Incentive-Compatible Kidney Exchange in a Slightly Semi-Random Model

AVRIM BLUM, Toyota Technological Institute at Chicago, USA

PAUL GÖLZ, Carnegie Mellon University, USA

Motivated by kidney exchange, we study the following mechanism-design problem: On a directed graph (of transplant compatibilities among patient–donor pairs), the mechanism must select a simple path (a chain of transplantations) starting at a distinguished vertex (an altruistic donor) such that the total length of this path is as large as possible (a maximum number of patients receive a kidney). However, the mechanism does not have direct access to the graph. Instead, the vertices are partitioned over multiple players (hospitals), and each player reports a subset of her vertices to the mechanism. In particular, a player may strategically omit vertices to increase how many of her vertices lie on the path returned by the mechanism.

Our objective is to find mechanisms that limit incentives for such manipulation while producing long paths. Unfortunately, in worst-case instances, competing with the overall longest path is impossible while incentivizing (approximate) truthfulness, i.e., requiring that hiding nodes cannot increase a player's utility by more than a factor of $1 + o(1)$. We therefore adopt a semi-random model where a small ($o(n)$) number of random edges are added to worst-case instances. While it remains impossible for truthful mechanisms to compete with the overall longest path, we give a truthful mechanism that competes with a weaker but non-trivial benchmark: the length of any path whose subpaths within each player have a minimum average length. In fact, our mechanism satisfies even a stronger notion of truthfulness, which we call *matching-time incentive compatibility*. This notion of truthfulness requires that each player not only reports her nodes truthfully but also does not stop the returned path at any of her nodes in order to divert it to a continuation inside her own subgraph.

CCS Concepts: • **Theory of computation** → **Algorithmic mechanism design**; *Random network models*.

Additional Key Words and Phrases: kidney exchange; mechanism design; semi-random models; beyond worst-case analysis

1 INTRODUCTION

For many years, the patients in need of a kidney transplant have far outnumbered the organs available from deceased donors [24]. While many patients can find a friend or relative willing to donate a live organ to them, medical incompatibilities frequently prevent such a direct donation. Kidney-exchange platforms address this problem by matching patients to kidneys across multiple patient–donor pairs. The overarching goal in designing these platforms is to maximize the number of patients who receive transplants, i.e., the *welfare* of the matching. The input to the mechanism can be represented as a *compatibility graph*, a directed graph whose nodes represent patient–donor pairs, and where an edge indicates that the first pair's donor can donate to the second pair's patient.

Traditionally, research on kidney exchange has focused on matchings[1] consisting of disjoint 2-cycles and 3-cycles. Cycles are a natural choice because a paired donor can only be expected to donate a kidney if her patient receives one; the short length of cycles allows surgeries to proceed simultaneously and thus prevents donors from reneging. In this body of work, compatibility graphs are often modeled by classes of densely connected random graphs. If these random graphs are sufficiently large, 2- and 3-cycles can match essentially all pairs that one could hope to match [3]. However, these properties of the random model do not reflect the observation that, in real kidney-exchange platforms, more complex matchings can match substantially more patient–donor pairs [2].

Indeed, subsequent work [2, 11] has highlighted that welfare can be increased substantially by one such form of matchings in particular: long chains starting at non-directed altruistic donors. These donors, which we simply refer to as *altruists* in the rest of the paper, are donors willing to donate a kidney without requiring a transplant in exchange. When an altruist donates a kidney, a compatible patient can receive a kidney *before* her donor donates to another patient. Thus, even if one of the donors should renege, no pair will have donated without having received a kidney in return, which means that simultaneous transplantations are no longer required. Since donors still renege very infrequently [10], an altruistic donor can in principle be at the beginning of an arbitrary-length chain of donations.[2]

In this paper, we study the problem of finding such a long path starting from a given altruist. For this path to obtain high welfare, the path must usually extend across patient–donor pairs in multiple hospitals. However, finding such a path across hospitals is complicated by the fact that the centralized path-finding mechanism cannot directly observe the patient–donor pairs at each hospital and must rely on hospitals to accurately report them. Since hospitals feel primarily bound to their own patients' well-being, a hospital might choose not to register certain pairs if this will lead the platform to match more of its other pairs [4]. Thus, the problem of finding long paths in kidney exchange is one of mechanism design, which means that we search for mechanisms that incentivize hospitals to truthfully report all their clients, which is known as *incentive compatibility*.[3]

In fact, we propose the stronger notion of *matching-time incentive compatibility*, in which hospitals have additional opportunities for manipulation after the mechanism has determined a path. At any node on the path belonging to a hospital, the hospital may receive the kidney and then use it to divert the original path to a continuation within its own subgraph. Matching-time incentive compatibility requires that hiding vertices, diverting the path, or a combination of both does not pay off. We believe this stronger notion of incentive compatibility, which generalizes the widely-used incentive property of individual rationality, to be of independent interest. Throughout

---

[1]As is customary in the kidney-exchange literature, we use the word "matching" differently from most of computer science. A matching is a set of edges in a directed graph such that each node has indegree and outdegree at most one, and for each node (except for the altruist, which we define later), the outdegree is at most the indegree.

[2]As of February 2021, one particular chain has led to 114 successful transplants and is still ongoing.

[3]For ease of exposition, we assume that hospitals can hide nodes, not individual edges. However, as discussed in Section 6, our results generalize to the case where edges can also be hidden.
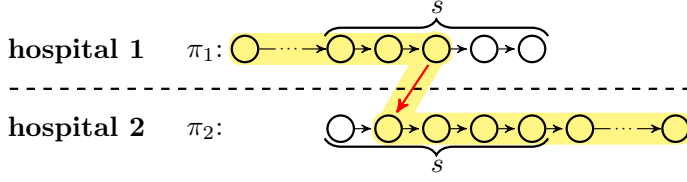
Fig. 1. Illustration for stitching together two paths π_1 and π_2. If the red edge exists, we can stitch the paths together to obtain the yellow path.

this paper, we consider $(1 + o(1))$-approximate versions of these incentive properties. That is, if manipulation can increase a hospital's utility by at most a factor of $(1 + o(1))$, we assume that hospitals can be convinced to participate truthfully, which is in the common interest.

On worst-case graphs, simple counterexamples show that welfare maximization and incentive compatibility are incompatible (Section 3.1). This impossibility holds even for generous approximate notions of welfare maximization and incentive compatibility. On the other extreme, if the compatibility graphs were purely random Erdős-Rényi graphs, the problem would become easy for most parameter settings: either there are no long paths or there is a Hamiltonian cycle. Unfortunately, in practice, kidney-exchange compatibility graphs do not resemble Erdős-Rényi graphs due to issues such as sensitization and the dynamic nature of how patients enter and exit the system.

Adopting a semi-random approach, we show that one can obtain interesting positive results by adding a small number of random edges to worst-case graphs. Notably, edges may be added with probability $1/n^c$ for $1 < c < 2$, meaning that only a vanishing fraction of vertices will be the endpoint of a random edge. To our knowledge, these are the first positive results for semi-random graphs with edge probabilities in $o(1/n)$.

In contrast to purely random models, our semi-random model captures a wide range of compatibility graphs, including very sparse and heterogeneous ones. While the random edges do not directly model a clinical phenomenon,[4] we interpret them as a mild regularity condition on the graph, expressing that a given path in one hospital's subgraph and a given path in another hospital's subgraph are likely to be linked by at least one edge. Since these random edges are only assumed to exist between pairs of different hospitals, the hospital subgraphs themselves can be purely worst case, somewhat like in the semi-random model of Makarychev, Makarychev, and Vijayaraghavan [23].

1.1 Our Techniques and Results

Even in our semi-random model, maximum welfare is at odds with incentive compatibility. This is shown by the family of instances in Section 3.2, in which it is impossible to compete with the optimal path without creating incentives for misrepresentation. As we discuss in Section 3.3, these instances are problematic because any high-welfare path must visit one hospital's subgraph many times but spend few steps there each time, a scenario which incentivizes hospitals to hide outgoing edges to other hospitals. It is natural, then, that we can overcome this impossibility result by benchmarking our mechanism against the longest path such that, for each hospital, the subpaths the hospital's subgraph have some minimum average length.

