High Satisfaction in Thin Dynamic Matching Markets

Johannes Bäumler, Martin Bullinger, Stefan Kober, Donghao Zhu
Technical University of Munich
{johannes.baeumler, martin.bullinger, stefan.kober, donghao.zhu}@tum.de

Abstract

Dynamic matching markets are an ubiquitous object of study with applications in health, labor, or dating. There exists a rich literature on the formal modeling of such markets. Typically, these models consist of an arrival procedure, governing when and which agents enter the market, and a sojourn period of agents during which they may leave the market matched with another present agent, or after which they leave the market unmatched. One important focus lies on the design of mechanisms for the matching process aiming at maximizing the quality of the produced matchings or at minimizing waiting costs.

We study a dynamic matching procedure where homogeneous agents arrive at random according to a Poisson process and form edges at random yielding a sparse market. Agents leave according to a certain departure distribution and may leave early by forming a pair with a compatible agent. The objective is to maximize the number of matched agents. Our main result is to show that a mild guarantee on the maximum sojourn time of agents suffices to get almost optimal performance of instantaneous matching, despite operating in a *thin* market. This has the additional advantages of avoiding the risk of market congestion and guaranteeing short waiting times. We develop new techniques for proving our results going beyond commonly adopted methods for Markov processes.

1 Introduction

Matching problems are the basis of one of the most vibrant research areas at the intersection of economics, operations research, and computer science with an abundance of applications encompassing labor markets, school choice, child adoption, dating platforms, allocating advertisements, ride-sharing, or barter exchange, which includes in particular kidney exchange. Therefore, it is not surprising that an extensive stream of research deals with modeling and decision making in matching markets. In reality, many matching scenarios have a dynamic flavor in the sense that agents arrive and get matched over time, and we contribute to the research on such variants of matching markets.

The central concern in matching markets is to find a matching of high quality. This leads to the algorithmic challenge of deciding when and how to match agents. In particular, this means a careful timing of performing matches. A repeatedly discussed and observed economic paradigm is that it can be beneficial to defer decisions (such as pairing two agents to form a match) to gain longterm benefits from improved later decisions based on an increased range of possibilities or larger knowledge. In other words, delaying decisions *thickens* a market, and especially in the research on matching markets, there are a lot of theoretical results promoting market thickness (see, e.g., Akbarpour et al., 2020; Baccara et al., 2020; Loertscher et al., 2022; Emek et al., 2016; Shimer and Smith, 2001; Leshno, 2021, and our extensive discussion in the related work). Often, this means that agents have to wait in the market for some time to make optimal decisions. Still, under certain conditions, it is possible that instantaneous matching algorithms have a good performance (Ünver, 2010; Ashlagi et al., 2019b, 2022). The reason for this is that in these models, a thick market arises even under quick matching decisions, and therefore, there is no reason to 'wait and thicken the market'. By contrast, it is not indisputable whether market thickness is always desirable due to the risk of congestion (Roth, 2018). In fact, the analysis of real-life data indicates that market thickness may lead to bad outcomes with respect to the obtained size (Li and Netessine, 2020) and quality (Fong, 2020) of matchings.

Our main result observes the high quality of matchings computed by instantaneous decisions in a *thin* market. More precisely, we consider a stochastic model following Akbarpour et al. (2020) and Anderson et al. (2017), where agents arrive and depart from a market governed by certain random distributions. In particular, if agents are guaranteed to stay in the market for any given (uniform) minimum amount of time, then there is no need to wait with matching decisions. This stands in contrast to the landmark result by Akbarpour et al. (2020) which show that the loss through instantaneous matching decisions is high if the agents' departure follows an exponential distribution. Hence, we find a surprising path to circumvent this bad prediction: only tiny tweaks to the departure times of agents make a huge difference. Why is the good performance possible in our model? The reason is that agents' matching opportunities are sufficiently evenly distributed. Then, for agents in the market, a guaranteed sojourn time means that they cannot be completely neglected within matching decisions when new agents arrive. Are our assumptions realistic? Consider for example matching markets that require a fee to enter. We believe that it is unreasonable that agents would enter such markets without giving the market an opportunity to match them. For instance, when agents look for a ride sharing service in a large market, assuming that agents wait for a short amount of time, say one minute, to be matched is sufficient for the good performance of our algorithm. While Akbarpour et al. (2020) show that the information gain of the exact *realizations* of departures has high value (if the departure times are exponentially distributed), we show that we do not need such information if we can make mild assumptions about the departure behavior of the agents. Hence, we can additionally avoid both cost and ethical concerns of the collection of departure information (Reese et al., 2015). To summarize, our work identifies a new reasonable cause to promote instantaneous matching decisions that goes beyond market thickness.

Apart from avoiding the danger of congestion, instantaneous matching has the advantage that agents do not sojourn in the market unnecessarily. We show that the total waiting time of the agents is the integral of the pool size over time. Hence, we derive a short total waiting time of the agents by proving that the market stays small for most of the time. This stands in contrast to a trade-off between matching quality and waiting time observed in dense markets (Mertikopoulos et al., 2020). In other words, transitioning to the goal of minimizing loss in a *sparse* market can circumvent this conflict of aims.

The rich literature on matching markets comprises a large number of formal models. These differ in their assumptions on the essence of agents, the matching technology, and procedural specifications as well as regarding the measurement of the desirability of matchings. First, agents may be homogeneous or heterogeneous. The latter is often due to having certain applications in mind. For instance, agents can be partitioned into two classes like workers and firms in labor markets, children and schools in school choice, or children and potential adoptive parents in adoption markets. Moreover, agents might assume various additional properties such as being hard or easy to match, which play for instance an important role in kidney exchange (Ashlagi et al. (2019b), Ashlagi et al. (2022), see also Loertscher et al. (2022) for a more general trade market model with superior and inferior buyers and sellers). Second, the matching technology specifies the possible matchings. For instance, it might be prohibited that agents of the same class match (which is very reasonable for the above examples of *bipartite* instances), or the possible matches could be given endogenously by a compatibility relation (which might even be deduced from an exogenous factor like affiliation to some type). Third, procedural specifications concern the chronological process of the model. Agents might arrive stochastically, in fixed time steps, or even according to an adversary. Then, they stay in the market for some time after which they depart according to some procedure. Some stylistic models even assume an indefinite sojourn of agents until they are matched. Finally, we have to measure the quality of

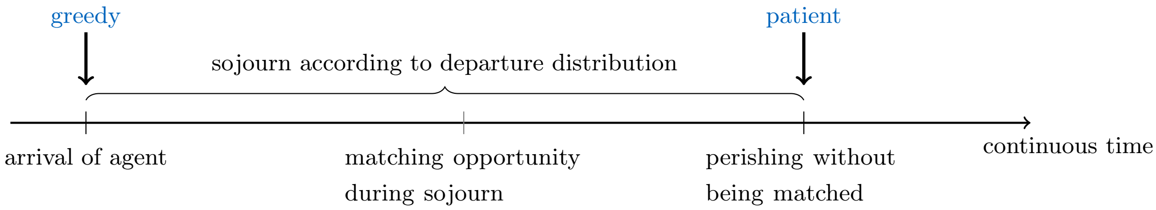


Figure 1: Illustration of the matching process. Agents arrive stochastically according to a Poisson process. During their sojourn in the matching market, they can be matched with contemporaneous agents and leave the market. They perish according to a predefined departure distribution. Specific matching policies are highlighted in blue. The greedy algorithm seeks to match an agent instantaneously at arrival, while the patient algorithm waits for the last moment to match.

matchings. The two most common approaches considered in the literature are measuring the quality of a matching simply by its cardinality or by maximizing a more complex predefined objective function, which is often defined specifically for the given model and the application in mind. The former is appropriate in models where sparse markets evolve from a compatibility relationship of the agents or in settings such as kidney exchange where losing an agent is severe. The latter often requires a careful consideration of model parameters and may quickly lead to complex measures. Objective functions usually include the quality of matches, but can also consider loss suffered by spending time in the market.

In the present paper, we follow the work by Akbarpour et al. (2020) and Anderson et al. (2017) and consider a matching market with ex ante homogeneous agents, that is, agents are a priori identical but, given a snapshot of the current market, they differ in their departure information and potential matches. An overview of our procedural specifications is given in Figure 1. Agents arrive stochastically according to a Poisson process and stay for a maximum time according to a departure distribution. Our focus lies on the departure information of the agents, which we treat as a variable component of our model. Much of our analysis concerns predicting the behavior of the market given certain fairly general assumptions on the departure information. The matching technology is given by a compatibility relationship based on the Erdős-Rényi random graph model (Erdős and Rényi, 1960), and is determined by independent Bernoulli random variables deciding on the compatibility of arriving agents with existing unmatched agents. We express our results with respect to the arrival rate of compatible agents, that is, the expected number of compatible agents arriving in one time unit. Since we consider growing arrival rates leading to decreasing compatibility probabilities, our considered markets are mostly sparse. As a consequence, we measure the performance of an algorithm by the cardinality of the matching it achieves, i.e., we seek to minimize the obtained *loss*, defined as the fraction of agents departing from the market unmatched.

