that is safe for all sets of size $(1/2-\varepsilon)n$, as desired.

First, we define a collection of sets that can potentially remain unmatched ("bad" sets). Let B_ε denote the set of all sets $S \subset U$ of size $(1/2-\varepsilon)n$ such that there exists a linear order π that is unsafe for S.

Second, for a given set S and a linear order π we identify a sufficient condition for π to be safe for S. Let $S' \subset S$ be the lowest αn vertices in S (according to π), let v' be the last vertex in S' (i. e., the vertex with rank αn in S'), and let P be the set of vertices in $V-S'$ that precede v' in π. We claim that if the size of P is smaller than the size of $N(S')$ (the neighbors of S'), then π is safe for S. To see this, assume by way of contradiction that π is unsafe for S'. This implies that every vertex in $N(S')$ is matched to a vertex in $V-S'$. Since there are strictly less than $|N(S')|$ vertices in $V-S'$ that precede v', at least one of the vertices in $N(S')$ must be matched to a vertex higher than v'. But, this vertex has a neighbor in S' with lower rank, contradicting the greedy process.

We now proceed by establishing the following three lemmas:

- *Few bad sets lemma:* the size of B_ε is at most $n_B = n_B(\varepsilon)$.
- *Expansion lemma:* given a set $S \subset V$ and parameters α, β, the probability (over a random choice of π) that the lowest αn vertices in S have less than βn neighbors is at most $p = p(\alpha, \beta)$.
- *Good order lemma:* given a set $S \subset V$ and parameters α, β, the probability (over a random choice of π) that the $(\alpha n)^{th}$ lowest vertex in S is higher than βn vertices in $V \setminus S$ is at most $q = q(\alpha, \beta)$.

The three lemmas are combined as follows. For a given set S, due to the sufficient condition identified above, it follows from the union bound that the probability that a random linear order π is unsafe for S is at most $p+q$. Applying the union bound once more over all bad sets (at most n_B sets, as implied by the few bad sets lemma), implies that the probability that a random linear order π is unsafe for some set of size $(1/2-\varepsilon)n$ is at most $n_B(p+q)$. Thus, to conclude the proof, it remains to find parameters such that $n_B(p+q) < 1$.

The good order lemma is independent of the graph structure. In contrast, the expansion lemma and the few bad sets lemma rely heavily on the structure of the graph. As it turns out, Hamiltonian graphs have properties that enable us to establish the two lemmas with good parameters.

6.2 Full proof of Theorem 6.2

We use $H(\cdot)$ to denote the binary entropy function, i. e., given $p \in (0,1)$,

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p). \tag{6.8}$$

Fact 6.3 (Stirling's Approximation)**.** *As $n \to \infty$,*

$$n! = (1+o(1))\sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n. \tag{6.9}$$

Using Stirling's Approximation, one can derive the following bound.

Fact 6.4. *For n and $k = pn$ for some constant $p \in (0,1)$, as $n \to \infty$,*

$$\binom{n}{k} = 2^{(H(p)+o(1))n}. \tag{6.10}$$

We first establish the good order lemma, which is independent of the graph structure.

Lemma 6.5 (Good order lemma)**.** *Let $\alpha < \beta < 1$, $\rho = \frac{1}{2} + \varepsilon$ for some $\varepsilon > 0$ and $\bar{\rho} = 1 - \rho$ such that $\frac{\beta}{\alpha} > \frac{\rho}{\bar{\rho}}$. Let $S \subset V$ be a set of size $\bar{\rho}n$. The probability that in a random linear order π there are at least βn vertices of $V \setminus S$ before αn vertices from S is at most*

$$2^{-(H(\alpha+\beta)-H(\frac{\alpha}{\bar{\rho}})\bar{\rho}-H(\frac{\beta}{\rho})\rho-o(1))n}.$$

Proof. We first analyze the case that in the first $(\alpha+\beta)n$ vertices in π there are *exactly* αn vertices from S. The number of possibilities for this case is $\binom{\bar{\rho}n}{\alpha n}\binom{\rho n}{\beta n}$.

Let $\beta' = \beta + x$ and $\alpha' = \alpha - x$. By the conditions on α, β and ε, we have that $\frac{\beta'}{\alpha'} \geq \frac{\beta}{\alpha} \geq \frac{\rho}{\bar{\rho}}$. Therefore,

$$\beta'\bar{\rho} \geq \alpha'\rho \Rightarrow \beta'\bar{\rho} - \alpha'\beta' \geq \alpha'\rho - \alpha'\beta' \Rightarrow \frac{(\bar{\rho}-\alpha')}{\alpha'} \cdot \frac{\beta'}{(\rho-\beta')} \geq 1 \tag{6.11}$$

$$\Rightarrow \frac{(\bar{\rho}n - \alpha'n + 1)}{\alpha'n} \cdot \frac{\beta'n+1}{(\rho n - \beta'n)} > 1 \iff \frac{\binom{\bar{\rho}n}{\alpha'n}\binom{\rho n}{\beta'n}}{\binom{\bar{\rho}n}{\alpha'n-1}\binom{\rho n}{\beta'n+1}} > 1. \tag{6.12}$$