The random edges in our model are so sparse that they, by themselves, do not form paths of significant length. The main insight, however, is that they allow us to "stitch together" sufficiently

---

[4]One medical justification for these random edges could be that some exchanges perform blood-type incompatible transplantations when tissue-type profiles are well compatible [17]. Since we insert random edges only for a $o(1)$ fraction of agents, such phenomena need not be frequent to support our model.

long paths of worst-case edges inside different hospitals. Indeed, suppose that p is the probability of a random edge existing, and that s is a parameter such that $s^2 p \gg 1$. As illustrated in Fig. 1, let there be two paths π_1, π_2 of length greater than s, each internal to a different hospital. Then, the expected number of edges from the the s last nodes of π_1 to the first s nodes of π_2 is high. If any of these edges do exist, we can use them to "stitch" π_1 and π_2 together. Then, nearly all nodes of π_1 and π_2 will lie on the stitched path except at most $2s$ nodes in the suffix of π_1 and the prefix of π_2.

Roughly speaking, our mechanism has the following structure: From the graph reported by the hospitals, the mechanism chooses a set of disjoint paths within each hospital so that total length is large, and so that truthful reporting maximizes the total length of a hospital's chosen paths. Then, the mechanism stitches these internal paths together, which will succeed with high probability. Successfully stitched paths preserve the welfare and hospital utility except for small stitching losses, which gives approximate guarantees on welfare and incentive compatibility.

Since our mechanism can stitch long enough paths with high probability, it can select the internal paths of hospital i (mostly) based on the subgraph reported by hospital i. As a result, each hospital wants to reveal all its nodes to increase the total length of its chosen paths. If, by contrast, the mechanism chose its path by how well nodes are connected via cross edges, hospitals might hide connections to other hospitals to sway the mechanism to match more of their own nodes instead.

Challenges in translating this high-level structure into concrete mechanisms include:

The choice of original paths. The paths within each hospital initially selected by the mechanism must compete with the benchmark in terms of welfare, and must lend themselves to stitching.

The number of paths per hospital. Stitching only works if we alternate the hospitals on a chain, so we cannot stitch too many paths by a single hospital.

Internal paths too short for stitching. The altruist might only be able to reach long paths via an initial sequence of short internal paths connected by worst-case cross edges. In this case, we do have to consider worst-case cross edges in deciding where to stitch while also ensuring that incentives are limited.

Our main result is the following:

Theorem 8. *Let there be at least two hospitals. Let s, p be parameters varying in n such that $p\,s^2 \in \omega\big(\log(n)^5\big)$. Let $AvgOpt^s$ be the longest path from the altruist such that, for each hospital, the average length of the subpaths in this hospital's subgraph is at least s (or the hospital does not own any nodes on the path).*

Then, there is a mechanism that, with high probability, produces a path of length at least $\big(1 - o(1)\big)\,|AvgOpt^s|$ assuming truthful responses, and that ensures that each hospital can only increase its utility by lower-order terms through manipulation, including matching-time manipulation.

The rest of this paper is organized as follows: In Section 1.2, we discuss related work on kidney exchange and semi-random models. In Section 3, we present lower bounds on the welfare guarantees achievable by incentive-compatible mechanisms, both in the worst case and in our semi-random model. We also explain why our welfare benchmark restricting the average length of subgraphs for each hospital is a natural way to circumvent these impossibilities. In Section 4, we present a simple mechanism that gives lower guarantees in terms of welfare and incentive properties, and also displays key ideas of our full mechanism, which we present in Section 5. Section 6 discusses alternative modeling assumptions and how they would affect the problem. We conclude in Section 7.

1.2 Related Work

There is a rich body of research on algorithm and mechanism design for kidney exchange [1, 3, 6, 7, 19, 26], much of which we have already mentioned. Even though an influential paper by

Ashlagi and Roth [4] specifically identified hospital incentives and altruist-initiated chains as the two main challenges in increasing the welfare of kidney exchange in practice, we know of no existing positive results accounting for both aspects, which is perhaps understandable in light of the robustness of worst-case results as those in Section 3.1.[5] If matchings are restricted to unions of 2-cycles, Ashlagi, Fischer, Kash, and Procaccia [1] give an incentive-compatible mechanism with good welfare guarantees, but their approach does not extend to richer matchings.

Whenever a problem is intractable for worst-case inputs and specific input distributions are hard to justify, semi-random models can allow for positive results while retaining substantial generality. In most works that apply semi-random models, the motivating intractability is computational complexity. For instance, Blum and Spencer [8] pioneered semi-random models in the context of 3-coloring, which is NP hard in the worst case. One of their models starts from an adversarially chosen 3-colored graph and then, independently for each pair of differently colored nodes, toggles the existence of an edge between the two nodes with some small probability p. If $p \geq n^{-0.6+\epsilon}$ (and all colors have $\Omega(n)$ nodes), an efficient algorithm 3-colors the resulting graph with high probability. Since then, many authors have investigated other computational problems in related models [15, 16, 22]; we refer the reader to Feige [14] for a survey of the literature. Semi-random models are also used in smoothed analysis [13, 25], again driven by complexity considerations. As in our work, one can understand positive results as saying that worst-case instances have to be constructed in a very brittle way, which hopefully makes them unlikely to appear in practice.

Only in few cases have semi-random models been applied for desiderata other than complexity. The most prominent example is the use of random-order models for online algorithms [18, 20, 21], where a random arrival order makes the problem more tractable in an information-theoretic sense. In certain versions of the secretary problem, for instance, an adversary chooses a set of quality scores, but the order in which the algorithm sees these scores is chosen uniformly at random. Here, assuming a random arrival order is not a tool for easier computation; instead, it increases the best approximation ratio of *any* algorithm from $1/n$ to $1/e$. Babaioff, Immorlica, and Kleinberg [5] combine these ideas with mechanism design in order to design auctions that are incentive compatible when a set of buyers with worst-case valuations arrive in random order. The fact that monetary transfers are available in this setting gives their mechanisms a flavor decidedly different from ours.

We are not aware of existing work on semi-random models that gives positive results with edge probabilities as low as ours. As two examples of work giving relatively low edge probabilities, Makarychev et al. [23] find planted balanced cuts for edge probabilities in $polylog(n)/n$ for balanced cuts; Chen, Sanghavi, and Xu [9] recover clusters for a similar range of probabilities in a different semi-random model. By contrast, our approach can handle edge probabilities as low as $1/n^{2-\epsilon}$.

Our work is most closely related to that of Blum et al. [7], who design mechanisms for kidney exchange in a different semi-random model. Their model is semi-random in that, in a worst-case compatibility graph, the nodes are allocated randomly to the hospitals. For matchings composed of disjoint constant-length cycles, their mechanism guarantees optimal welfare and individual rationality up to lower-order terms, a much weaker notion than matching-time incentive compatibility. Blum et al. acknowledge that, unfortunately, their approach "does not seem to extend beyond individual rationality," and they provide an impossibility result showing that individual rationality cannot be obtained if long paths from altruistic donors are allowed.

An argument closely related to the stitching technique described in Section 1.1 was used by Dickerson and Sandholm [12] in the proof of their Proposition 2. However, we make much heavier

---

[5]While one paper [19] claims to overcome these impossibilities in a repeated setting, there are concerns about the validity of their results, which we have raised with the authors.

use of this technique: Whereas Dickerson and Sandholm perform a single stitching between two specified subgraphs, we construct a long path consisting of many stitchings across multiple subgraphs.