Following the approach and terminology by Akbarpour et al. (2020), we focus on two matching procedures assuming the global perspective of a social planner responsible for a centralized execution of the matchings. The first one, called *greedy algorithm*, corresponds to the instantaneous implementation of matchings, whenever agents arrive at the market. In other words, whenever an agent arrives at the market, she either forms an edge with some existing agent and leaves the market right away as part of a pair, or she joins a pool of agents of which no pair can be matched. Once added to the pool, an agent can solely wait to be matched with a newly arriving agent, or until she perishes.

The second matching procedure, called *patient algorithm*, assumes that the exact departure times of agents are known to the social planner. Hence, once the point of departure arrives, there is one last chance for the agent to be matched before perishing. Therefore, the patient algorithm waits to match an agent until the very last moment and only matches upon the danger of losing an agent. Consequently, the only possibility of loss is if an agent at the edge of departure maintains no relationships with other agents in the pool. The main result by Akbarpour et al. (2020) is that the patient algorithm outperforms the greedy algorithm in case of exponentially distributed departure times.

As discussed before, our main result shows an exponentially small loss of the greedy algorithm under certain assumptions on the distribution of the departure time of agents, while agents simultaneously stay in the market for a short time.

By contrast, the patient algorithm causes agents to wait long, because at least half of the agents have to wait for their maximum possible duration to stay. Still, our analysis of the patient algorithm complements the findings by Akbarpour et al. (2020) and shows an exponentially small loss for a constant departure time. Hence, if exact departure points are known and waiting time is secondary, then the patient algorithm also performs very well.

We complement our theoretical results by a series of simulations. As outlined above, our theoretical findings consider the limit case for growing arrival rates. However, the error terms vanish at rapid rates and the simulations also confirm our results for a rather small arrival rate indicating their robustness. Moreover, we test further specifications of the greedy algorithm. In both of our algorithms, the decision to form a matching is done by selecting a partner uniformly at random among the possible matches. This already yields approximately optimal results as our lower bounds for the performance of any greedy algorithm show. On the other hand, we perform simulations for the case of a tie-breaking in favor of partners that have the smallest remaining sojourn, finding that this still yields a small improvement. Interestingly, our simulations provide evidence that the loss of the greedy mechanism and the patient mechanism are identical under a constant departure time. In a concluding section, we shed light on theoretical reasons for this observation. There, we also discuss extensions of our work to the case of heterogeneous agents.

On the technical side, we develop novel techniques to obtain our results. While most of the existing literature heavily relies on a steady state analysis of the pool size (in particular Akbarpour et al., 2020; Anderson et al., 2017; Ashlagi et al., 2022), this quantity is not Markovian for general departure distributions, and we do not make use of a stationary distribution inherent to our process. Instead, we perform a detour by a careful direct analysis obtaining uniform bounds over time for the pool size which in turn lead to bounds for the loss. Similarly, our analysis of the patient algorithm demands new techniques and our key idea is a case analysis for the near past of a fixed point in time, performing a comparison of the critical case with an urn model.

2 Related Literature

There exist several streams of literature dealing with dynamic models of matching markets. Closest to our work are the articles by Anderson et al. (2017) and Akbarpour et al. (2020). We extend their model and results offering new insights when greedy matching performs well. Anderson et al. (2017) seek to model barter exchange (having in mind the important application of kidney exchange) and the model is essentially a directed version of the model by Akbarpour et al. (2020). Arriving agents form directed relationships/edges with probability p, however they arrive in fixed time intervals for an indefinite sojourn, and the matching technology even allows for matching a larger group of agents or chains originating from altruistic donors. The goal is to minimize total waiting time and the main result is that allowing more matching possibilities significantly reduces waiting time. If the matching technology is restricted to pairs, then their model closely resembles our model, where compatibility is interpreted as having both directed edges present, amounting to a compatibility probability of p^2. The waiting time in their result matches the waiting time that we observe for the greedy algorithm. Therefore, we complement the result by Anderson et al. (2017) for different procedural specifications. Kakimura and Zhu (2021) consider a bipartite version of this model, where the agents of the two partition classes arrive with two different Poisson rates and they also consider the greedy and patient algorithm. Performing a Markov chain analysis, they confirm the superiority of the patient algorithm in a two-sided market, but they observe a similar performance of the two algorithms for a one-sided market where one class of the agents are merely objects.

Much of the literature on dynamic matching markets originates from research in theoretical computer science on the online formation of bipartite matchings (Karp et al., 1990). There, the primary goal is to maximize the cardinality of the matching formed instantaneously, while procedural specifications are secondary and algorithms are measured by their performance against an adversarial arrival of agents. With the purpose of modeling ad allocation, the model was generalized by Mehta et al. (2007), and a version with stochastic arrival of agents was considered subsequently (Feldman et al., 2009; Manshadi et al., 2012). An adversarial arrival of agents is also considered by Emek et al. (2016) who consider an objective function combining the compatibility quality of matched agents and their waiting time, and by Ashlagi et al. (2019a) who consider a model with arrivals and departures after constant time.

The paradigm of waiting to create further possibilities, i.e., the desire of market thickness, observed by Akbarpour et al. (2020) is ubiquitous in the literature. This phenomenon often occurs in markets with heterogeneous types and unbounded sojourn of agents, and we identify two main circumstances for its appearance. The first encompasses markets with the concurrent objectives to maximize matching quality while minimizing waiting time, and the second concerns settings, where agents are offered matching opportunities which they can accept or deny, similar to the secretary problem (Ferguson, 1989; Bearden, 2006).

Regarding the former, Baccara et al. (2020) model child adoption by a bipartite matching market with two agent types and identify thresholds to decide when to conduct low-quality matches. In a similar vein, Loertscher et al. (2022) consider a bipartite trade model with a superior and inferior type of buyers and sellers, and determine how many agents to keep in the market. Interestingly, they find that few stored agents lead to a good market performance, but this result has to be interpreted with respect to their market design which requires to match many agents early. Blanchet et al. (2022) consider a bipartite market with exponentially distributed departure times. Their focus is on the distribution of matching values and they identify certain threshold rules which achieve a desired market thickness. In models with adversarial arrival of agents, Emek et al. (2016), Collina et al. (2020), and Pavone et al. (2020) consider the fine-tuned timing of matching decisions.

In the second type of markets, the challenge is to balance information gains through rejecting matches with the danger of turning down a promising match. There, Leshno (2021) studies a queuing model applied to an assignment problem, and finds that it is beneficial to decline mismatches. This can have both positive aspects for the declining agent who might obtain a better object, and for other agents in the market who might get served earlier. Furthermore, a recent paper by Agarwal et al. (2021) considers a matching market with arrivals and departures in discrete time steps and they investigate cutoff strategies for accepting a match. Loosely related, Shimer and Smith (2001) consider a model for labor markets where heterogeneous agents incur a cost for searching a suitable match. They find that agents of high productivity should wait to improve their matching quality. Also, note that the matching problem in the above paper by Pavone et al. (2020) is equivalent to an adversarial version of the secretary problem.

There is far less evidence for the antagonistic principle of taking instantaneous decisions, with the notable exception of the research on kidney exchange, originating from seminal contributions by Roth et al. (2004, 2005). In this line of research, agents usually stay indefinitely in the market and the performance is measured by means of the total waiting time. Ünver (2010) anticipates the conclusion by Anderson et al. (2017), namely the optimality of greedy matching. In his model, the role model for Anderson et al. (2017), he considers arrivals by a Poisson process and an indefinite sojourn, but the model contains agents of homogeneous types and the waiting time objective is specifically designed for the given setting. Follow-up work by Ashlagi et al. (2019b) offers a closer look at the case of easy-to-match and hard-to-match agents and their prioritization in greedy-type algorithms. These theoretical findings are in accordance with simulations based on real-world data (Ashlagi et al., 2018). Ashlagi et al. (2022) show the optimality of greedy matching in a similar setting. All of this work has in common that the reason for the optimality of greedy matching is market *thickness*. By contrast, our work presents the optimality of an instantaneous algorithm measured by the *loss* of agents in a sparse market that is *thin* (for most of the time). Very recently, Kerimov et al. (2022) consider a model with heterogeneous types whose compatibility is globally fixed. They show that, under certain conditions on the compatibility structure, greedy matching is optimal with respect to regret.

Finally, Aveklouris et al. (2021) consider a bipartite model with heterogeneous agents where the objective function is a tradeoff of matching value and waiting costs. Since the model is quite complex model, a Markovian analysis is not feasible. They circumvent this obstacle by finding optimal matching rates by solving an optimization problem, and identify an algorithm that mimics these rates. Interestingly, this paper is the only related paper we are aware of which investigates the influence of the departure distribution.