It follows that $\binom{\bar{\rho}n}{\alpha n}\binom{\rho n}{\beta n} > \binom{\bar{\rho}n}{\alpha'n}\binom{\rho n}{\beta'n}$ for every $\alpha' < \alpha$ and $\beta' > \beta$ such that $\alpha + \beta = \alpha' + \beta'$. Therefore, the probability to have at most αn vertices from S in the first $(\alpha+\beta)n$ vertices in π is at most

$$\frac{\alpha n \cdot \binom{\bar{\rho}n}{\alpha n}\binom{\rho n}{\beta n}}{\binom{n}{(\alpha+\beta)n}} = \frac{2^{(H(\frac{\alpha}{\bar{\rho}})\bar{\rho}+H(\frac{\beta}{\rho})\rho+o(1))n}}{2^{(H(\alpha+\beta)+o(1))n}} \tag{6.13}$$

$$= 2^{-(H(\alpha+\beta)-H(\frac{\alpha}{\bar{\rho}})\bar{\rho}-H(\frac{\beta}{\rho})\rho-o(1))n}, \tag{6.14}$$

where the first equality follows by Fact 6.4. ☐

Let $\rho = 1/2 + \varepsilon$ for some constant $\varepsilon > 0$, and $\bar{\rho} = 1 - \rho = 1/2 - \varepsilon$. The next lemma will be used in order to prove the few bad sets lemma and the expansion lemma. It uses the existence of a Hamiltonian cycle in the graph in order to claim that most sets will have a large number of neighbors. Therefore, a random set will have a large expansion. In addition, there will be few sets of size $(1/2-\varepsilon)n$ with fewer than $(1/2+\varepsilon)n$ neighbors (i. e., few bad sets).

Lemma 6.6. *Let $\alpha \in (0,1/2)$ and $\beta \in (\alpha,1)$ be two constants such that $\delta = \beta - \alpha < \alpha/2$. The number of sets S of size αn where $|N(S)| \leq \beta n$ is at most*

$$2^{(\alpha H(\delta/\alpha)+(1-\alpha)H(\delta/(1-\alpha))+o(1))n}.$$

Proof. Consider a Hamiltonian cycle $H = (v_1, u_1, v_2, u_2, \ldots, v_n, u_n, v_1)$, where $\{v_i\}_{i\in[n]} = V$ and $\{u_i\}_{i\in[n]} = U$. Let S be a set of vertices from V of cardinality αn. Note that in the cycle H, each vertex v of S has two neighbors, where one of these neighbors is joined with an adjacent vertex from V in the cycle. Therefore, the number of neighbors of a sequence of k consecutive vertices of V in H is $k+1$. Thus, the set $N(S)$ is of size αn plus the number of consecutive blocks of vertices from V chosen.

We bound the number of ways to pick at most δn consecutive blocks of vertices from V. We first bound the number of ways to pick *exactly* δn such blocks. In this case, the αn chosen elements have to be within δn blocks. The number of ways to partition αn elements to δn non-empty blocks is $\binom{\alpha n-1}{\delta n-1}$. After deciding the number of elements in each block, we need to figure out their location along the Hamiltonian cycle. $(1-\alpha)n$ elements reside outside of the blocks of the chosen αn elements. We need to choose the location of the first block in H (for which there are n possibilities), and then the number of elements between each pair of consecutive blocks where two blocks are separated by at least one element. The latter is equivalent to dividing $(1-\alpha)n$ elements into δn non-empty segments for which there are $\binom{(1-\alpha)n-1}{\delta n-1}$ possibilities. Overall, there are $n\binom{\alpha n-1}{\delta n-1}\binom{(1-\alpha)n-1}{\delta n-1}$ such possibilities[6].

For $\delta' < \delta$, one can similarly prove the bound of

$$n\binom{\alpha n-1}{\delta' n-1}\binom{(1-\alpha)n-1}{\delta' n-1} \tag{6.15}$$

which is smaller than

$$n\binom{\alpha n-1}{\delta n-1}\binom{(1-\alpha)n-1}{\delta n-1} \tag{6.16}$$

by our conditions on α and δ. Overall, we can bound the number of ways to pick at most δn consecutive blocks of vertices from V by

$$\begin{aligned}
\delta n^2\binom{\alpha n-1}{\delta n-1}\binom{(1-\alpha)n-1}{\delta n-1} &< \delta n^2\binom{\alpha n}{\delta n}\binom{(1-\alpha)n}{\delta n} && (6.17)\\
&= 2^{o(1)n}\cdot 2^{(H(\frac{\delta}{\alpha})+o(1))\alpha n}\cdot 2^{(H(\frac{\delta}{(1-\alpha)})+o(1))(1-\alpha)n} && (6.18)\\
&= 2^{(\alpha H(\frac{\delta}{\alpha})+(1-\alpha)H(\frac{\delta}{(1-\alpha)})+o(1))n}, && (6.19)
\end{aligned}$$

where the first equality follows Fact 6.4. □

The expansion and few bad sets lemmas are obtained as direct corollaries of Lemma 6.6.

[6]Notice there is some overcounting in this argument, but this bound suffices for our purpose.