2 MODEL

Let there be a finite, directed base graph $G_{base} = (V, E_{base})$, where $n := |V|$. Let the vertices V be partitioned into subsets V_i for each of finitely many hospitals i. Let E_p denote the random variable ranging over subsets of $\bigcup_{i \neq j}(V_i \times V_j)$, where each potential edge is included in E_p independently with some given *edge probability* p. While our results hold for a wider range of p, we are primarily interested in very low values of p, which might, for example, scale as $1/n^c$ for $1 < c < 2$. The *compatibility graph* is defined by adding the random edges to the base graph, i.e., as $G := (V, E_{base} \cup E_p)$. Let one node in G be labeled as the altruist α. The hospital owning the altruist node is called the *altruist owner* and is denoted by a. We consider an *instance* to be the base graph together with node ownership, the value of p, and the identity of the altruist.

Each hospital i reports a subset $V_i' \subseteq V_i$ to the mechanism. If $V_i' = V_i$, we say that i reports *truthfully*. The mechanism then learns about all edges in the subgraph of G induced by $\bigcup V_i'$, which means that the mechanism knows this subgraph, node ownership and the identity of the altruist,[6] but does not have access to the remaining nodes and edges, to whether edges are random or deterministic, and to the value of p. Based on this knowledge, the mechanism selects a (simple) path starting at the altruist.

Let the *length* $|\pi|$ of a path π be its number of nodes, and let the *total length* $||P||$ of a set of vertex-disjoint paths P be the sum of their lengths. We say that paths are *disjoint* when they are vertex disjoint. We refer to the length of the path produced by the mechanism as the *welfare*, and to the number of vertices in this path that belong to a certain hospital as the *utility* of this hospital.

We say that edges between nodes of the same hospital are *internal edges*; all other edges are *cross edges*. When cross edges are considered as part of a path, we call them *hops*. A path entirely consisting of internal edges of hospital i is called an *internal* or *i-internal* path, and we refer to the maximal internal subpaths of a path as the *segments* of the path.

The welfare of a mechanism can be measured relative to multiple benchmarks, all of whom we define on the full compatibility graph (rather than just the base graph). We define our strongest benchmark **Opt** as the longest path starting at the altruist. We also define two weaker benchmarks. First, for a given s, let the *long-segments benchmark*, **SOpt^s**, be the longest path from the altruist such that every segment has at least length s, i.e., such that no hop appears too close to another hop, or to the start or end of the path. Second, for some s, let the *high-averages benchmark*, **AvgOpt^s**, be the longest path from the altruist such that each hospital has either no segments in the path, or its mean segment length is at least s. Either of the two weaker benchmarks might not be defined; in these cases, consider the benchmark as the empty path. Note that, on every compatibility graph, the total length of $AvgOpt^s$ and $SOpt^s$ is monotone nonincreasing in s, and that, for every compatibility graph and for every s, $||SOpt^s|| \leq ||AvgOpt^s|| \leq ||Opt||$.

We search for a mechanism that, with high probability over the random edges of a given instance, satisifies the following desiderata:

Efficiency: If all hospitals report truthfully, the welfare of the mechanism is at least a $(1 - o(1))$ fraction of the welfare of a given benchmark.

---

[6]For well-definedness, we assume that the altruist is always reported ($\alpha \in V_a'$). This is without loss of generality: If the altruist is hidden, the mechanism cannot output anything and, thus, no hospital can receive more utility from the mechanism. The more promising strategy of using the altruist for internal matching rather than bringing it to the mechanism will be captured by the definition of matching-time incentive compatibility even without hiding nodes.

Incentive Compatibility: If all hospitals report truthfully except for hospital i, i's utility from the path produced by the mechanism is at most a $(1 + o(1))$ fraction of its utility when all hospitals report truthfully.

Individual Rationality: Under truthful reports, the utility of the altruist owner is always at least a $(1 - o(1))$ fraction of the length of its longest internal path π_{IR} starting at the altruist.

Matching-Time Incentive Compatibility: As for regular incentive compatibility, let π be the path produced by the mechanism when all hospitals except for possibly i report truthfully. Not only must i's utility from π be at most a $(1 + o(1))$ fraction of its truthful utility, but the same must be true for any path π' created by diverting π at one of i's nodes v into an internal continuation.[7]

Note that matching-time incentive compatibility implies incentive compatibility. Matching-time incentive compatibility also generalizes individual rationality since the altruist owner may obtain her longest internal path π_{IR} starting at the altruist by reporting all her nodes and diverting the path at the altruist into π_{IR}.

In Section 6, we discuss the effects of alternative modeling choices on our problem.

3 LOWER BOUNDS

In this section, we give lower bounds on the welfare of all mechanisms that are either approximately individually rational or approximately incentive compatible. For brevity, we present the arguments for individual rationality here and defer the proofs for incentive compatibility to Appendix A of the full version of this paper.

Consider any mechanism in which, with high probability, the altruist owner can increase her utility by at most $o(n)$ by dropping out and matching internally.[8] If compatibility graphs are worst case with no added random edges, we show in Section 3.1 that the hypothetical mechanism will fall short of even our suboptimal benchmarks (even for some $s \in \Theta(n)$) by a gap in $\Theta(n)$ with constant probability. In Section 3.2, we then show that even with randomness, the hypothetical mechanism still cannot compete with the welfare of Opt, again by a gap in $\Theta(n)$ with constant probability.

The second of these lower bounds illustrates why our positive results can compete with the suboptimal benchmarks $SOpt^s$ and $AvgOpt^s$, but not the full Opt. More subtly, these analyses also show that purely efficiency-maximizing mechanisms can create large incentives for manipulation, resulting in hospitals being unlikely to participate truthfully or at all. We show in Section 3.3 that, in instances such as those in Theorem 2, our mechanism actually gives higher utility to each hospital than the pure welfare maximization after accounting for hospital behavior due to incentives. Finally, we explain how the lower bound in Theorem 2 guided our choice of suboptimal benchmarks.

3.1 Worst-case graphs

Theorem 1. *Let $p = 0$, i.e., suppose that no random edges are added to the worst-case base graph. For each deterministic mechanism, there exists a family of instances such that, on every instance, the mechanism will obtain $\Theta(n)$ less welfare than $SOpt^{n/4}$ or violate individual rationality by $\Theta(n)$. (Note that, if the mechanism falls short of $SOpt^{n/4}$, it also falls short of the stronger benchmarks $SOpt^s$ and $AvgOpt^s$ for any $s \leq n/4$.)*

---

[7]Formally: For some node $v \in V_i$ on π, write $\pi = \pi_1\, v\, \pi_2$ for the prefix and suffix, and let π_{ext} be an internal path of hospital i, starting at v, disjoint from π_1, and not necessarily contained in the nodes reported by i. Then, $\pi' = \pi_1 \pi_{ext}$.

[8]Note that our mechanisms fall into this class of mechanisms because, as we will show, hospitals can gain at most an $o(1)$ fraction of the utility obtained under truthful reporting by manipulating, and because the utility obtained under truthful reporting is at most n.
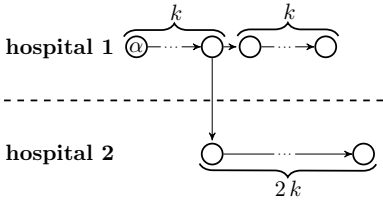
Fig. 2. Without random edges, individually rational mechanisms cannot compete with $SOpt^{n/4}$ and $AvgOpt^{n/4}$.
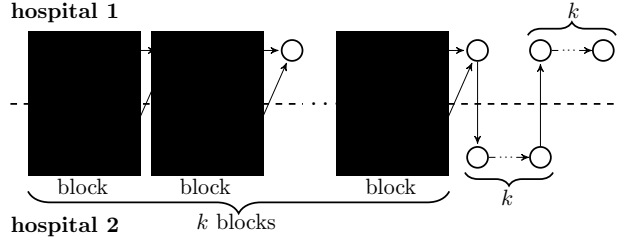

Fig. 3. Even with random edges, individually rational mechanisms cannot compete with maximum welfare.