3 Model

Following the models by Anderson et al. (2017) and Akbarpour et al. (2020), we consider a continuous-time matching market in the time window $[0,T]$ for some maximum time $T > 0$. Agents arrive with Poisson rate m and enter the market. We label the agents $v_1, v_2, \ldots$ in the order of their arrival. Given times $s, t \in [0,T]$ with $s < t$, we denote by A_t the set of *agents arriving at time* t and by $A_{[s,t]}$ and $a_{[s,t]}$ the set and number of agents arriving in the time interval $[s,t]$, respectively. At her arrival, an agent enters the *pool*, i.e., the set of agents in the market. The pool at time t is a random variable Z_t, and we denote its size by z_t, i.e., $z_t = |Z_t|$. We assume that the market is initially empty, i.e., $Z_0 = \emptyset$.

Once an agent enters the pool, we decide on her *compatibility* with agents already in the pool by independent Bernoulli random variables with parameter p. Hence, compatible agents arrive at Poisson rate $d = p \cdot m$, and we call this quantity the *density parameter*. Formally, we model compatibility as follows. We assume that agent v_i carries an independent Bernoulli $(p)^{\otimes \mathbb{N}}$ distributed random variable $U(v_i)$. There, $\otimes \mathbb{N}$ denotes the product space. Let the pool before the arrival of v_i consist of the agents $\{v_{l_1}, \ldots, v_{l_K}\}$, where K is the size of the pool just before the arrival of v_i, and $1 \leq l_1 < l_2 < \ldots < l_K < i$ are all integer numbers. At arrival, the agent v_i forms an edge with v_{l_j} if and only if $U_j(v_i) = 1$, where $U_j(v_i)$ is the j-th component of $U(v_i)$. So in particular, only the first K many components of $U(v_i)$ are of interest.

While an agent is in the pool, she can be matched with a compatible agent in the pool, and they leave the pool together. Otherwise, the agent leaves the pool after some independent random *maximum sojourn time* X_i has passed, which is distributed according to some measure μ with support contained in $[0, +\infty]$. The maximum sojourn time is also called *departure time* (but some caution is needed here, as an agent can also 'depart' from the market due to being matched). The last point, where an agent joining at time t can be matched is at time $t + X_i$, and we call an agent *critical* at the moment when she reached her last time to be matched. The time an agent spends in the pool is called her *sojourn time* and we say that an agent *perishes* if she leaves the pool unmatched.

Agents are matched by certain *matching algorithms* (or *matching policies*), which decide when two agents present in the pool leave the market as a pair. In principle, multiple pairs could be matched at the same time (as it is often done in batching algorithms, for instance for kidney exchange), and no agent can be part of more than one pair. Two important matching policies are the greedy algorithm and patient algorithm (Akbarpour et al., 2020). The *greedy algorithm* GDY is defined by the policy that every agent arriving at the pool is matched to a compatible agent in the pool uniformly at random and instantaneously, or joins the pool if there is no compatible agent. On the other hand, the *patient algorithm* PAT is defined by having agents wait until they reach their maximum sojourn time at which they match with a compatible agent uniformly at random, or perish if no such agent exists. More formally, if agent v_i arrives at time t, then she waits until time $t + X_i$, and may only leave the market during this time when she is matched with a compatible critical agent selecting her as a match. If this does not happen, she checks the pool Z_{t+X_i} for compatible agents once she is about to perish, and selects a match if possible. Note that, for every time t, it holds almost surely that at most one agent arrives or perishes at this time, and we can therefore disregard a case analysis for the possibility of multiple simultaneous matching events.

Next, we want to define the objective of minimizing the loss of agents. Therefore, given an algorithm ALG and an arrival rate m, we consider the number of agents matched until time T, as defined by $\mathrm{ALG}(m,T) = \{v_i \colon v_i \text{ matched until time } T\}$. Then, the loss of ALG until time T is defined by

$$\mathbf{L}_{\mathrm{ALG}}(m,T) = \frac{\mathbb{E}[|A_{[0,T]} - \mathrm{ALG}(m,T) - Z_T|]}{mT}.$$

In this expression, the denominator is equal to the expected number of agents arriving in the time window $[0,T]$. Note that the loss of an algorithm also depends on the density parameter d and the distribution μ of the departure times. For the sake of notational simplicity, we exclude these from our notation, but add the departure distribution as a subscript of probabilities or expectations in danger of ambiguity. We are interested in the limiting behavior for m and T tending to infinity. Therefore, we define

$$\mathbf{L}_{\mathrm{ALG}} = \limsup_{m,T\to\infty} \mathbf{L}_{\mathrm{ALG}}(m,T).$$

Note that this limiting behavior is mostly cosmetic. Error terms depending on m and T decay quickly in our theoretical analysis, and our simulations show the robustness of our results for reasonably small arrival rates.

4 Analysis of the Greedy Algorithm

In this section, we perform our analysis of the greedy algorithm. We defer some technical proofs to the appendix.

4.1 Outline of the Analysis

The main problem in the analysis of any other model than the one with exponentially distributed departure times is the lack of Markovianity of the pool size. Knowing at which exact time points the agents in the current pool arrived, already reveals some information at which time points in the future we expect more or less agents to get critical and perish. So the knowledge of the entire history $(z_t)_{t\leq t_0}$ typically reveals more information than the pool size z_{t_0}, which shows that $(z_t)_{t\geq 0}$ is not a Markov chain in general. Hence, we cannot simply analyze a stationary distribution in order to control the pool size, but have to perform a direct analysis of the pool size to be able to investigate the loss of the greedy algorithm.

Determining an upper bound for the pool size follows from the following idea: Assume that the pool has a size of Cm for some constant $C > \frac{\ln(2)}{d}$ and an agent arrives. Then, there is a probability of

$$\left(1 - \frac{d}{m}\right)^{Cm} \overset{\text{Lemma C.1}}{\leq} e^{-Cd} < \frac{1}{2}$$

that the new agent joins the pool causing the pool size to increase by 1. On the other hand, there is a probability of strictly more than $\frac{1}{2}$ that the agent matches and thus the pool size decreases by 1. Consequently, for a sufficiently large pool size, the arrival of new agents decreases the pool size in average. Besides the arrival of new agents, existing agents in the pool may also perish and cause a further decrease of the pool size. This analysis shows that we do not expect the pool size to be much larger than $\frac{\ln(2)}{d}m$ for most of the time. This key step is shown in Corollary 4.3 and Lemma 4.4. An important technique for the former result is to compare an arbitrary departure time with the case when agents do not perish at all, i.e., with the case of an infinite departure time. For this, we will create a coupling of the respective markets.

Given a uniform bound on the pool size, we can also bound the loss under various assumptions on the departure time. For instance, assume a constant departure after 1 time unit. Further, suppose that a new agent, called v, enters the pool at time t without being matched right away and that the pool size is bounded from above by $\frac{\ln(2)}{d}m$ for all times in $[t, t+1]$. This agent will stay for (at most) one time unit and each agent that arrives during this time has a probability of approximately (up to constant multipliers) $\frac{d}{m}$ to match with v. Hence, the probability that v perishes without being matched by another arriving agent is bounded by an expression of the form $\left(1 - \Theta\left(\frac{d}{m}\right)\right)^m$ which explains the exponentially small loss under a constant departure time. By generalizing these arguments, we provide lower and upper bounds on the loss for different distributions of departure times. We find quite general conditions that distinguish cases when we are guaranteed either an exponentially small or an inverse linear loss.

4.2 Comparison of Departure Distributions

In this section, we compare the pool sizes under the Greedy algorithm for different distributions of departure times and under different initial conditions.

In the first lemma, we use a coupling method to compare the evolution of the pool size under any departure time with the case of an infinite departure time. Therefore, let μ_∞ be the degenerate measure with $\mu_\infty(\{+\infty\}) = 1$. Given a time t, denote by Z_t^∞ and z_t^∞ the pool and the pool size for departure times according to μ_∞, respectively. Note that our coupling lemma leaves a lot of freedom for the starting pool, and does for instance not require empty pools in the beginning.

Lemma 4.1. *Let μ be any probability measure on $[0, +\infty]$. Consider the pools Z_t and Z_t^∞ at time t for the greedy algorithm with departure times distributed according to μ and μ_∞, respectively, and let Z_0 and Z_0^∞ be two starting pools with $z_0 \le z_0^\infty$. Then, there exists a coupling of Z_t and Z_t^∞ such that*

$$z_t \le z_t^\infty + 1 \tag{1}$$

for all $t \ge 0$.

Proof. We consider the procedure where we have two pools to which two identical copies of agents arrive according to Poisson rate m. The first pool acts according to the greedy algorithm under measure μ and initial pool Z_0 and the second one according to the greedy algorithm under measure μ_∞ and initial pool Z_0^∞. Recall that, attached to each arriving agent v, there is a random vector $U(v) \in \{0,1\}^{\mathbb{N}}$ storing the random compatibility with other agents, as described in Section 3. In particular, if the pool is of size K and an agent v arrives, then the agent directly forms an edge if and only if $U_j(v) = 1$ for some $j \in \{1, \ldots, K\}$. If there are compatible agents in the pool, then the agent v selects a partner among them chosen uniformly at random.