Lemma 6.7 (Few bad sets lemma for Hamiltonian graphs)**.** *Let* ε *be a constant such that* $\varepsilon < 0.1$. *The number of bad sets in any Hamiltonian graph is at most*

$$|B_\varepsilon| \le 2^{\left(\bar{\rho}H(\frac{2\varepsilon}{\bar{\rho}})+\rho H(\frac{2\varepsilon}{\rho})+o(1)\right)n}.$$

Proof. Notice that if a set S of size $\bar{\rho}n = (1/2-\varepsilon)n$ has more than ρn neighbors, it cannot be left unmatched, since at least one of its neighbors will not be matched to $V \setminus S$. By Lemma 6.6, the number of such sets is at most $2^{\left(\bar{\rho}H(\frac{2\varepsilon}{\bar{\rho}})+\rho H(\frac{2\varepsilon}{\rho})+o(1)\right)n}$. □

We note that this lemma is not true for general graphs. An example of a graph that admits $2^{n/4}$ bad sets is given in Proposition 6.9.

Lemma 6.8 (Expansion Lemma for Hamiltonian graphs)**.** *Consider a set* $S \subset V$ *of size* $\bar{\rho}n$ *and parameters* α, β. *The probability that the lowest* αn *vertices in* S *have fewer than* βn *neighbors is at most*

$$2^{\left(-H(\frac{\alpha}{\bar{\rho}})+\alpha H(\frac{\delta}{\alpha})+(1-\alpha)H(\frac{\delta}{(1-\alpha)})+o(1)\right)\bar{\rho}n}.$$

Proof. Consider a set S of size $\bar{\rho}n$, and the first αn vertices in S in a random linear order. This set is just a random subset of S of size αn. The number of choices of such subset is

$$\binom{\bar{\rho}n}{\alpha n} = 2^{(H(\frac{\alpha}{\bar{\rho}})+o(1))\bar{\rho}n}.$$

Notice that we can apply Lemma 6.6 with set S and bound the number of subsets of S with "small" expansion, even though S is a subset of V, because the same proof applies with respect to a subset of vertices in only one side of a Hamiltonian graph. Therefore, the number of subsets of size αn of S with at most βn neighbors is at most

$$2^{(\alpha H(\frac{\delta}{\alpha})+(1-\alpha)H(\frac{\delta}{(1-\alpha)})+o(1))\bar{\rho}n}.$$

Combining the above, we get that the probability that a random set of αn vertices of S have at most βn neighbors is at most

$$2^{(-H(\frac{\alpha}{\bar{\rho}})+\alpha H(\frac{\delta}{\alpha})+(1-\alpha)H(\frac{\delta}{(1-\alpha)})+o(1))\bar{\rho}n}. \qquad \square$$

Now that we have established the three lemmas, we are ready to prove Theorem 6.2.

Proof of Theorem 6.2. Setting $\varepsilon = 0.0012$, $\alpha = 0.245$ and $\beta = 0.3675$ (and $\rho = \frac{1}{2}+\varepsilon$, $\bar{\rho} = 1-\rho$), we get that these parameters satisfy the conditions for Lemmas 6.5, 6.8 and 6.7.

Applying Lemma 6.7, we get that the size of B_ε is at most $n_B \le 2^{0.044n}$. Applying Lemma 6.8, we get that the probability that the lowest αn vertices of a set of size $\bar{\rho}n$ have fewer than βn neighbors is at most $p \le 2^{-0.86n}$. Applying Lemma 6.5, we get that the probability that for a set S of size $\bar{\rho}n$ the αn-th vertex in a random π comes after βn vertices of $V - S$ is at most $q \leq 2^{-0.45n}$. Combining these three bounds, we get that the probability there exists a set of size $\bar{\rho}n$ unmatched by a random π is at most $n_B(p+q) < 1$, therefore, there must be a π that matches at least one vertex in each set of size $\bar{\rho}n$, and the proof follows. □

This approach can also be used to show that a random linear order is guaranteed to to match more than half of the vertices in every regular graph. On the other hand, Theorem 9.1 in Section 9 shows that one cannot hope to get $\rho > 3/4$ with a random linear order in regular graphs.

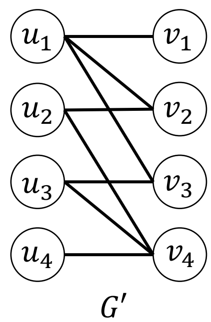

Figure 4: The graph G' admits two bad sets, namely $\{v_1, v_2\}$ and $\{v_1, v_3\}$.

The following proposition gives an example of a bipartite graph that admits many bad sets.

Proposition 6.9. *There exists a bipartite graph $G(U,V;E)$ with n vertices on each side, which admits $2^{n/4}$ bad sets.*

Proof. Consider the bipartite graph G' (with four vertices on each side) depicted in Figure 4. The sets $V_1 = \{v_1, v_3\}$ and $V_2 = \{v_1, v_2\}$ are bad sets. Indeed, under $\pi = \{v_4, v_3, v_2, v_1\}$, the set $\{v_1, v_2\}$ remains unmatched if u_1 and u_2 arrive before u_3 and u_4. Similarly, under $\pi = \{v_4, v_2, v_3, v_1\}$, the set $\{v_1, v_3\}$ remains unmatched if u_1 and u_2 arrive before u_3 and u_4. Now, consider the graph G that is composed of $n/4$ disjoint copies of G'. Since each copy admits two bad sets, G admits $2^{n/4}$ bad sets. □

7 Iterative process

A natural approach for establishing the existence of a good linear order π is the following iterative process of “upgrading” unmatched vertices. Given a linear order $\pi : V \to [n]$ and a linear order $\sigma : U \to [n]$, let $M_G[\sigma, \pi]$ be the result of the greedy matching where vertices in U arrive in order σ (from high to low) and each vertex $u \in U$ is matched to its highest (under π) neighbor (or left unmatched if all its neighbors are already matched).