Proof. For any $k \in \mathbb{N}_{\geq 1}$, consider the instance given in Fig. 2. Clearly, $SOpt^{n/4}$ will select the path of length $3k = 3/4 n$. If the mechanism selects a path of length at most $2k$, it falls short of the optimal welfare by at least $k = n/4$. Else, the path must visit some of hospital 2's nodes, and therefore can visit at most k nodes in hospital 1. However, hospital 1 is the altruist owner and can match all her $2k$ nodes internally outside of the mechanism. Thus, the mechanism must either fall short of the welfare of $SOpt^{n/4}$ by $n/4$, or it must give hospital 1 an incentive of at least $n/4$ to not participate in the mechanism. $\qquad\square$

Appendix A of the full version shows an analogous result for incentive compatibility (rather than individual rationality) with respect to $SOpt^{n/6}$.

At a high level, the graphs that give rise to these results are problematic because they force the mechanism to choose between a branch with higher welfare and a branch that satisfies key individual hospitals (in Theorem 1, the altruist owner), who, if unsatisfied, could remove the high-welfare option altogether by not participating or hiding critical nodes. Intuitively, our semi-random model gives positive results because, when p is large enough, there are enough non-adversarial cross-edges for the mechanism to avoid these problematic decision points in the graph. For example, in the instance in Fig. 2, if random edges allow for the stitching of hospital 1's $2k$-length path in front of the path of hospital 2, then obtaining high welfare and satisfying hospital 1 become compatible.

3.2 Competing with Optimal Welfare in the Semi-Random Model

Unfortunately, the low number of random edges we allow for in our model do not entirely prevent these incompatibilities between incentives and welfare. In particular, within our model, we can still force the mechanism to make these difficult choices on a "small scale" — that is, between segments that are so short that we likely cannot stitch them using random edges. These small-scale choices become a problem for our guarantees in instances where the mechanism must make them repeatedly. We use this observation in the proof of Theorem 2 to construct instances in which our mechanism cannot compete with Opt while maintaining individual rationality. The corresponding result for incentive compatibility is shown in Appendix A of the full version of this paper.

Theorem 2. *Fix any $p \in o(1/n)$ and any deterministic mechanism. Then, there exists a family of instances such that, with $\Omega(1)$ probability, the mechanism will on each instance either obtain $\Theta(n)$ less welfare than Opt or violate individual rationality by $\Theta(n)$.*

Proof. For any $k \in \mathbb{N}_{\geq 1}$, consider the base graph of size $8k + 1$ in Fig. 3. Even relying only on deterministic edges, there is always a path of length $6k + 1$, which can be obtained by branching

into hospital 2's subgraph in every block (k times one node from hospital 1 and three nodes from hospital 2) and going through both segments of length k at the end.

Observe that the number of random edges added to this base graph is binomially distributed with success probability p and at most n^2 trials. With probability $1/2$, the number of random edges is at most its rounded expectation, which is less than $p\,n^2 + 1 \in o(n)$. We therefore condition on the assumption that the number of random edges is at most $p\,n^2 + 1$ for the remainder of the proof.

If the mechanism matches at most $11/4\,k$ of hospital 1's nodes with probability at least $1/2$, then with total probability $1/4$ there is an incentive of at least $3\,k + 1 - 11/4\,k = k/4 + 1 \geq n/32$ for hospital 1 to match outside of the mechanism, concluding the proof of this case.

Else, we additionally condition on the event of more than $11/4\,k$ of hospital 1's nodes being matched. At least $7/4\,k$ of these nodes must lie within the blocks. For the sake of contradiction, assume that the mechanism obtains welfare of at least $(6 - 1/4)\,k + 1 = 23/4\,k + 1$. Then, at least $15/4\,k$ nodes must be matched within the blocks. Since there are $o(n)$ random edges in total, at most $o(n)$ blocks have some incoming random edge. Call the number of these blocks b. In each of the $k - b$ remaining blocks, the mechanism has to decide between the upper and lower branch, i.e., it cannot take more than one of hospital 1's nodes while taking more than 3 nodes in total. Let ℓ be the number of blocks in which the upper branch is chosen. Since at least $7/4\,k$ of hospital 1's nodes are matched, it must hold that $(b + \ell)\,3 + (k - b - \ell)\,1 \geq 7/4\,k$, and thus that $\ell \geq 3/8\,k - b$. Since at least $15/4\,k$ nodes are matched in total, it must hold that $b\,6 + \ell\,3 + (k - b - \ell)\,4 \geq 15/4\,k$ and thus that $\ell \leq 1/4\,k + 2\,b$. Since $b \in o(n) = o(k)$, for large n, these inequalities are in contradiction. We conclude that, with total probability at least $1/4$, welfare is at least $k/4$ lower than $6\,k + 1$, which is a lower bound on the length of Opt. □

3.3 Discussion of Lower Bounds

In the previous two sections, we have shown that both individual rationality and incentive compatibility imply that a mechanism has to match substantially fewer nodes than the longest feasible path. It is natural to wonder whether these incentive properties are worth the loss in welfare.

However, if strong incentives push hospitals towards deviation, the cost to welfare can be significantly higher. Indeed, the example in Fig. 3 illustrates that maximizing welfare without concern for incentives can lead to lower welfare than is given by our mechanisms. By reasoning similar to that used in Theorem 2, a welfare-maximizing mechanism will (with high probability) give the altruist owner $\Theta(n)$ incentive to match outside of the mechanism, giving the altruist owner utility $3\,k + 1$ and hospital 2 utility 0. By contrast, one can verify that, for $s \leq (n - 1)/8$, both mechanisms that we develop in this paper (with high probability) choose a path giving the altruist owner utility in $(4 - o(1))\,k$ and hospital 2 utility in $(1 - o(1))\,k$. Except for $o(n)$ nodes more or fewer, this path will visit all of hospital 1's nodes in the blocks, and all the nodes in the k-length segments of both hospitals. Thus, using our mechanisms rather than naïve welfare maximization benefits both hospitals in this situation.

In the rest of this section, we explain why we consider the long-segments and high-averages benchmarks to be natural responses to the lower bound in Theorem 2. We start with the simple observation that, in our semi-random model, it is advantageous for a hospital to have long internal paths. This is because long internal paths are more likely to have cross-edges that are "useful" to the mechanism — that is, cross-edges that the mechanism could use to increase overall welfare. When a hospital has useful cross-edges to and from its internal paths, the mechanism will incorporate many of the hospital's nodes into a high-welfare path, an outcome which gives utility to the hospital while increasing overall welfare, thereby aligning the objectives of this hospital and the mechanism.

Unfortunately, long internal paths do not guarantee a hospital high utility in the mechanism, because a hospital's long paths give it utility only to the extent that the mechanism can use them to

create welfare. This problem arises in the example in Fig. 3, in which hospital 1 has a long internal path, but yet ends up getting $\Omega(n)$ less utility than it would receive by leaving the mechanism altogether. What happens to hospital 1 in this example illustrates the broader problem addressed by our benchmarks: to even approximate maximum welfare, the mechanism's path must use many short segments of hospital 1's internal path, leaving the remainder chopped into pieces that are (with high probability) too short to be stitched in later. For each of these short segments, then, hospital 1 gains little utility (one node) while losing a comparable amount of potential utility (2 nodes). By this reasoning, any path with welfare close to Opt is unacceptable to hospital 1, which would be better off matching its internal path outside of the mechanism.

Our benchmark $SOpt^s$ addresses this problem by enforcing that, for each hospital, all their segments in the path have length at least s. We will choose s to be much larger than the maximum amount of nodes on a long path that might become useless for stitching due to a single segment. By the reasoning above, this restriction guarantees that the utility each hospital gains from segments included in the mechanism's path outweighs the potential loss in stitchable paths. The high-averages benchmark $AvgOpt^s$ is based on a slightly more refined version of the argument for $SOpt^s$, which recognizes the fact that it suffices if the utility from *all* segments outweighs the losses afflicted by *all* segments, rather than each segment paying off individually. Obviously, there exist base graphs on which obtaining more utility than our benchmarks would not be in conflict with incentives. However, our benchmarks allow us to make a welfare guarantee for all base graphs, which should be high for graphs with long internal paths and ensures that the mechanism makes use of a wide range of opportunities for matching between hospitals.