We show the assertion (1) by a case distinction. The statement is clear for the case $t = 0$. For $t > 0$, the only interesting times are the ones where agents arrive, as agents can only perish from the first pool. Hence, z_t^∞ cannot get smaller without the arrival of a new agent.

If $z_t < z_t^\infty$ and an agent arrives at time t, then z_t can increase by at most 1 and z_t^∞ can decrease by at most 1, so (1) is still satisfied after the arrival of the agent. If $z_t = z_t^\infty$ and an agent arrives at time t, then the agent matches under the measure μ if and only if she matches under the measure μ_∞, so (1) is also satisfied after the arrival of the agent. If $z_t = z_t^\infty + 1$

and an agent arrives, then if the agent does not match with an element in the first pool, she also does not match with an element in the second pool, so both pools increase in size. So it is not possible that z_t gets larger while z_t^∞ gets smaller at the same time which shows that (1) is always satisfied. $\square$

In particular, Lemma 4.1 implies that when running the greedy algorithm without perishing (infinite departure time) and one starts with two different initial pool sizes, the process that started with the lower pool size, can surpass the other one by at most 1.

A key feature of the measure μ_∞ is that the pool size actually is Markovian under this measure, as the evolution of the pool size depends only on the current size of the pool and the arriving agents. The transition rates $(r_{k\to k+1})_{k\in\mathbb{N}_{\geq 0}}$ and $(r_{k\to k-1})_{k\in\mathbb{N}_{\geq 1}}$ are given by

$$r_{k\to k+1} = m\left(1-\frac{d}{m}\right)^k, \; r_{k\to k-1} = m\left[1-\left(1-\frac{d}{m}\right)^k\right]. \tag{2}$$

Interestingly, we have that $r_{0\to 1} = m$ and, for all $k \geq 1$, that $r_{k\to k+1} + r_{k\to k-1} = m$. Thus, if the pool size is larger than 0, then the process waits an exponential time with expectation $1/m$ and then jumps to $k+1$ or $k-1$ with corresponding probabilities. At 0, the process also waits for an exponential time with expectation $1/m$ and then jumps $+1$. In particular, the transition probabilities $(p(k,l))_{k,l\in\mathbb{N}_{\geq 0}}$ of the underlying discrete-time Markov chain are simply given by

$$p(k,k+1) = \left(1-\frac{d}{m}\right)^k \text{ and } p(k,k-1) = 1-\left(1-\frac{d}{m}\right)^k \tag{3}$$

for $k \in \mathbb{N}_{\geq 1}$, $p(0,1) = 1$, and all other probabilities are 0. These transition probabilities give rise to an irreducible and reversible Markov chain (every Markov chain with state space $\mathbb{N}_{\geq 0}$ making jumps between nearest neighbors only is reversible). We claim that there exists a stationary distribution ρ for this Markov chain. Indeed, one can construct a stationary measure with the aid of the measure $\tilde\rho$ defined by $\tilde\rho(0) = 1$ and $\tilde\rho(k+1) = \rho(k)\frac{p(k,k+1)}{p(k+1,k)} = \prod_{i=0}^k \frac{p(i,i+1)}{p(i+1,i)}$. This is *not* a probability measure, yet. However, the measure $\tilde\rho$ can be normalized as $\sum_{k=0}^\infty \prod_{i=0}^k \frac{p(i,i+1)}{p(i+1,i)} < \infty$, which follows because $p(i,i+1) < 1/3$ for all large enough i.

The existence of a stationary distribution tells us that the pool size does not go off to infinity as time grows. The balance equation of the stationary distribution for $j \in \mathbb{N}_{\geq 0}$ reads

$$\rho(j)p(j,j+1) = \rho(j+1)p(j+1,j), \tag{4}$$

and we will give bounds on the stationary distribution in Lemma 4.2 below.

4.3 Pool Size Bounds for the Greedy Algorithm

In this section, we provide bounds for the pool size of the greedy algorithm. The key idea is to distinguish different ranges of the pool size constraint by carefully chosen constants. For this, we define the constants $C_1 = C_1(m)$, $C_2 = C_2(m)$, and $C_3 = C_3(m)$ by

$$C_1 = 1 + \frac{10}{\ln(2)\ln(m)}, \; C_2 = C_1 + 2\frac{d\ln(m)^2}{m\ln(2)}, \text{ and } C_3 = C_1 + 4\frac{d\ln(m)^2}{m\ln(2)}.$$

First, we show how to control the stationary measure under infinite departure times. As a consequence, we obtain bounds for the pool size under an arbitrary departure distribution.

Lemma 4.2. *Let ρ be the stationary measure of the pool size without perishing. Then,*

$$\rho\left(\left(\frac{C_1\ln(2)m}{d} + \frac{3}{2}\ln(m)^2, \infty\right) \cap \mathbb{N}\right) \leq m^{-9}$$

for m large enough.

Proof. Let $j \geq \frac{C_1 \ln(2) m}{d}$. Then, we have

$$p(j, j+1) = \left(1 - \frac{d}{m}\right)^j \leq \left(1 - \frac{d}{m}\right)^{\frac{C_1 \ln(2) m}{d}} \leq \frac{1}{2} e^{-\frac{10}{\ln(m)}}.$$

On the other hand, we also have that $p(j, j-1) \geq 1/2$. Inserting this into (4), we get

$$\rho(j+1) = \rho(j) \frac{p(j, j+1)}{p(j+1, j)} \leq e^{-\frac{10}{\ln(m)}} \rho(j).$$

Thus, we get inductively

$$\rho\left(\lceil \frac{C_1 \ln(2) m}{d} \rceil + n\right) \leq e^{-n \frac{10}{\ln(m)}} \rho\left(\lceil \frac{C_1 \ln(2) m}{d} \rceil\right) \leq e^{-n \frac{10}{\ln(m)}}.$$

Summing this over different values of n, we get for m large enough

$$\begin{aligned} \sum_{k > \frac{C_1 \ln(2) m}{d} + \frac{3}{2} \ln(m)^2} \rho(k) &\leq \sum_{k = \lceil \ln(m)^2 \rceil}^{\infty} \rho\left(\lceil \frac{C_1 \ln(2) m}{d} \rceil + k\right) \leq \sum_{k = \lceil \ln(m)^2 \rceil}^{\infty} e^{-k \frac{10}{\ln(m)}} \\ &= e^{-\lceil \ln(m)^2 \rceil \frac{10}{\ln(m)}} \sum_{k=0}^{\infty} e^{-k \frac{10}{\ln(m)}} \leq m^{-10} \frac{1}{1 - e^{-\frac{10}{\ln(m)}}} \leq m^{-10} \frac{\ln(m)}{5} \leq m^{-9}. \end{aligned}$$

There, we used that $1 - e^{-\frac{10}{\ln(m)}} \geq \frac{5}{\ln(m)}$ for m large enough in the second to last step. □

Combining the bounds under infinite departures with our coupling result, we can bound the pool size of the greedy algorithm under an arbitrary departure time.

Corollary 4.3. *Assume that the departure time is distributed according to an arbitrary probability measure μ. Then, it holds for all times t and arrival rates m large enough that*

$$\mathbb{P}_\mu\left(z_t > \frac{C_2 \ln(2)}{d} m\right) \leq m^{-9}.$$

Proof. Assume that Z_0 is the empty pool and Z_0^∞ has size distributed according to the stationary distribution ρ. Then $z_0 \leq z_0^\infty$, so we can couple the processes according to Lemma 4.1 such that $z_t \leq z_t^\infty + 1$. The pool size z_t^∞ is still distributed according to ρ by stationarity, so we get that for large enough m

$$\mathbb{P}_\mu\left(z_t > \frac{C_2 \ln(2)}{d} m\right) \leq \mathbb{P}_{\mu_\infty}\left(z_t^\infty > \frac{C_1 \ln(2)}{d} m + \frac{3}{2} \ln(m)^2\right) \leq m^{-9}.$$

There, the first inequality follows from Lemma 4.1 and the second inequality from Lemma 4.2. □

Our second key insight is that if the pool size is sufficiently small, then the probability of reaching a much larger pool size within one time unit is very low.