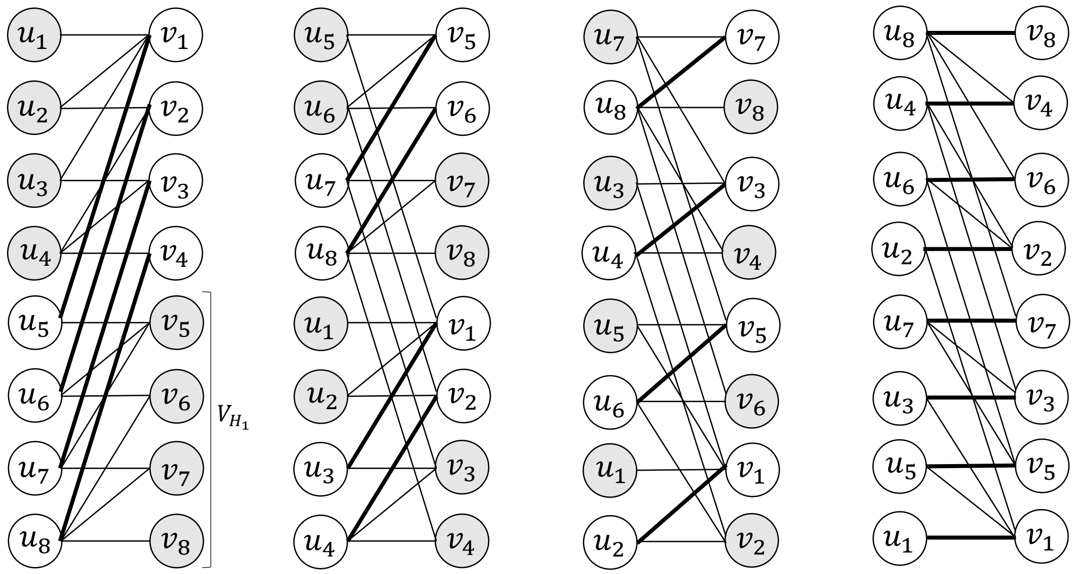

Figure 5: An iterative process where unmatched vertices are given priority. In every iteration thick edges are in the matching; gray vertices are unmatched.

Fix an arbitrary linear order π_1 on V, and let σ_1 be a linear order on U minimizing the greedy matching.[7] Let $M_1 = M_G[\sigma_1, \pi_1]$ be the result of the greedy matching under linear order σ_1 and π_1. If $|M_1|/n$ is some constant greater than $1/2$, then terminate with linear order π_1. Otherwise, partition V into the set V_{H_1} of unmatched vertices (H for *high*, as they will be placed high in the next iteration) and the set V_{L_1} of matched vertices (L for *low*, as they will be placed low in the next iteration).

Consider now a linear order π_2 in which V_{H_1} precedes V_{L_1} (preserving the internal order between vertices in V_{H_1} and similarly between vertices in V_{L_1}), and let σ_2 be a linear order on U minimizing the resulting greedy matching. Let $M_2 = M_G[\sigma_2, \pi_2]$. If $|M_2|/n$ is some constant greater than $1/2$, then terminate with linear order π_2. Else, partition V into the set V_{H_2} of unmatched vertices and the set V_{L_2} of matched vertices, and consider a linear order π_3 in which V_{H_2} precedes V_{L_2} (preserving internal orders). Continue this iterative process until the linear order π_k obtained ensures a matching greater than a half.

The intuition behind this approach is that the unmatched vertices need some "help" in order to be matched, and we provide this help in the form of prioritizing them over matched vertices. One might hope that this process will reach a good linear order within a constant number of iterations. Unfortunately, we show an example where the process goes through $\log n$ iterations before it first obtains a linear order

[7] It is unclear whether σ_1 can be computed in polynomial time. The related problem of computing a minimum maximal matching in bipartite graphs is known to be NP-hard [2]. However, here we consider the existence problem.

ensuring a matching that exceeds $n/2$.

The construction of the graph is inductive. The base is $G_0(U_0,V_0;E_0)$, with two vertices u,v and a single edge between them. For every $i = 1,2,\ldots$, $G_i(U_i,V_i;E_i)$ is such that $|U_i| = |V_i| = 2^i$; it is obtained by taking two (disjoint) copies of G_{i-1}, with additional edges of the form (u_j,v_j) for every u_j from one copy of G_{i-1} to v_j in the second copy of G_{i-1}. An example of G_3 is presented in Figure 5(a). The iterative process is depicted in Figure 5(a)-(d). In all iterations preceding the last one, exactly $n/2$ vertices are matched in the worst σ.