4 A SIMPLE INCENTIVE-COMPATIBLE MECHANISM

It remains to show that we indeed can compete with the high-averages benchmark while providing matching-time incentive compatibility. To build intuition for the full result, we begin with a simple mechanism, which only satisfies traditional incentive compatibility (but not individual rationality and matching-time incentive compatibility) and only competes with the long-segments benchmark.

Designing mechanisms would be fairly easy if our graphs were Erdős-Rényi style graphs with high edge probability, because the mechanism could simply construct a path including essentially all nodes, and no hospital would have an incentive to deviate. Unfortunately, the low probability with which our random edges appear prevents us from applying this idea. Instead, suppose that, for each hospital i, we fix a set $Paths_i$, which contains disjoint i-internal paths that have length much larger than $2\,s''$ for some s''. Then, consider these paths as the nodes in a meta graph, where the meta graph has an edge between nodes π_1 and π_2 iff G has an edge from one of the last s'' nodes of path π_1 to one of the first s'' nodes of path π_2. Thus, edges in the meta graph exist whenever at least one of the s''^2 potential random edges between π_1 and π_2 exists in G. These edges occur independently between meta edges and with probability $1 - (1 - p)^{s''^2}$. Therefore, if s'' is large enough, the meta graph will be densely connected, and we will be able to stitch all paths in G into a single long path.

The mechanism built from these intuitions is shown in Mechanism 1. It takes in the reported graph including node ownership and the identity of the altruist and a parameter s. Recall that the altruist owner is denoted by a.

Initially, the mechanism defines for each hospital i a set $Paths_i$ of disjoint i-internal paths of equal length $s' \gg 2\,s''$ (line 5). The main complication standing in the way of stitching $\bigcup_i Paths_i$ is that random edges only allow for the stitching of paths in different hospitals. Thus, no hospital can contribute more than half (rounded up for a) of all paths in $\bigcup_i Paths_i$. To deal with this constraint, in line 7, the mechanism redefines $Paths_j$ for the hospital with the largest number of paths such that

---

Mechanism 1: Long-segments mechanism

---

input : reported graph of size $n \gg 1$, parameter $s \gg \log(n)^2$
output : success indicator, path π

1   $s' \leftarrow s/\log n$, $s'' \leftarrow s/\log(n)^2$
2   **if** $\nexists$ *internal path of length s' from α* **then**
3   **return** "success", minimum path $\{\alpha\}$
4   **foreach** *hospital i* **do**
5     $Paths_i \leftarrow$ largest set of disjoint i-internal paths each of length exactly s'; for $i = a$, one path must start at α.
6   $j \leftarrow \text{argmax}_i |Paths_i|$, break ties by hospital id.
7   $Paths_j \leftarrow$ set of disjoint j-internal paths of max. total length subject to (1) the number of paths being at least $\max_{i \neq j} |Paths_i|$ and at most $\sum_{i \neq j} |Paths_i| + \mathbb{1}\{j = a\}$, (2) all paths having size $\geq s'$, and (3) if $j = a$, one path starting at α.
8   order the paths $\bigcup_i Paths_i$ in a hospital-alternating sequence such that the path starting at α is the first in the sequence.
9   check if paths can be stitched in order by finding an edge from the last s'' nodes of the tth path in the sequence to the first s'' nodes of the $(t + 1)$th path for all t.
10   **if** *this stitching succeeds* **then**
11   **return** "success", stitched path π
12   **else**
13   **return** "failure", minimum path $\{\alpha\}$

---

the hospital now has at most as many paths as all other hospitals combined (plus one, for a). In line 8, the mechanism orders all paths in $\bigcup_i Paths_i$ in a sequence such that the path owners alternate. We will show further below that this is always possible by applying the following combinatorial lemma, whose proof we defer to Appendix B of the full paper:

Lemma 3. *Let there be a finite number of balls with colors in $[c]$, such that there are $n_1 \geq n_2 \geq \cdots \geq n_c$ many balls of each color. If $n_1 \leq \sum_{i \geq 2} n_i$, the balls can be arranged in a cyclical sequence such that no two adjacent balls have the same color. In particular, they can be ordered in a linear, color-alternating sequence such that an arbitrary ball can be chosen as the first.*

For our choices of parameters, we will show that, with high probability, the meta graph will be so dense that the chosen order of $\bigcup_i Paths_i$ forms a path in the meta graph, which means that the paths in $\bigcup_i Paths_i$ can be stitched in order. Assuming that this stitching succeeds, the stitched path is returned in line 11; in the low-probability event of failure, the mechanism returns in line 12.

We now comment on the well-definedness of Mechanism 1. In line 1, we ignore the rounding of s' and s''. Since we assume s to be large, this does not influence the asymptotics of our analysis. The check in line 2 ensures that line 5 is well-defined. Line 7 is well-defined because an appropriately chosen subset[9] of the original $Paths_j$ satisfies the requirements. For line 8, we need to show that such an ordering exists. If $|Paths_j| \leq \sum_{i \neq j} |Paths_i|$, Lemma 3 directly gives us a sequence. Else, by line 7, it must be that $j = a$ and $|Paths_j| = \sum_{i \neq j} |Paths_i| + 1$. Then, Lemma 3 allows to order all paths in $\bigcup_i Paths_i$ except for the one starting at α[10] such that the sequence starts with a hospital other than the altruist owner. Then, adding the altruist path to the front of the sequence will produce a

---

[9]That is, if $j \neq a$, any subset of size $\max_{i \neq j} |Paths_i|$; else, any subset of size $\min\left(1, \max_{i \neq j} |Paths_i|\right)$ that includes the path starting at α.

[10]If the path starting at α is the only path in $\bigcup_i Paths_i$, an ordering exists trivially.

valid hospital-alternating sequence. Finally, in line 9, for large n, note that $s' = \log n\, s'' \gg 2\, s''$, so that the nodes used for stitching into and out of a path do not overlap.

With these things out of the way, our theorem is the following:

Theorem 4. *Let there be at least two hospitals, and let s, p be parameters varying in n such that $p\, s^2 \in \omega\big(\log(n)^5\big)$. Then, for any sequence of instances with such n and p, running Mechanism 1 on the full graph succeeds with high probability. In this case, call the resulting path π. It must hold that $|\pi| \geq (1 - o(1))\, |SOpt^s|$ and, by hiding vertices, any hospital i can receive at most utility $(1 + o(1))\, |\pi \cap V_i|$ from the mechanism.*

To prove this theorem, we show in the following lemmas that the mechanism indeed succeeds on the truthful graph with high probability (Lemma 5), that it achieves high welfare in this case (Lemma 6), and that, if the mechanism succeeds on the truthful graph, manipulation can only increase utility by lower-order term (Lemma 7).

Lemma 5. *If the parameters p and s vary in n such that $p\, s^2 \in \omega\big(\log(n)^5\big)$, Mechanism 1 succeeds with high probability under truthful reports.*

PROOF. If the branch in line 2 is taken, the mechanism succeeds deterministically. Else, it will get to line 9. Even if we only consider the random edges, each individual stitching fails with probability at most $(1-p)^{(s'')^2}$. Adding the deterministic edges can only decrease this probability. Using a union bound over at most n paths, we conclude that the stitching — and thus the mechanism — fails with probability at most

$$n\,(1-p)^{s''^2} \leq \exp(\log n - p\, s^2/\log(n)^4) \to 0. \qquad \square$$

Lemma 6. *Suppose that $s \in \omega\big(\log(n)^2\big)$. Whenever Mechanism 1 succeeds on the truthful reports, it achieves welfare of at least $(1 - o(1))\, |SOpt^s|$.*

PROOF. Assuming that all hospitals are truthful, consider the path $SOpt^s$. In lines 4 and 5, every path segment in $SOpt^s$ of length $\ell \geq s$ could be cut up into $\lfloor \ell/s' \rfloor \geq \ell/s' - 1 \geq (1 - 1/\log n)\, \ell/s'$ disjoint paths of length s', which means that $Paths_i$ must contain at least $(1-1/\log n)\, |SOpt^s \cap V_i|/s'$ many paths. Thus, the total length of each set $Paths_i$ is at least $(1 - 1/\log n)\, |SOpt^s \cap V_i|$.