Lemma 4.4. *Assume that the departure time is distributed according to an arbitrary probability measure μ. Further, assume that Z is a pool of size at most $C_2 \frac{\ln(2)}{d} m$. Then, for large enough m, it holds that*

$$\mathbb{P}_\mu\left(z_t \leq C_3 \frac{\ln(2) m}{d} \text{ for all } t \leq 1 \;\middle|\; Z_0 = Z\right) \geq 1 - m^{-9}.$$

Proof. We seek an upper bound for the probability that the pool size reaches $C_3\frac{\ln(2)}{d}m$ when starting at a size of at most $C_2\frac{\ln(2)}{d}m$. Let $M = \lceil C_2\frac{\ln(2)}{d}m\rceil$. When the pool has a size of at least M and an agent arrives, then the probability that the agent joins the pool, i.e., that she is not matched directly, equals

$$\left(1-\frac{d}{m}\right)^M \leq \left(1-\frac{d}{m}\right)^{C_2\frac{\ln(2)}{d}m} \leq e^{-C_2\ln(2)} \leq \frac{1}{2}e^{-\frac{10}{\ln(m)}}. \tag{5}$$

There, we used Lemma C.1 in the second inequality. Whenever a new agent arrives, the pool size can increase or decrease by 1, where we have a tendency towards a decrease of the pool size. Thus, in particular, if the pool starts at size M, the probability that the pool size increases beyond $C_3\frac{\ln(2)}{d}m$ before decreasing to $C_2\frac{\ln(2)}{d}m$ is bounded by the probability of this event for the case of a random walk with drift to the left of at least $\frac{1}{2}e^{-\frac{10}{\ln(m)}}$. The euclidean distance between $C_2\frac{\ln(2)}{d}m$ and $C_3\frac{\ln(2)}{d}m$ is $2\ln(m)^2$. Hence, there are more than $\ln(m)^2$ integer points between these values, for $m \geq 3$. We can apply Lemma C.3 with $\frac{1}{2} - \varepsilon = \frac{1}{2}e^{-\frac{10}{\ln(m)}}$ in order to bound the probability that the Markov chain with drift crosses such an interval. Thus, the probability of this event is bounded by

$$\left(2\frac{1}{2}e^{-\frac{10}{\ln(m)}}\right)^{\ln(m)^2} = m^{-10}.$$

If the pool size starts at some value less than $\frac{C_2\ln(2)}{d}m$ and then grows to some value greater than $\frac{C_3\ln(2)}{d}m$ then at some time interval in between the pool size starts with size M and reaches size $\frac{C_3\ln(2)}{d}m$ before going back to $M-1$.

Assume that k agents arrive during the time interval $[0,1]$, i.e., $a_{[0,1]} = k$. For $i \in \{1,\ldots,k\}$, the probability that the pool size is M when the i-th agent arrives and then increases beyond $\frac{C_3\ln(2)}{d}m$ before going to $M-1$ is bounded by m^{-10} by the previous calculations. Thus, by a union bound, the probability that the pool size increases beyond $\frac{C_3\ln(2)}{d}m$ is bounded by km^{-10}. Conditioning on the number of arrivals in the time interval $[0,1]$, we get

$$\begin{aligned}
&\mathbb{P}_\mu\left(z_t > \frac{C_3\ln(2)m}{d} \text{ for some } t \leq 1 \;\middle|\; Z_0 = Z\right) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_\mu\left(z_t > \frac{C_3\ln(2)m}{d} \text{ for some } t \leq 1 \;\middle|\; Z_0 = Z, a_{[0,1]=k}\right)\mathbb{P}\left(a_{[0,1]} = k\right) \\
&\leq \sum_{k=1}^{\infty} km^{-10}\mathbb{P}\left(a_{[0,1]} = k\right) = m^{-9}
\end{aligned}$$

as $a_{[0,1]}$ is a Poisson random variable with expectation value m. □

4.4 Loss Bounds for the Greedy Algorithm

Before we leverage the bounds on the pool size to obtain bounds on the loss, we provide a useful lemma that lets us express the loss as an integral over the probability of perishing. This lemma holds for arbitrary matching algorithms and will be applied again for the patient algorithm. Its proof follows by an application of Mecke's equation.

Lemma 4.5. *Let a matching algorithm ALG be given and assume that departure time is distributed according to an arbitrary probability μ. Then,*

$$\mathbf{L}_{ALG}(m,T) = \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\textit{an agent arriving at time } t \textit{ perishes before time } T\right)dt,$$

and

$$\mathbf{L}_{ALG} = \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\textit{an agent arriving at time } t \textit{ perishes}\right) dt.$$

Finally, we can apply our insights to bound the loss of the greedy algorithm. We start with an assumption on the measure that particularly covers the case of constant unit waiting times. We want to remark that all obtained loss bounds are also true for ex ante heterogeneous agents which have departure times with respect to different probability measures, all satisfying the respective assumptions.

Theorem 4.6. *Assume that the departure time is distributed according to a probability measure μ with $\mu([0,1)) = 0$. For $d \geq 2$, it holds that*

$$\mathbf{L}_{GDY} \leq e^{-\frac{d}{2\ln(2)}}.$$

Proof. By Lemma 4.5, we have

$$\mathbf{L}_{\text{GDY}} = \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes}\right) dt. \tag{6}$$

Now we define a certain good event $\mathcal{G}_t$ on which the probability that an agent perishes is relatively low. Let

$$\mathcal{G}_t = \left\{z_t \leq \frac{C_2\ln(2)}{d}m\right\} \cap \left\{z_s \leq \frac{C_3\ln(2)}{d}m \forall s\in[t,t+1]\right\} \cap \left\{a_{[t,t+1]} \geq m - 30\ln(m)m^{1/2}\right\}.$$

We first want to bound the probability of the complement $\mathcal{G}_t^c$. By a union bound, we get

$$\begin{aligned}\mathbb{P}_\mu\left(\mathcal{G}_t^c\right) \leq & \mathbb{P}_\mu\left(z_t > \frac{C_2\ln(2)}{d}m\right) + \mathbb{P}_\mu\left(a_{[t,t+1]} < m - 30\ln(m)m^{1/2}\right) \\ & + \mathbb{P}_\mu\left(\left\{z_s \leq \frac{C_3\ln(2)}{d}m \text{ for all } s\in[t,t+1]\right\}^c \cap \left\{z_t \leq \frac{C_2\ln(2)}{d}m\right\}\right).\end{aligned}$$

The probability of the first summand is bounded by m^{-9} by Corollary 4.3. For the second summand, note that $a_{[t,t+1]}$ is a Poisson random variable with parameter m and thus we can use Lemma C.2 obtaining

$$\mathbb{P}_\mu\left(a_{[t,t+1]} < m - 30\ln(m)m^{1/2}\right) = \mathbb{P}_\mu\left(a_{[t,t+1]} < m\left(1 - \frac{30\ln(m)}{m^{1/2}}\right)\right) \leq e^{-\frac{m}{3}\left(\frac{30\ln(m)}{m^{1/2}}\right)^2} \leq m^{-10}.$$

There, the last two inequalities hold for m large enough. The last summand is bounded by

$$\mathbb{P}_\mu\left(z_s > \frac{C_3\ln(2)}{d}m \text{ for some } s\in[t,t+1] \;\middle|\; z_t \leq \frac{C_2\ln(2)}{d}m\right) \leq m^{-9}$$

by Lemma 4.4. Together, this shows that $\mathbb{P}_\mu\left(\mathcal{G}_t^c\right) \leq m^{-8}$ for large enough m.

Now assume that the event $\mathcal{G}_t$ holds and an agent a arrives at time t. We would like to derive the probability that a does neither get matched directly nor between time t and $t+1$. Whenever a new agent v arrives at time $s \in (t, t+1)$ and at a pool of size z_s, then the probability that v matches with a equals

$$\frac{1}{z_s}\left[1 - \left(1 - \frac{d}{m}\right)^{z_s}\right].$$

Indeed, there is a probability of $1 - \left(1-\frac{d}{m}\right)^{z_s}$ that the agent a matches at arrival, and all vertices in the pool have the same probability of getting matched. The function $x \mapsto \frac{1}{x}\left[1-\left(1-\frac{d}{m}\right)^x\right]$

is decreasing for $x > 0$ which can be seen as follows. Write $b = 1 - \frac{d}{m}$. The derivative of this function is given by $\frac{b^x - xb^x \ln(b) - 1}{x^2}$. As $1 - x\ln(b) \leq e^{-x\ln(b)} = b^{-x}$ we already have that

$$\frac{b^x - xb^x \ln(b) - 1}{x^2} = \frac{b^x}{x^2}\left(1 - x\ln(b) - b^{-x}\right) \leq 0$$

which shows that the function is decreasing. So on the event $\mathcal{G}_t$ we have

$$\frac{1}{z_s}\left[1 - \left(1 - \frac{d}{m}\right)^{z_s}\right] \geq \frac{d}{C_3 \ln(2)m}\left[1 - \left(1 - \frac{d}{m}\right)^{\frac{C_3 \ln(2)m}{d}}\right]$$

for all $s \in [t, t+1]$. As $C_3 > 1$ we have that

$$\left(1 - \frac{d}{m}\right)^{\frac{C_3 \ln(2)m}{d}} \leq \left(1 - \frac{d}{m}\right)^{\frac{\ln(2)m}{d}} \overset{\text{Lemma C.1}}{\leq} \frac{1}{2}$$

and thus the probability that the agent v matches with a is at least

$$\frac{1}{z_s}\left[1 - \left(1 - \frac{d}{m}\right)^{z_s}\right] \geq \frac{d}{C_3 \ln(2)m}\left[1 - \left(1 - \frac{d}{m}\right)^{\frac{C_3 \ln(2)m}{d}}\right] \geq \frac{d}{2C_3 \ln(2)m}.$$