8 Stable matching

The max-min greedy matching problem is also related to the stable matching problem [10] with *partial preferences*. In a stable matching scenario, every vertex in U has a preference order over the vertices in V, and every vertex in V has a preference order over the vertices in U. Given a matching M, a pair of vertices $u \in U, v \in V$ is said to constitute a blocking pair if u and v prefer each other over their partners in M. A matching is said to be stable if no pair of vertices u,v constitutes a blocking pair. The stable matching problem is extended to partial preferences by including an option of being unmatched in the preference order, which can be preferred over being matched to some other vertex. This implies that if being unmatched is preferred over being matched to v by some vertex u, then u being matched to v cannot be stable.

Consider a graph G and linear orders π and σ over V and U, respectively. In the corresponding stable matching with partial preferences instance, a vertex $v \in V$ sets its preferences over vertices in $N(v) \subseteq U$ according to the ranking in π, and prefers being unmatched over vertices in $U \setminus N(v)$. The analogous preferences are set for vertices in U with respect to vertices in V and σ. We show the following.

Proposition 8.1. *The partial preference orders defined by the graph G and linear orders π,σ imply the existence of a unique stable matching. Moreover, the outcome of the greedy matching process is the unique stable matching.*

Proof. First, we show that the preferences defined by G, π and σ imply a unique stable matching. Let $\succ_\pi$ and $\succ_\sigma$ be the preference orders defined by π and σ, where a higher ranking implies a higher preference. Let $u_1 \succ_\sigma u_2 \succ_\sigma \ldots \succ_\sigma u_n$ and $v_1 \succ_\pi u_2 \succ_\pi \ldots \succ_\pi v_n$ be the U-side and V-side vertices ordered according to $\succ_\sigma$ and $\succ_\pi$, respectively. Consider two *different* matchings, M_1 and M_2 that are stable with respect to the preferences defined by G, π and σ. Consider iterating over the U-side vertices according to their ordering until we reach a vertex u_i with a different outcome in M_1 and M_2. There are two scenarios:

- u_i is matched to v_j in M_1 and to v_k for $k \neq j$ in M_2. Assume without loss that $k < j$; that is, $v_k \succ_\pi v_j$ and also $v_k \succ_{u_i} v_j$. In this case we claim that (u_i,v_k) is a blocking pair in M_1. To see this, we first notice v_k is not matched to any u_ℓ with $\ell < i$ in M_1. Otherwise, u_i would not have been the first U-side vertex for which M_1 and M_2 differ. Therefore, either v_k is matched to some u_ℓ in M_1 for which $u_i \succ_\sigma u_\ell$ (and therefore $u_i \succ_{v_k} u_\ell$), or v_k is unmatched in M_1. In both cases, since $v_k \succ_{u_i} v_j$, then both u_i and v_k *strictly* improve their outcome in M_1 by matching to each other.

- u_i is matched in one of the matchings and is unmatched in the other. Without loss of generality, u_i is matched to v_k in M_2 and is unmatched in M_1. As in the first scenario, since the matchings are identical until we reach u_i, either v_k is unmatched in M_1, or it is matched to some u_ℓ for which $u_i \succ_{v_k} u_\ell$. This again implies that (u_i, v_k) is a blocking pair in M_1 as both vertices improve their outcome by matching to each other.

To show that the greedy matching process results in the unique stable matching, we argue there are no blocking pairs. Let M be the resulting matching, and suppose towards a contradiction that (u_i, v_j) is a blocking pair. This means that (*i*) the edge (u_i, v_j) exists in G, and (*ii*) v_j is not matched to a higher ranked vertex than u_i in M. But this implies that when it is u_i's turn in the greedy matching process, v_j is an option. Therefore, it would not end up with an outcome, inferior to being matched to v_j, a contradiction. □

Thus, the max-min greedy matching problem is equivalent to choosing a linear order π over V such that for every linear order σ over U, the unique stable matching with respect to π and σ obtains a large matching.

9 Additional results for regular graphs

The following theorem shows that one cannot hope to get $\rho > 3/4$ with a random linear order in regular graphs.

Theorem 9.1. *For every $\varepsilon > 0$ and sufficiently large d, there are d-regular graphs G for which a random linear order π results in $\rho \leq \frac{3}{4} + \varepsilon$.*

Proof. Consider a d-regular bipartite graph $G(U,V;E)$, where $U = U_1 \cup U_2$, $V = V_1 \cup V_2$, $|U_1| = |V_1| = d$, $|U_2| = |V_2| = d-1$, and $|U| = |V| = 2d-1 = n$. The edge set is $(U_1 \times V_2) \cup (U_2 \times V_1) \cup M$, where M is a perfect matching between U_1 and V_1. Hence the degree of every vertex is d, and (U_2, V_2) forms a balanced bipartite independent set.

Let $Q_1 \subset V$ (a random variable) be the set of first d vertices under the random linear order π, and let $Q_2 = V \setminus Q_1$. Then, $E[|V_2 \cap Q_2|] = (\frac{d-1}{n})^2 n > \frac{n}{4} - 1$. Moreover, as the value of $|V_2 \cap Q_2|$ is concentrated around its mean, then for sufficiently large d (say, $d = \Omega\left(1/\varepsilon^2\right)$)), the actual value of $|V_2 \cap Q_2|$ is above $(1-\varepsilon)n/4$ with high probability. Hence the bounds we prove hold with high probability, and not just in expectation.