Some additional reasoning is needed to show that the same remains true after $Paths_j$ is redefined in line 7. Namely, $\sum_{i \neq j} |Paths_i| + \mathbb{1}\{j = a\}$ could be lower than the number of s''-length paths needed to cover a $(1 - 1/\log n)$ fraction of nodes in j's segments in $SOpt^s$. Suppose that this is the case, and let $SOpt^s$ contain r many segments in total.

Suppose for now that $j \neq a$. Then, hospital j cannot have owned more than $r/2$ of the segments in $SOpt^s$ since the ownership alternates between segments and the altruist owns the first segment. By the reasoning of the first paragraph of this proof, the at least $r/2$ many segments belonging to other hospitals mean that the $Paths_i$ for all $i \neq j$ together contain at least $\log n\, r/2$ many paths, which, for large n, is larger than $r/2$. Then, the redefinition of $Paths_j$ in line 5 permits at least $r/2$ paths of *minimum* length s', which allows to cover all nodes in $SOpt^s \cap V_j$.

Else, i.e., if $j = a$, essentially the same reasoning applies. Hospital j cannot have owned more than $\lceil r/2 \rceil$ of the segments in $SOpt^s$, and the $Paths_i$ for $i \neq j$ will together contain at least $\log n\, \lfloor r/2 \rfloor$ paths. For large n, this number is at least $\lceil r/2 \rceil - 1$, which means that the redefinition of $Paths_j$ allows for at least $\lceil r/2 \rceil$ paths; again, all nodes in j's segments of $SOpt^s$ could be covered by a set of paths satisfying the constraints in the redefinition of $Paths_j$. Thus, whether $j = a$ or not, the total length of $Paths_j$ continues to be at least $(1 - 1/\log n)\, |SOpt^s \cap V_j|$.

Through the stitching process, the length of each path in $\bigcup_i Paths_i$ (which has size at at least s') might decrease by up to $2\, s''$, so each hospital i receives utility of at least $(1 - 2/\log n)\,(1 - $

$1/\log n) |SOpt^s \cap V_i| \geq (1 - 3/\log n) |SOpt^s \cap V_i|$ from the returned path. Adding up over all hospitals, it follows that the social welfare is at least a $(1 - 3/\log n)$ fraction of $|SOpt^s|$. □

Lemma 7. *Again, suppose that $s \in \omega(\log(n)^2)$. Assuming that Mechanism 1 succeeds under truthful reports and returns a path π, each hospital can only gain $(1 + o(1)) |\pi \cap V_i|$ utility by hiding vertices.*

PROOF. If the mechanism returns a trivial path in line 3 under truthful reporting, hiding nodes cannot change this outcome. If a manipulation causes the mechanism to return in lines 3 or 13, the mechanism returns a subpath of the truthful path, and no hospital can gain from this manipulation. We may thus assume that the mechanism successfully returns in line 11 under truthful reports and only consider manipulations that make the mechanism return in line 11 as well.

First, consider a hospital i who cannot form enough paths of length s' in line 5 to become the special hospital j. Let r be the number of paths in $Paths_i$ under truthful reporting. Since the mechanism succeeded, truthful reporting guarantees a utility of at least $r (s' - 2 s'')$ to hospital i. Since no manipulation can increase the number of chosen paths in $Paths_i$ and since only nodes in $\bigcup_i Paths_i$ can end up in a stitched path, the possible utility is bounded above by $r s' = 1/(1 - 2/\log n) r (s' - 2 s'')$. Thus, the benefits of manipulation are of lower order.

Second, consider a hospital j who, under truthful reporting, will have the highest number of paths. Fix the updated $Paths_j$ under truthful reporting. If a manipulation by j does not change the fact that the hospital is chosen as j in line 6, any such manipulation can only reduce the total size of $Paths_j$ in line 7. Again, such a change can only be advantageous in terms of the stitching loss, which is lower-order as above.

If, however, the manipulation turns this hospital into one of the hospitals i who are not chosen in line 6, the manipulation must lead to $Paths_i$ as chosen in line 5 containing at most $\max_{i' \neq i} |Paths_{i'}^s|$ paths. Thus, $Paths_i$ contains fewer paths than the updated $Paths_j$ under truthful reporting, and each path has smaller length. It follows that the total length of $Paths_i$ decreases in the manipulation, which means that the hospital may at most gain lower-order utility from a more advantageous stitching. □

Note that, when we defined s' as $s/\log n$ and s'' as $s/\log n$, any other function $f(n) \in \omega(1)$ could have replaced the logarithm. By choosing a function f that grows very slowly, the precondition of Theorem 4 can be slightly weakened to $\lim_{n \to \infty} p s^2 / f(n)^4 - \log n \to \infty$.

5 MATCHING-TIME INCENTIVE COMPATIBILITY WITH HIGHER WELFARE

As we showed in the previous section, Mechanism 1 guarantees incentive compatibility while competing with the long-segments benchmark in terms of welfare. While this positive result is encouraging, we would like to strengthen both the incentive properties and the efficiency of our mechanism. Before explaining our stronger Mechanism 2 in more detail, we explain why these properties are needed and how we modify the original mechanism to obtain them.

5.1 Strengthening Mechanism 1: Desiderata and Changes

Stronger notions of incentive compatibility. Mechanism 1 guarantees incentive compatibility, i.e., that hiding nodes cannot increase a hospital's utility from the path returned by the mechanism. Unfortunately, it is not clear that the real path of transplantations will be what the mechanism suggests. Indeed, at any point in the execution of the transplantation chain, a hospital might pretend that the chain ended at one of its vertices because the donor reneged, and then use the donated kidney to make more internal matches. Even without such foul play, the lack of individual rationality in Mechanism 1 might disincentivize the altruist owner from participating in the mechanism.

Mechanism 2: High-averages mechanism

---

input : reported graph of size $n \gg 1$, parameter $s \gg \log(n)^2$
output: success indicator, path π

1 $s' \leftarrow s/\log(n)^2$

2 **foreach** *hospital i* **do**

3     $paths_i \leftarrow \log n \times$max. number of disj. *i*-internal paths such that average path length $\geq s$.

4 say that hospitals *i* with $paths_i > 0$ are *active*.

5 **if** *a inactive or if only a is active* **then**

6   **return** "success", π_{IR}

7 $j \leftarrow \text{argmax}_i |Paths_i|$, break ties by hospital id.

8 $paths_j \leftarrow \min(paths_j, \sum_{i \neq j} paths_i)$

9 **foreach** *active hospital i* **do**

10     $Paths_i \leftarrow$ set of up to $paths_i$ disjoint *i*-internal paths of maximum total length.

11 from each $Paths_i$, remove all paths of length less than $4\,s'$

12 **foreach** *active hospital i* **do**

13   **repeat** $paths_i - |Paths_i|$ **times**

14   split a suffix of length $4\,s'$ off the end of longest path in $Paths_i$; add to $Paths_i$.

15 $Exp \leftarrow \emptyset, v \leftarrow \alpha$

16

17

18

19

20

21

22

23

24

25   ,

26 **if** $|Exp| < s'$ **then**

27   **return** "success", π_{IR}

28 **else if** *some neighbor of $v \in Exp$ is in first s' nodes of some $\pi \in \bigcup_i Paths_i$* **then**

29   **return** stitch2()

30 **else**

31   **return** "failure", π_{IR}

Clearly, to satisfy individual rationality, the mechanism must be more generous to the altruist owner than in line 3 of Mechanism 1, choosing a long internal path instead of just the altruist. Mechanism 2 does so, most relevantly in lines 6 and 27, where π_{IR}, the longest internal path starting at the altruist, is returned.