Using $a_{[t,t+1]} \geq m - 30\ln(m)m^{1/2}$ and the assumption on the support of μ, implying that an agent can only perish after a sojourn of at least one time unit, we can further bound the probability that the agent a perishes from above by

$$\left(1 - \frac{d}{2C_3 \ln(2)m}\right)^{m - 30\ln(m)m^{1/2}} \underset{m\to\infty}{\longrightarrow} e^{-\frac{d}{2\ln(2)}}.$$

Consider the event $\mathcal{A}_t = \{\text{an agent arriving at time } t \text{ perishes}\}$. To extend the former inequality without the restriction of $\mathcal{G}_t$, we compute

$$\begin{aligned}
\mathbb{P}_\mu(\mathcal{A}_t) &= \mathbb{P}_\mu(\mathcal{A}_t \mid \mathcal{G}_t)\,\mathbb{P}_\mu(\mathcal{G}_t) + \mathbb{P}_\mu(\mathcal{A}_t \mid \mathcal{G}_t^c)\,\mathbb{P}_\mu(\mathcal{G}_t^c) \\
&\leq \mathbb{P}_\mu(\mathcal{A}_t \mid \mathcal{G}_t) + \mathbb{P}_\mu(\mathcal{G}_t^c) \leq \mathbb{P}_\mu(\mathcal{A}_t \mid \mathcal{G}_t) + m^{-8} \underset{m\to\infty}{\longrightarrow} e^{-\frac{d}{2\ln(2)}}.
\end{aligned}$$

Inserting this into (6) finishes the proof. □

Interestingly, the assumption in the previous theorem on a minimum departure time of one time unit is only cosmetic. Indeed, for obtaining an exponentially small loss in d, we merely have to assume a uniform lower bound on the minimum departure time of an agent. However, transitioning to more general departure times also affects the constant appearing in the exponent of the loss. In the next lemma, we show some rather general scaling invariance for the connection between loss and distribution of the departure times. To state the result, we need to express the loss with respect to parameters that we usually suppress. Therefore, given a probability measure ρ and a density parameter κ, let $\mathbf{L}_{\mathrm{ALG}}(\rho, \kappa)$ denote the loss of ALG with respect to departure times distributed according to ρ and density parameter κ.

Lemma 4.7. *Let μ be a probability measure on $[0, +\infty]$, let $c > 0$, and let ν be the probability measure defined by $\nu([a, b]) = \mu([ca, cb])$ for all $a, b \in [0, +\infty]$. Then,*

$$\mathbf{L}_{GDY}(\nu, d) = \mathbf{L}_{GDY}(\mu, cd).$$

Combining Theorem 4.6 and Lemma 4.7, we obtain the following corollary.

Corollary 4.8. *Let $\varepsilon > 0$ and assume that the departure time is distributed according to a probability measure μ with $\mu([0,\varepsilon)) = 0$. For $d \geq 2$, we have*

$$\mathbf{L}_{GDY} \leq e^{-\frac{\varepsilon d}{2\ln(2)}}.$$

Our next goal is to provide a similar lower bound on the greedy algorithm that shows that our analysis is optimal up to a small constant factor in the exponent.

Lemma 4.9. *Let $\varepsilon, \delta > 0$ and assume that the departure time is distributed according to a probability measure μ with $\mu([0,\varepsilon]) \geq \delta$. Then,*

$$\mathbf{L}_{GDY} \geq \frac{\delta}{2} e^{-\varepsilon d}.$$

We want to use Lemma 4.9 to give two direct corollaries. The first one deals with the case of constant departure time, and we restrict attention to the case $\mu(\{1\}) = 1$. The second one considers probability measures with non-negligible probability for a quick departure. This also generalizes the result by Akbarpour et al. (2020) for exponentially distributed departure times.

Corollary 4.10. *Assume that the departure time is given by the degenerate measure μ with $\mu(\{1\}) = 1$. Then,*

$$\mathbf{L}_{GDY} \geq \frac{1}{2} e^{-d}.$$

The second corollary follows by taking d with $\frac{1}{d} \leq \varepsilon_0$ and applying Lemma 4.9 for $\varepsilon = \frac{1}{d}$.

Corollary 4.11. *Let $c > 0$, $\varepsilon_0 > 0$, and assume that the departure time is distributed according to a probability measure μ with $\mu([0,\varepsilon]) \geq c\varepsilon$ for all $\varepsilon \leq \varepsilon_0$. Then, for d large enough, it holds that*

$$\mathbf{L}_{GDY} \geq \frac{c}{6d}.$$

4.5 Waiting Times under the Greedy Algorithm

In this section, we will consider the waiting time of agents, i.e., the time that agents spend in the pool. Therefore, we consider the process in a fixed time window $[0, T]$. Let W denote the *total waiting time* of all agents up to time T, i.e., the sum over all agents of the time spent in the pool until time T.

For the patient algorithm, at least half of the agents stay for their maximum possible sojourn. Hence, under unit waiting times and for arrivals until time $T-1$, the expected waiting time is at least $1/2$, a fixed constant. The situation is different for the greedy algorithm. Approximately half of the agents get matched at arrival, but the waiting time of agents added to the pool is unclear. We will show that, in the greedy algorithm under an arbitrary departure time, $\mathbb{E}[W] \in \Theta\left(\frac{mT}{d}\right)$, i.e., the average waiting time is of order $\Theta\left(\frac{1}{d}\right)$.

Interestingly, the expected total waiting time can be expressed by an integral over the expected pool size.

Proposition 4.12. *Assume that the departure time is distributed according to an arbitrary probability measure μ. Then,*

$$\mathbb{E}_\mu[W] = \int_0^T \mathbb{E}_\mu[z_s]\, ds.$$

Proof. Let V be the set of agents that arrive up to time T. By Fubini's theorem, we have

$$W = \sum_{v \in V} \int_0^T \mathbf{1}_{v \in Z_s} ds = \int_0^T \sum_{v \in V} \mathbf{1}_{v \in Z_s} ds = \int_0^T z_s ds.$$

There, $\mathbf{1}$ denotes the indicator function. Another application of Fubini's theorem yields that

$$\mathbb{E}_\mu [W] = \int_0^T \mathbb{E}_\mu [z_s] \, ds.$$

□

We can use this relationship to give an upper bound on the total waiting time.

Proposition 4.13. *Assume that the departure time is distributed according to an arbitrary probability measure μ. Then, for large enough m, under the greedy algorithm, it holds that*

$$\mathbb{E}_\mu [W] \leq \frac{6mT}{5d}.$$

Proof. Using Proposition 4.12, we would like to get an upper bound on $\mathbb{E}_\mu [z_s]$. The key idea is to consider the Markov chain $(Z_t^\infty)_{t \geq 0}$ with infinite departure time started at stationarity and apply Lemma 4.1. By Lemma C.1, we have $p(k, k+1) = \left(1 - \frac{d}{m}\right)^k \leq 1/3$ for $k \geq \frac{\ln(3)m}{d}$ for the transition probabilities with respect to μ_∞. Thus, the stationary distribution satisfies $\rho(k+1) = \frac{p(k,k+1)}{p(k+1,k)} \rho(k) \leq \frac{1}{2} \rho(k)$ for these k. However, this already implies that when z_0^∞ is distributed according to the stationary distribution, then

$$\begin{aligned}
\mathbb{E}_{\mu_\infty} [z_0^\infty] &= \sum_{k=1}^{\infty} \rho(k) k = \sum_{k=1}^{\lceil \frac{\ln(3)m}{d} \rceil + 10} \rho(k) k + \sum_{\lceil \frac{\ln(3)m}{d} \rceil + 11}^{\infty} \rho(k) k \\
&\leq \lceil \frac{\ln(3)m}{d} \rceil + 10 + \sum_{k=11}^{\infty} 2^{-k} \left(\lceil \frac{\ln(3)m}{d} \rceil + k \right) \\
&\leq \lceil \frac{\ln(3)m}{d} \rceil + 10 + 2^{-10} \lceil \frac{\ln(3)m}{d} \rceil + 1 \leq 1.01 \lceil \frac{\ln(3)m}{d} \rceil + 11.
\end{aligned}$$

Thus, we can apply Lemma 4.1 to obtain

$$\mathbb{E}_\mu [W] = \int_0^T \mathbb{E}_\mu [z_s] \, ds \leq \int_0^T \mathbb{E}_{\mu_\infty} [z_s^\infty + 1] \, ds \leq T \left(1.01 \lceil \frac{\ln(3)m}{d} \rceil + 12 \right) \leq \frac{6mT}{5d}$$

where the last inequality holds for m large enough. □

We also prove a lower bound on the total waiting time. However, this lower bound needs minor additional assumptions on the measure of the departure times. Its proof uses a similar technique as the proof of Lemma 4.5.