Observe that there is a perfect matching M' between U_1 and Q_1: For every vertex $u \in U_1$, if its matched vertex under M is in $V_1 \cap Q_1$ then match u to this vertex, and otherwise match u to an arbitrary vertex in $V_2 \cap Q_1$. The number of vertices in U_1 whose partner under M is not in Q_1 equals $|V_2 \cap Q_1|$, and so this gives a perfect matching. For $1 \leq j \leq d$, let u_j denote the vertex in U_1 that is matched to the jth vertex in Q_1 under M' according to π. Choose a linear order σ over U that begins with $u_1, \ldots, u_d$, and is followed by the vertices of U_2 (unlike the vertices of U_1, the vertices of U_2 can be arranged in an arbitrary order). With this σ, all vertices of U_1 will be matched with Q_1, and consequently no vertex of $V_2 \cap Q_2$ will be matched. This leaves $(1-\varepsilon)n/4$ vertices unmatched, proving the theorem. □

We also establish a few impossibility results for regular graphs of low degree.

Theorem 9.2. *The following hold.*

- *There exists a* 3*-regular bipartite graph G for which* $\rho(G) = \frac{5}{7}$.
- *There exists a* 4*-regular bipartite graph G for which* $\rho(G) = \frac{10}{13}$.

The proof relies on the incidence graphs of finite projective planes. A projective plane consists of a set of lines and a set of points, where every two lines intersect in a single point, every two points are incident to a single line, and there are four points, no three of which are on a line. A projective plane defines a bipartite graph $G(U,V;E)$, called its *incidence graph*, where every vertex $u \in U$ corresponds to a point in the plane, every vertex $v \in V$ corresponds to a line, and there exists an edge between u and v if the point corresponding to u is incident to the line corresponding to v.

Proof. For the first result, we show that $\rho = \frac{5}{7}$ for incidence graph of the Fano plane. The Fano plane is a projective plane consisting of 7 points and 7 lines, with 3 points on every line and 3 lines through every point. Let $G(U,V;E)$ be the incidence graph of the Fano plane. This is a 3-regular bipartite graph. Let $N(V')$ denote the set of neighbors of a set $V' \subset V$. For every set $V' \subset V$ of size $|V'| = 2$ we have $|N(V')| = 5$. We show below that for every such V' there exists a perfect matching between $N(V')$ and $V \setminus V'$. Hence one can choose a linear order σ over U whose first 5 vertices are $N(V')$ that will match the vertices of $V \setminus V'$ one by one. Thereafter, the vertices of V' will remain unmatched. By Hall's condition, it suffices to show that for every set $U' \subset N(V')$ of size $|U'| \leq 5$ we have $|N(U')| \geq |U'| + 2$ (so that Hall's condition holds with respect to the set $V \setminus V'$). Indeed, for every set U' of size 1, $|N(U')| = 3$, for every set U' of size 2, $|N(U')| = 5$, for every set U' of size ≥ 3, $|N(U')| \geq 6$, and for every set U' of size 5, $|N(U')| = 7$. These statements can be quickly verified using the symmetries of the Fano plane. It follows that $\rho(G) = 5/7$.

The second result follows by a similar argument. It is known that there exists a projective plane consisting of 13 points and 13 lines, with 4 points on every line and 4 lines through every point. We claim that $\rho = \frac{10}{13}$ for the incidence graph $G(U,V;E)$ of this projective plane. By the properties of a projective plane, for every set $V' \subset V$ such that $|V'| = 3$, it holds that $|N(V')| \in \{9, 10\}$. We show below that for every such V' there exists a perfect matching between $N(V')$ (and possibly an additional vertex u in case $|N(V')| = 9$) and $V \setminus V'$. Hence one can choose a linear order σ over U whose first 10 vertices are $N(V')$ (possibly with the additional vertex) that will match the vertices of $V \setminus V'$ one by one. Thereafter, the vertices of V' will remain unmatched. By Hall's condition, it suffices to show that for every set $U' \subset N(V')$ such that $|U'| \leq 10$ it holds that $|N(U')| \geq |U'| + 3$ (so that Hall's condition holds with respect to the set $V \setminus V'$). Indeed, for every set U' of size 1, $|N(U')| = 4$, for every set U' of size ≥ 2, $|N(U')| \geq 7$, for every set U' of size ≥ 5, $|N(U')| \geq 11$, for every set U' of size 9, $|N(U')| \geq 12$, and for every set U' of size ≥ 10, $|N(U')| = 13$. It follows that $\rho(G) = 10/13$. □

10 Open problems

In this paper we proved various bounds regarding the max-min greedy matching problem. Our main result was in showing that for every bipartite graph G, one can efficiently compute a linear order π such that for every σ, ρ is greater than a constant strictly greater than $1/2$ (namely, greater than 0.51). In [4], they present a G such that for every π, there is a σ such that $\rho \leq 2/3$. The main open question left by this paper is to close the gap between 0.51 and $2/3$. Moreover, is it possible to compute the optimal π efficiently?