More importantly, while the requirement in Mechanism 1 that all paths in $Paths_i$ for $i \neq j$ have fixed length s' is convenient for analysis, it can incentivize hospitals to divert the path at the end of their last segment to extend the length of the segment. To address this shortcoming of Mechanism 1, our new mechanism must be more flexible in how it selects the $Paths_i$, perhaps in a way that more closely resembles the redefinition of $Paths_j$ in line 7 of Mechanism 1. However, while allowing for

---

Procedure stitch1

---

1    $\sigma' \leftarrow$ path from α through Exp to v, then σ

2   let $\langle \pi, \pi_2, \ldots, \pi_r \rangle$ be a hospital-alternating ordering of $\bigcup_i Paths_i$ starting at π such that the owner of π_r is not the owner of π.

3   **if** $|\pi| < 2\,s'$ **then**

4       $seq \leftarrow \langle \sigma', \pi_2, \ldots, \pi_r \rangle$

5   **else**

6       $seq \leftarrow \langle \sigma', \pi_2, \ldots, \pi_r, \pi \rangle$

7   for each t, try find an edge between the last s' nodes in the tth path in seq to the first s' nodes in the $(t+1)$th path.

8   **if** *this stitching succeeds* **then**

9   **return** "success", stitched path

10   **else**

11   **return** "failure", π_{IR}

Procedure stitch2

---

1    $\sigma \leftarrow$ the path from α through Exp to v

2   arrange $\bigcup_i Paths_i$ in hospital-alternating sequence starting with π.

3   try to stitch the sequence of paths in order, using up to s' nodes each time; add the already established stitch between σ and π to the front.

4   **if** *this stitching succeeds* **then**

5   **return** "success", stitched path

6   **else**

7   **return** "failure", π_{IR}

---

a wider range of $Paths_i$, we still need to ensure that we can stitch the resulting paths, i.e., that all paths in $\bigcup_i Paths_i$ have a certain minimum length, and that no hospital has more paths than all other hospitals combined. Mechanism 2 accomplishes this by first determining a number of paths $paths_i$ for each hospital i (lines 2, 3, 7 and 8). It then initializes $Paths_i$ by selecting up to $paths_i$ paths in a way that maximizes the total length of this set without any further constraints (line 10), and modifies $Paths_i$ afterwards to make sure that its paths have an adequate minimum length and that their number is exactly $paths_i$ (lines 11 to 14). We will have to show that this always succeeds on large enough graphs, and that welfare and incentives are preserved.

Higher welfare. On the welfare side, a severe limitation of Mechanism 1 is that, whenever the altruist is not at the start of an internal path of length at least s', the mechanism returns a trivial path. This shortcoming is reflected in the benchmark $SOpt^s$, which is also trivial in these cases. On the one hand, the lack of welfare in these scenarios is understandable because random edges can no longer reliably connect the altruist to long paths in other hospitals. On the other hand, there might be deterministic cross edges that can play this role, possibly going back and forth between different hospitals for a while before reaching long paths. If the altruist has great connectivity to paths, but these paths happen to be in other hospitals' subgraphs, we would still like our mechanism to make use of this connectivity. The high-average benchmark $AvgOpt^s$ is a way to capture this potential. In particular, it allows the number of segments in each hospital to be arbitrarily high as long as some of these segments are long enough to make the stitching worthwhile, as we discussed in Section 3.3.

In contrast to Mechanism 1, no path in $\bigcup_i Paths_i$ has to start at the altruist in Mechanism 2. Thus, rather than stitching immediately, Mechanism 2 starts a graph search (from line 15 onward), which explores nodes reachable from the altruist until it finds a node that exists on a path in the set of predetermined paths $\bigcup_i Paths_i$. Not only is this node required to lie on a predetermined path, however; it also must also be far enough from the end of its predetermined path that, with high probability, we can stitch in another path after it. If this search successfully finds a node on a predetermined path π, its suffix and prefix might get stitched in separately. Since a matching-time manipulation could exploit this fact to visit a $(paths_i + 1)$th path, all nonzero $paths_i$ must be chosen large enough in line 3 that a single additional path gains only lower-order amounts of utility.

Another complication is that this graph search cannot explore too many nodes because, if the initial path becomes too long, its contribution to some hospitals' utility might no longer be lower-order. As a result, hospitals might hide nodes to force the graph search to succeed later, which might increase the length of this initial path. To counter this, the mechanism returns with a failure in line 31 if the search is unsuccessful after exploring more than s' nodes. Fortunately, this failure mode is unlikely to happen under truthful reports due to the random edges. Additional cases we consider are those in which the search reaches a node on a path, but that node is too close to its end to start stitching (line 24), or if too few nodes are reachable from the altruist (line 27).

5.2 Main Result

We can now state our main result in full:

Theorem 8. *Let there be at least two hospitals, and let s, p be parameters varying in n such that $p\,s^2 \in \omega\big(\log(n)^5\big)$. Then, for any sequence of instances with such n and p, running Mechanism 2 under truthful reports succeeds with high probability. In this case, call the resulting path π. Then, it must hold that $|\pi| \geq \big(1 - o(1)\big)\,|AvgOpt^s|$ and that, by any manipulation captured by matching-time incentive compatibility, i can receive at most utility $\big(1 + o(1)\big)\,|\pi \cap V_i|$.*

The proof of this theorem is deferred to Appendix C of the full version of this paper. There, we begin by going through the steps of the algorithm, which also serves to show that Mechanism 2 is well defined. These observations then enable us to prove that Mechanism 2 succeeds with high probability, obtains high welfare, and is matching-time incentive compatible. Taken together, these lemmas establish our main theorem.

6 ALTERNATIVE MODELING ASSUMPTIONS

As is often the case in the design of semi-random models, there is a range of alternatives to the modeling choices we made in Section 2. In this section, we discuss how some of these alternative assumptions would affect our mechanism-design problem.

For example, our model allows hospitals to hide nodes but not individual edges, a standard assumption in the existing kidney literature [1, 4, 7, 19, 26]. In fact, however, the proofs of incentive compatibility for both algorithms go through even if hospitals can hide not only their nodes but also internal edges and cross edges adjacent to their nodes. It follows that our mechanisms retain the same guarantees in this stronger model.

Additionally, we restricted our model to a single altruist and thus a single path, which simplifies the notation and makes the algorithmic problem more natural. For appropriate generalizations of the benchmark and matching-time incentive compatibility, we believe that Mechanism 2 and Theorem 8 can be extended to multiple altruists.

Concerning the use of semi-randomness, adversarial and random steps can be combined in many ways.[11] For a fixed set of nodes, fixed node ownership, a fixed altruist, and a fixed probability p, we can partition our semi-random model into four stages:

(1) **"I":** The adversary adds internal edges.
(2) **"A":** The adversary adds cross edges.
(3) **"R":** Random cross edges are added with probability p.
(4) **"B":** The efficiency benchmarks are computed.

The 24 permutations of the four phases "I", "A", "R", and "B" define different semi-random models, which we will discuss exhaustively. Note that some of these permutations define equivalent models since, if the adversarial phases "I" and "A" are adjacent, their relative order does not matter. Furthermore, we are only interested in models where the benchmark (phase "B") is computed after the internal edges and at least some cross edges are inserted.

Among the remaining models, there are some ("ARIB", "RIAB", "RIBA") in which the adversary can add internal edges adaptively to the random cross edges. Assume that each hospital owns $\Omega(n)$ vertices and that $p \in o(1/n)$. Then, in these semi-random models, the adversary can (with high probability) create variants of the worst-case instances in Fig. 2, effectively removing the benefit we get from randomness. Indeed, with high probability, the altruist does not have random outgoing edges, only a vanishing fraction of each hospital's nodes have random edges, and there exists an isolated cross edge from the altruist owner to another hospital. If this is the case, the adversary can build the internal paths using only nodes without random edges, and can use the isolated random cross edge to replicate the shape of the worst-case example. As a result, any mechanism that competes with $SOpt^{c \cdot n}$ in welfare[12] cannot give individual rationality or incentive compatibility with high probability. Because the argument used no adversarial cross edges, it equally applies to "ARIB", "RIAB", and "RIBA".