Proposition 4.14. *Let $c > 0$ and assume that the departure time is distributed according to a probability measure μ with $\mu([c, \infty]) > \frac{9}{10}$. Then, for any $d > \frac{1}{c}$, T, and m large enough, it holds that*

$$\mathbb{E}_\mu [W] \geq \frac{mT}{8d}.$$

5 Analysis of the Patient Algorithm

In this section, we consider the loss of the patient algorithm for the case of agents with a maximum sojourn time of exactly one time unit. We find an exponentially small loss, complementing the consideration of exponentially distributed departure times considered by Akbarpour et al. (2020). The central idea of the proof is to use the intuition that the probability of being matched is high if the pool size is high. Therefore, the key step is to show that the pool size is moderately high with very high probability. To this end, we perform a case analysis of the near past of a fixed point in time, considering four time intervals of length 1/3. The critical insight is that few agents arriving during the first time interval can be present at the beginning of the last time interval. To show this, we perform a comparison with an urn model.

Theorem 5.1. *Assume that the departure time is distributed according to the degenerate probability measure μ with $\mu(\{1\}) = 1$. Then, it holds that*

$$\mathbf{L}_{PAT} \leq e^{-\frac{d}{5}}.$$

Proof. As μ is fixed throughout the proof, we omit it as subscript. By Lemma 4.5 and unit departure times, we have

$$\mathbf{L}_{\text{PAT}} = \limsup_{m,T\to\infty} \frac{1}{T}\int_1^T \mathbb{P}\left(\text{an agent arriving at time } t-1 \text{ perishes}\right) dt.$$

Assume that the pool has a size of k at time t and an agent gets critical. Then, the probability that the agent perishes is given by $\left(1-\frac{d}{m}\right)^{k-1}$. Thus, in order to prove small loss, it suffices to show that the pool size is large with relatively high probability. Let $t \geq \frac{4}{3}$ be arbitrary. Our goal is to show that $z_t > \frac{m}{5}$ with high probability. For this, we condition on the pool size at time $t-\frac{1}{3}$. By the law of total probability, we have

$$\mathbb{P}\left(z_t \leq \frac{m}{5}\right) = \mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t-\frac{1}{3}} < \frac{m}{8}\right)\mathbb{P}\left(z_{t-\frac{1}{3}} < \frac{m}{8}\right) + \mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t-\frac{1}{3}} \geq \frac{m}{8}\right)\mathbb{P}\left(z_{t-\frac{1}{3}} \geq \frac{m}{8}\right) \tag{7}$$

and thus it suffices to bound each of the two summands above.

We start with the first one. Each agent that is in the pool at time $t-\frac{1}{3}$ can take at most one additional agent out of the pool when she gets critical. Thus, conditioned on the event where $z_{t-\frac{1}{3}} < \frac{m}{8}$, it holds that $z_t \geq a_{(t-\frac{1}{3},t)} - z_{t-\frac{1}{3}} \geq a_{(t-\frac{1}{3},t)} - \frac{m}{8}$. By independence of $a_{(t-\frac{1}{3},t)}$ and $z_{t-\frac{1}{3}}$, this already implies that

$$\mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t-\frac{1}{3}} < \frac{m}{8}\right) \leq \mathbb{P}\left(a_{(t-\frac{1}{3},t)} \leq \frac{m}{5} + \frac{m}{8} \mid z_{t-\frac{1}{3}} < \frac{m}{8}\right) = \mathbb{P}\left(a_{(t-\frac{1}{3},t)} \leq \frac{13}{40}m\right) \overset{\text{Lemma C.2}}{\leq} m^{-2}$$

for m large enough. This directly implies that

$$\mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t-\frac{1}{3}} < \frac{m}{8}\right)\mathbb{P}\left(z_{t-\frac{1}{3}} < \frac{m}{8}\right) \leq m^{-2}$$

for m large enough. Now let us consider the bound for the second addend in (7). Here we have that

$$\mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t-\frac{1}{3}} \geq \frac{m}{8}\right)\mathbb{P}\left(z_{t-\frac{1}{3}} \geq \frac{m}{8}\right) = \mathbb{P}\left(z_t \leq \frac{m}{5}, z_{t-\frac{1}{3}} \geq \frac{m}{8}\right). \tag{8}$$

We now condition on the pool size at time $t-\frac{1}{3}$ and on the number of arriving agents in certain intervals. This is necessary to avoid possible coherences between the evolution of the pool size and the arrival of new agents. For this, define the vector $a = (a_1, a_2, a_3) \coloneqq \left(a_{[t-\frac{4}{3},t-1)}, a_{[t-1,t-\frac{2}{3})}, a_{[t-\frac{2}{3},t-\frac{1}{3})}\right)$ to get

$$\mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} \ge \frac{m}{8}\right) = \sum_{k_1,k_2,k_3=0}^{\infty} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) \tag{9}$$

$$= \sum_{k_1,k_2,k_3=\lfloor m/3-m^{2/3}\rfloor}^{\lfloor m/3+m^{2/3}\rfloor} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right)$$

$$+ \sum_{\substack{(k_1,k_2,k_3)\in\mathbb{N}^3:\\ (k_1,k_2,k_3)\notin\left[\lfloor m/3-m^{2/3}\rfloor, \lfloor m/3+m^{2/3}\rfloor\right]^3}} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) .$$

The second summand in the above sum can be bounded by

$$\sum_{\substack{(k_1,k_2,k_3)\in\mathbb{N}^3:\\ (k_1,k_2,k_3)\notin\left[\lfloor m/3-m^{2/3}\rfloor, \lfloor m/3+m^{2/3}\rfloor\right]^3}} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right)$$

$$\le \sum_{\substack{(k_1,k_2,k_3)\in\mathbb{N}^3:\\ (k_1,k_2,k_3)\notin\left[\lfloor m/3-m^{2/3}\rfloor, \lfloor m/3+m^{2/3}\rfloor\right]^3}} \mathbb{P}\left(a = (k_1,k_2,k_3)\right)$$

$$\le \sum_{\substack{k\in\mathbb{N}:\\ k\notin\left[m/3-m^{2/3}, m/3+m^{2/3}\right]}} \mathbb{P}(a_1 = k) + \mathbb{P}(a_2 = k) + \mathbb{P}(a_3 = k)$$

$$= 3 \sum_{\substack{k\in\mathbb{N}:\\ k\notin\left[m/3-m^{2/3}, m/3+m^{2/3}\right]}} \mathbb{P}(a_1 = k) \overset{\text{Lemma C.2}}{\le} 6e^{-m^{1/3}} \le m^{-2}.$$

The equality follows because a_1, a_2, a_3 are identically distributed, and the last two inequalities hold for large enough m. Inserting this into (9), we see that

$$\mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} \ge \frac{m}{8}\right) \le \sum_{k_1,k_2,k_3=\lfloor m/3-m^{2/3}\rfloor}^{\lfloor m/3+m^{2/3}\rfloor} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) + m^{-2}$$

for all large enough m. Our next goal is to show that

$$\mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) \le m^{-2}\mathbb{P}\left(z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right), \tag{10}$$

as this already implies that

$$\mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} \ge \frac{m}{8}\right) \le \sum_{k_1,k_2,k_3=\lfloor m/3-m^{2/3}\rfloor}^{\lfloor m/3+m^{2/3}\rfloor} \sum_{l \ge \frac{m}{8}} \mathbb{P}\left(z_t \le \frac{m}{5}, z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) + m^{-2}$$

$$\le \sum_{k_1,k_2,k_3=\lfloor m/3-m^{2/3}\rfloor}^{\lfloor m/3+m^{2/3}\rfloor} \sum_{l \ge \frac{m}{8}} m^{-2}\mathbb{P}\left(z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) + m^{-2} \le 2m^{-2}.$$

By the law of conditional probability, it suffices to show

$$\mathbb{P}\left(z_t \le \frac{m}{5} \mid z_{t-\frac{1}{3}} = l, a = (k_1,k_2,k_3)\right) \le m^{-2} \tag{11}$$

for all $k_1, k_2, k_3 \in \left[\lfloor m/3 - m^{2/3} \rfloor, \lfloor m/3 + m^{2/3} \rfloor\right]$ and $l \geq \frac{m}{8}$, because this already implies (10).

For $i \in \{1, 2, 3\}$, let $v_1^i, v_2^i, \ldots, v_{k_i}^i$ be the k_i many agents that arrive during the time interval $\left[t - \frac{5-i}{3}, t - \frac{4-i}{3}\right)$. We define the random variable K_1 by

$$K_1 = \sum_{j=1}^{k_1} \mathbf{1}_{\left\{v_j^1 \text{ is in the pool at time } t - \frac{1}{3}\right\}},$$

where $\mathbf{1}$ denotes the indicator function. From our conditions, we already know that $K_1 \leq k_1$.