In addition to our result for general bipartite graphs, we also have results for interesting families of bipartite graphs. Specifically, we show that for regular graphs and for Hamiltonian graphs, we are able to compute linear orders with improved bounds. Namely, we show that for regular graphs, $\rho \geq 5/9 - O(1/\sqrt{d})$, and for Hamiltonian graphs, $\rho \geq 5/9$. Moreover, as opposed to general graphs, for these families of graphs, a random linear order guarantees that ρ is greater than a constant strictly greater than $1/2$ (by some tiny quantity). However, there are substantial gaps remaining for these families of graphs as well. The $2/3$ example from [4] is of a Hamiltonian graph, while the only upper bound known on the value of ρ for regular graphs is $8/9$. It will be interesting to see what are the best achievable bounds on ρ, and what are the best guarantees that a random π can achieve.

Finally, the results in our paper can be interpreted as welfare bounds that a seller can achieve by pricing items in full information markets with binary unit-demand valuations. One can ask the same question for more general classes of valuations; for instance, general unit-demand valuations. In this more general version of the problem, the bipartite graph is now weighted, and the seller, knowing the graph, sets prices over the right hand side. Then, the left hand side arrives in some order, and each arriving "buyer" on the left hand side chooses an available "item" on the right hand side that maximizes the weight of the edge *minus* the price of the item. For this more general problem, it is still unknown if the seller can set prices that guarantee that the weight of the resulting weighted matching is at least a ρ-fraction of the maximum weight matching for some constant $\rho > 1/2$. In fact, $2/3$ might actually be the correct answer for this more general problem.

Acknowledgments

A substantial part of this work was conducted in Microsoft Research. We are grateful to Amos Fiat and Sella Nevo for numerous discussions that contributed significantly to the ideas presented in this paper. We also thank Robert Kleinberg for helpful discussions.

References

[1] ALLAN BORODIN, DENIS PANKRATOV, AND AMIRALI SALEHI-ABARI: On conceptually simple algorithms for variants of online bipartite matching. *Theory Computing Sys.*, 63(8):1781–1818, 2019. Preliminary version in WAOA'17. [doi:10.1007/s00224-019-09916-0] 4

[2] MIROSLAV CHLEBÍK AND JANKA CHLEBÍKOVÁ: Approximation hardness of edge dominating set problems. *J. Combinat. Optim.*, 11(3):279–290, 2006. [doi:10.1007/s10878-006-7908-0] 25

[3] ILAN REUVEN COHEN AND DAVID WAJC: Randomized online matching in regular graphs. In *Proc. 29th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'18)*, pp. 960–979. SIAM, 2018. [doi:10.1137/1.9781611975031.62] 6

[4] VINCENT COHEN-ADDAD, ALON EDEN, MICHAL FELDMAN, AND AMOS FIAT: The invisible hand of dynamic market pricing. In *Proc. 17th ACM Conf. Econ. Comput. (EC'16)*, pp. 383–400. ACM Press, 2016. [doi:10.1145/2940716.2940730] 3, 8, 28, 29

[5] PAUL DÜTTING, MICHAL FELDMAN, THOMAS KESSELHEIM, AND BRENDAN LUCIER: Prophet inequalities made easy: Stochastic optimization by pricing non-stochastic inputs. In *Proc. 58th FOCS*, pp. 540–551. IEEE Comp. Soc., 2017. [doi:10.1109/FOCS.2017.56, arXiv:1612.03161] 7

[6] ALON EDEN, URIEL FEIGE, AND MICHAL FELDMAN: Max-min greedy matching. In *Proc. 22nd Internat. Conf. on Approximation Algorithms for Combinat. Opt. Probl. (APPROX'19)*, pp. 7:1–23. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2019. [doi:10.4230/LIPIcs.APPROX-RANDOM.2019.7] 1

[7] TOMER EZRA, MICHAL FELDMAN, TIM ROUGHGARDEN, AND WARUT SUKSOMPONG: Pricing multi-unit markets. *ACM Trans. Econ. Comput.*, 7(4):20:1–29, 2020. [doi:10.1145/3373715] 7

[8] URIEL FEIGE, R. RAVI, AND MOHIT SINGH: Short tours through large linear forests. In *Proc. 17th Integer Prog. Combinat. Optim. (IPCO'14)*, volume 8494 of *LNCS*, pp. 273–284. Springer, 2014. [doi:10.1007/978-3-319-07557-0_23] 15

[9] MICHAL FELDMAN, NICK GRAVIN, AND BRENDAN LUCIER: Combinatorial auctions via posted prices. In *Proc. 26th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'15)*, pp. 123–135. SIAM, 2015. [doi:10.1137/1.9781611973730.10] 5, 7

[10] DAVID GALE AND LLOYD S. SHAPLEY: College admissions and the stability of marriage. *Amer. Math. Monthly*, 69(1):9–15, 1962. [doi:10.2307/2312726] 3, 26

[11] GAGAN GOEL AND ARANYAK MEHTA: Online budgeted matching in random input models with applications to Adwords. In *Proc. 19th Ann. ACM–SIAM Symp. on Discrete Algorithms (SODA'08)*, pp. 982–991. SIAM, 2008. ACM DL. 6, 7