Perhaps the most interesting variant of our semi-random model is "IRAB", which differs from our model in that the adversarial cross edges may depend on the random cross edges. Note that this model is more general than ours and that our lower bounds apply as a result. Since Mechanism 1 does not consider the existence of cross edges at all before choosing which path to stitch, adversarial cross edges can only increase the probability of success, which means that Theorem 4 continues to hold. Unfortunately, the same cannot be said about Mechanism 2, which considers cross edges in the graph search. By adaptively adding cross edges, an adversary could guide the graph search into parts of the graph where stitching will fail. While a version of Mechanism 2 without the graph search can guarantee matching-time incentive compatibility for the long-segments benchmark, we do not have results for the high-averages benchmark in this alternative model.

The two remaining models, "IABR" and "IRBA", are variants of our main model and "IRAB", respectively. In these variants, the benchmark is computed not on the full compatibility graph, but on the graph with either the random or adversarial cross edges still missing. Because the welfare attained by these benchmarks will be lower than on the full graph, positive results immediately carry over. Since the proof of Theorem 2 did not make use of the fact that the optimal benchmark might increase due to random edges, this lower bound also applies to "IABR".

7 DISCUSSION

For studying the mechanism-design problem of finding long chains in kidney exchange, we proposed a semi-random model for compatibility graphs. Since this model only assumes the existence of a

---

[11]This is similar to work the work of Kolla et al. [22], who show how, in the context of unique games, four different models arise from adding randomness in different phases.

[12]For a small enough constant $c > 0$, depending on the number of nodes per hospital.

very low number of random edges, this assumption can be seen as a mild regularity condition on the compatibility graph. In this semi-random model, we developed mechanisms that limit incentives for manipulation to lower-order terms while competing with strong benchmarks in terms of welfare.

We do not claim that the mechanisms designed in this paper are ready for use in practice. Whether they are depends on questions not yet investigated: whether the asymptotic guarantees have traction in graphs of typical size, whether hospitals can be convinced to report their pairs truthfully despite possible lower-order gains from manipulation, and whether the high-averages benchmark gets reasonably close to the optimal welfare on practical kidney-exchange graphs. Moreover, our model abstracts from crucial aspects of the practical problem, including the existence of multiple chains and the fact that practical compatibility graphs evolve dynamically while transplantations on the chain are executed.

Thus, rather than proposing a complete solution, we hope that our work can serve as a stepping stone for future mechanisms coordinating kidney exchange with altruist-initiated chains. Specifically, we have demonstrated that, despite the ubiquitous impossibilities of this setting, kidney exchange with donor chains is a valid target for mechanism design — as long as one is willing to trade traditional worst-case analysis for beyond worst-case methodologies.

ACKNOWLEDGMENTS

REFERENCES

[1] I. Ashlagi, F. Fischer, I. A. Kash, and A. D. Procaccia. 2015. Mix and Match: A Strategyproof Mechanism for Multi-Hospital Kidney Exchange. *Games and Economic Behavior* 91 (2015), 284–296.

[2] I. Ashlagi, D. Gamarnik, M. A. Rees, and A. E. Roth. 2012. *The Need for (Long) Chains in Kidney Exchange*. Technical Report.

[3] I. Ashlagi and A. E. Roth. 2011. Individual Rationality and Participation in Large Scale, Multi-Hospital Kidney Exchange. In *Proceedings of the 12th ACM Conference on Electronic Commerce*. 321âĂŞ322.

[4] I. Ashlagi and A. E. Roth. 2012. New Challenges in Multihospital Kidney Exchange. *American Economic Review* 102, 3 (2012), 354–59.

[5] M. Babaioff, N. Immorlica, and R. Kleinberg. 2007. Matroids, Secretary Problems, and Online Mechanisms. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*. 434–443.

[6] P. Biró, D. F. Manlove, and R. Rizzi. 2009. Maximum Weight Cycle Packing in Directed Graphs, with Applications to Kidney Exchange Programs. *Discrete Mathematics, Algorithms and Applications* 1, 4 (2009), 499–517.

[7] A. Blum, I. Caragiannis, N. Haghtalab, A. D. Procaccia, E. B. Procaccia, and R. Vaish. 2017. Opting into Optimal Matchings. In *Proceedings of the 28th Annual ACM-SIAM Symposium on Discrete Algorithms*. 2351–2363.

[8] A. Blum and J. Spencer. 1995. Coloring Random and Semi-Random k-Colorable Graphs. *Journal of Algorithms* 19, 2 (1995), 204–234.

[9] Y. Chen, S. Sanghavi, and H. Xu. 2012. Clustering Sparse Graphs. In *Proceedings of the 25th Conference on Neural Information Processing Systems*. 2204–2212.

[10] N. Cowan, H. A. Gritsch, N. Nassiri, J. Sinacore, and J. Veale. 2017. Broken Chains and Reneging: A Review of 1748 Kidney Paired Donation Transplants. *American Journal of Transplantation* 17, 9 (2017), 2451–2457.

[11] J. P. Dickerson, A. D. Procaccia, and T. Sandholm. 2012. Optimizing Kidney Exchange with Transplant Chains: Theory and Reality. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 2. 711–718.

[12] J. P. Dickerson and T. Sandholm. 2017. Multi-organ exchange. *Journal of Artificial Intelligence Research* 60 (2017), 639–679.

[13] M. Englert, H. Röglin, and B. Vöcking. 2014. Worst Case and Probabilistic Analysis of the 2-Opt Algorithm for the TSP. *Algorithmica* 68, 1 (2014), 190–264.

[14] U. Feige. 2020 (in press). Introduction to Semi-Random Models. In *Beyond the Worst-Case Analysis of Algorithms*, T. Roughgarden (Ed.). Cambridge University Press.

[15] U. Feige and J. Kilian. 1998. Heuristics for Finding Large Independent Sets, with Applications to Coloring Semi-Random Graphs. In *Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science*. 674–683.

[16] U. Feige and R. Krauthgamer. 2000. Finding and Certifying a Large Hidden Clique in a Semirandom Graph. *Random Structures & Algorithms* 16, 2 (2000), 195–208.

[17] P. Ferrari, P. D. Hughes, S. J. Cohney, C. Woodroffe, S. Fidler, and L. DâĂŹOrsogna. 2013. ABO-Incompatible Matching Significantly Enhances Transplant Rates in Kidney Paired Donation. *Transplantation* 96, 9 (2013), 821–826.

[18] A. Gupta and S. Singla. 2020 (in press). Random-Order Models. In *Beyond the Worst-Case Analysis of Algorithms*, T. Roughgarden (Ed.). Cambridge University Press.

[19] C. Hajaj, J. P. Dickerson, A. Hassidim, T. Sandholm, and D. Sarne. 2015. Strategy-Proof and Efficient Kidney Exchange Using a Credit Mechanism. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*.

[20] C. Karande, A. Mehta, and P. Tripathi. 2011. Online Bipartite Matching with Unknown Distributions. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*. 587–596.

[21] C. Kenyon. 1996. Best-Fit Bin-Packing with Random Order. In *Proceedings of the 7th Annual ACM-SIAM Symposium on Discrete Algorithms*. 359–364.

[22] A. Kolla, K. Makarychev, and Y. Makarychev. 2011. How to Play Unique Games against a Semi-Random Adversary: Study of Semi-Random Models of Unique Games. In *Proceeding of the 52nd Annual IEEE Symposium on Foundations of Computer Science*. 443–452.

[23] K. Makarychev, Y. Makarychev, and A. Vijayaraghavan. 2012. Approximation Algorithms for Semi-Random Partitioning Problems. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing*. 367–384.

[24] F. McCormick, P. J. Held, and G. M. Chertow. 2018. The Terrible Toll of the Kidney Shortage. *Journal of the American Society of Nephrology* 29, 12 (2018), 2775–2776.

[25] D. A. Spielman and S.-H. Teng. 2004. Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time. *J. ACM* 51, 3 (2004), 385–463.

[26] P. Toulis and D. C. Parkes. 2015. Design and Analysis of Multi-Hospital Kidney Exchange Mechanisms Using Random Graphs. *Games and Economic Behavior* 91 (2015), 360–382.