[12] CHINMAY KARANDE, ARANYAK MEHTA, AND PUSHKAR TRIPATHI: Online bipartite matching with unknown distributions. In *Proc. 43rd STOC*, pp. 587–596. ACM Press, 2011. [doi:10.1145/1993636.1993715] 7

[13] RICHARD M. KARP, UMESH V. VAZIRANI, AND VIJAY V. VAZIRANI: An optimal algorithm for on-line bipartite matching. In *Proc. 22nd STOC*, pp. 352–358. ACM Press, 1990. [doi:10.1145/100216.100262] 2, 6

[14] BRENDAN LUCIER: An economic view of prophet inequalities. *SIGecom Exchanges*, 16(1):24–47, 2017. [doi:10.1145/3144722.3144725] 7

[15] MOHAMMAD MAHDIAN AND QIQI YAN: Online bipartite matching with random arrivals: An approach based on strongly factor-revealing LPs. In *Proc. 43rd STOC*, pp. 597–606. ACM Press, 2011. [doi:10.1145/1993636.1993716] 7

[16] VAHIDEH H. MANSHADI, SHAYAN OVEIS GHARAN, AND AMIN SABERI: Online stochastic matching: Online actions based on offline statistics. *Math. Oper. Res.*, 37(4):559–573, 2012. [doi:10.1287/moor.1120.0551] 7

[17] ARANYAK MEHTA: Online matching and ad allocation. *Found. Trends Theor. Comp. Sci.*, 8(4):265–368, 2013. [doi:10.1561/0400000057] 2

[18] JOSEPH NAOR AND DAVID WAJC: Near-optimum online ad allocation for targeted advertising. *ACM Trans. Econ. Comput.*, 6(3–4):16:1–20, 2018. Preliminary version in EC'15. [doi:10.1145/3105447] 14

[19] RENATO PAES LEME AND SAM CHIU-WAI WONG: Computing Walrasian equilibria: Fast algorithms and structural properties. *Math. Programming*, 179:343–384, 2018/2020. Preliminary version in SODA'17. [doi:10.1007/s10107-018-1334-9] 8

AUTHORS

Alon Eden
Postdoctoral Fellow
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA, USA
aloneden@seas.harvard.edu
https://aloneden.github.io

Uriel Feige
Professor
Department of Computer Science and Applied Mathematics
The Weizmann Institute
Rehovot, Israel
uriel.feige@weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~feige

Michal Feldman
Professor
Department of Computer Science
Tel Aviv University
Tel Aviv, Israel
and
Microsoft Research Israel
Herzliya, Israel
michal.feldman@cs.tau.ac.il
http://www.cs.tau.ac.il/~mfeldman

ABOUT THE AUTHORS

ALON EDEN is a postdoctoral fellow in the Harvard EconCS group under the supervision of Yiling Chen and David Parkes. He received his Ph. D. in Computer Science from Tel Aviv University in 2019 under the supervision of Michal Feldman and Amos Fiat. He was awarded the Michael B. Maschler Prize of the Israeli Chapter of the Game Theory Society for the best Ph. D. thesis of 2019. He received a Best Paper award at SAGT'17 and a Best Paper with Lead Student Author award at EC'19. He enjoys traveling outdoors with his family.

Born in Jerusalem, URIEL FEIGE earned his B. Sc. from the Technion, and his M. Sc. and Ph. D. from the Weizmann Institute of Science, both under the guidance of Adi Shamir. After conducting postdoctoral research at Princeton University and at the IBM T.J. Watson Research Center, he joined the Weizmann Institute in 1992. From 2004-2007, he took leave from the Weizmann Institute to work with the Theory Group at Microsoft Research, and he continues to serve as a visiting researcher in Microsoft Research, Herzliya. His main research areas are algorithms, computational complexity, and algorithmic game theory. His work was recognized by some awards, including the 2001 Gödel Prize and the 2005 SIAM Outstanding Paper Prize. He enjoys playing the piano, some sports activities, and spending time with his family.

MICHAL FELDMAN is a Professor of Computer Science in the Blavatnik School of Computer Science at Tel Aviv University and a visiting scholar at Microsoft Research (MSR) Herzliya. Her research spans computer science, economics, game theory and operations research. She received her Ph. D. from the University of California at Berkeley in 2005. She was a visiting professor at Harvard University and Microsoft Research New England (2011-13). She is an Associate Editor of *Games and Economic Behavior, Mathematics of Operations Research, ACM Transactions on Economics and Computation*, and the *Journal of Computer and System Sciences.* She served as Vice Chair of ACM SIGecom, and PC chair of the leading conference on Economics and Computation (ACM EC 2015). She received multiple Best Paper Awards, and the TAU Rector's Award for excellence in teaching. She is an alumna of the Global Young Academy of Sciences, and the Israeli Young Academy of Sciences (and its management committee). She is the recipient of coveted international grants and fellowships, including two ERC grants of the European Research Council, Marie Curie IOF, Alon, ISF and Amazon Research Awards. She was listed in Forbes' list of the 50 most influential women in Israel (2016), and in the The Marker's list of 40 under 40 (2014).