In the next step, we make use of the following observation: For each agent x that arrives during the interval $\left[t - \frac{4}{3}, t - 1\right)$ and each agent y that arrives during the time interval $\left[t - 1, t - \frac{1}{3}\right)$, the chances of agent y of making it to the pool at time $t - \frac{1}{3}$ are at least as high as for agent x, as the agent x witnesses at least as many critical agents as y. Moreover, since $k_1, k_2, k_3 \in \left[\lfloor \frac{m}{3} - m^{2/3} \rfloor, \lfloor \frac{m}{3} + m^{2/3} \rfloor\right]$, it holds that $\frac{k_1}{k_1 + k_2 + k_3} \leq \frac{2}{5}$ for m large enough. We will assume that $\frac{k_1}{k_1 + k_2 + k_3} \leq \frac{2}{5}$ henceforth. Thus, we can infer that

$$\mathbb{E}\left[K_1 \mid z_{t - \frac{1}{3}} = l, a = (k_1, k_2, k_3)\right] \leq l \frac{k_1}{k_1 + k_2 + k_3} \leq \frac{2}{5} l.$$

We want to translate this bound on K_1, that holds in expectation, to a bound that holds with high probability, i.e., our next goal is to show that $K_1 < l/2$ with high probability. The distribution of K_1 highly depends on the arrival/departure-process, as agents that arrive at a later point in time typically witness less critical agents up to time $t - \frac{1}{3}$. We expect K_1 to be maximal, when all agents getting critical in the time frame $\left[t - \frac{4}{3}, t - \frac{1}{3}\right)$ get critical just before time step $t - \frac{1}{3}$, as all agents arriving in the time frame $\left[t - \frac{4}{3}, t - \frac{1}{3}\right)$ have the same chance of making it to the pool $Z_{t - \frac{1}{3}}$ in this case. In all other cases, agents that arrive earlier always have a lower chance of making it to the pool. Thus, we can bound the probability of $K_1 > \frac{l}{2}$ with the outcome of a different experiment, namely an urn experiment, that we describe now.

Let $n = k_2 + k_3$ and consider an urn with $N = k_1 + k_2 + k_3 = k_1 + n$ balls in it, where k_1 of them are red and $k_2 + k_3$ of them are blue. We draw $l \leq N$ balls out of this urn without replacement. Let $\tilde{K_1}$ be the number of red balls taken. We defer the technical proof of the next steps to two lemmas discussed in the appendix. By Lemma B.1, the random variable $\tilde{K_1}$ stochastically dominates K_1, so we need to bound the probability that $\tilde{K_1} \geq l/2$. By Lemma B.2, it holds that $\mathbb{P}\left(\tilde{K_1} \geq \frac{l}{2}\right) \leq 2m\, 0.98^{m/8}$. Consequently,

$$\mathbb{P}\left(K_1 \geq \frac{l}{2}\right) \leq \mathbb{P}\left(\tilde{K_1} \geq \frac{l}{2}\right) \leq 2m 0.98^{m/8} \leq 0.998^m \leq m^{-3},$$

where the last two inequalities hold for m large enough. If $K_1 \leq \frac{l}{2}$, then $z_t \geq z_{t - \frac{1}{3}} - 2K_1 + a_{(t - \frac{1}{3}, t)} \geq a_{(t - \frac{1}{3}, t)}$. So, we can conclude in particular that

$$\begin{aligned}
&\mathbb{P}\left(z_t \leq \frac{m}{5} \mid z_{t - \frac{1}{3}} = l, a = (k_1, k_2, k_3)\right) \\
&\leq \mathbb{P}\left(K_1 > \frac{l}{2} \mid z_{t - \frac{1}{3}} = l, a = (k_1, k_2, k_3)\right) + \mathbb{P}\left(a_{(t - \frac{1}{3}, t)} < \frac{m}{5} \mid z_{t - \frac{1}{3}} = l, a = (k_1, k_2, k_3)\right) \\
&\overset{\text{Lemma C.2}}{\leq} m^{-3} + \mathbb{P}\left(a_{(t - \frac{1}{3}, t)} < \frac{m}{5}\right) \leq m^{-2}
\end{aligned}$$

for m large enough, which finally shows (11). For the second inequality we used that $a_{(t - \frac{1}{3}, t)}$ is independent of $z_{t - \frac{1}{3}}$ and $a = (a_1, a_2, a_3)$. Together with (7), this already implies that

$$\mathbb{P}\left(z_t \leq \frac{m}{5}\right) \leq m^{-1}$$

for m large enough. With this result we get that

$$\begin{aligned}&\mathbb{P}\,(\text{an agent arriving at time } t-1 \text{ perishes }) \leq \mathbb{P}\,(\text{an agent getting critical at time } t \text{ perishes }) \\ &= \mathbb{P}\left(\text{an agent getting critical at time } t \text{ perishes } \mid z_t \geq \frac{m}{5}\right)\mathbb{P}\left(z_t \geq \frac{m}{5}\right) \\ &+ \mathbb{P}\left(\text{an agent getting critical at time } t \text{ perishes } \mid z_t < \frac{m}{5}\right)\mathbb{P}\left(z_t < \frac{m}{5}\right) \\ &\leq \left(1-\frac{d}{m}\right)^{\frac{m}{5}-1} + m^{-1} \underset{m\to\infty}{\longrightarrow} e^{-\frac{d}{5}}.\end{aligned}$$

□

6 Simulations

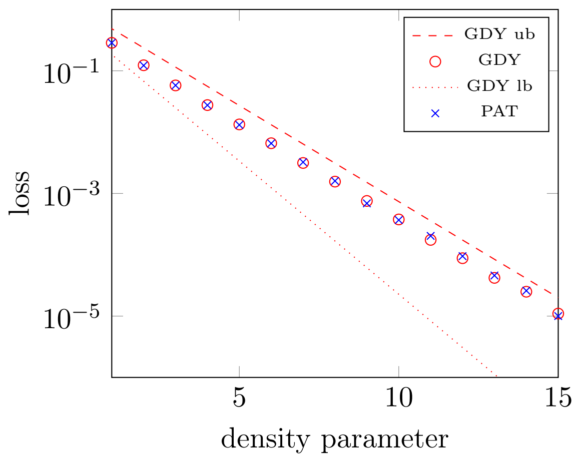
In the previous sections, we have provided strong theoretical bounds on the loss of the greedy and patient algorithm, particularly for the case of constant departure times. These bounds are guaranteed to hold for m and T tending to infinity and give an indication on the market behavior for fixed m or T. In order to complete the picture, we provide simulated results for different input values. Note that matching markets are often of large scale in reality, see, e.g., Ünver (2010) for a discussion of kidney exchange markets and Baccara et al. (2014) for a discussion of child adoption markets. Their data on the poolsize (greater than 50 000) is confirmed by recent numbers from the Organ Procurement and Transplantation Network and the US Census 2020, respectively. We investigate the market behavior with respect to key characteristics like the percentage of unmatched agents or the average waiting time of agents for realistically sized and randomly generated examples. In particular, this analysis gives insights on the significance of error terms in practical examples.

In order to simulate the market behavior, we start with an empty pool at time 0, i.e., no agents are present initially. Then, agents join the pool at Poisson rate m. This means in particular that the difference between the arrival time of two successively arriving agents is distributed with respect to an exponential distribution with expectation $\frac{1}{m}$, i.e., the arrival of the next agent can always be simulated by evaluating a random variable. Every agent is assigned a maximum sojourn time at her arrival to the market according to a predetermined distribution. During the simulation, a matching algorithm is employed and used to determine pairs of agents from the pool to be matched.

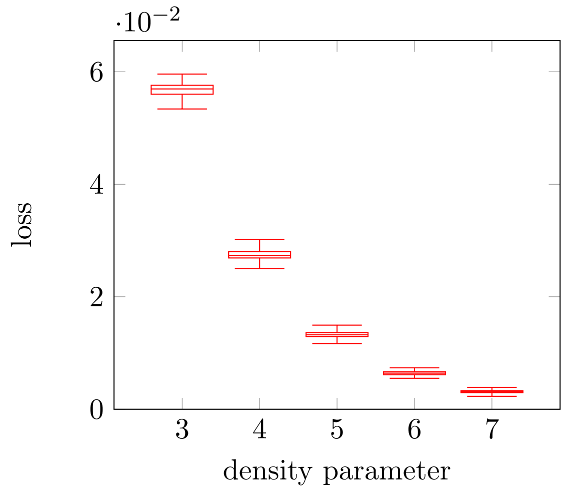
Our main focus lies on the greedy algorithm with constant unit departure times, for which we evaluate the average proportion of unmatched agents and the waiting time. The results of our simulations are depicted in Figure 2. First, we consider the average proportion of unmatched agents for the two algorithms in Figure 2a. The simulations indicate an exponentially small loss with respect to d for both algorithms, which is in line with the bounds obtained in our theoretical findings. Interestingly, the average proportion of unmatched agents in both algorithms appears to be surprisingly similar, hinting at a possible connection between two basic algorithms using neither structural information of the compatibility relation nor timing information beyond arrivals and critical states of agents. In Figure 2b, we display a detailed analysis of the average proportion of unmatched agents of the greedy algorithm under small density parameters by means of box-and-whisker plots.[1] We can see that there is little spread, indicating the robustness of the theoretically proved statements for very sparse markets.

Instead of choosing a match uniformly at random from the set of compatible agents, it is also possible to use further structural or timing information in order to perform a selection. If the social planner receives additional information on the maximum remaining sojourn time of

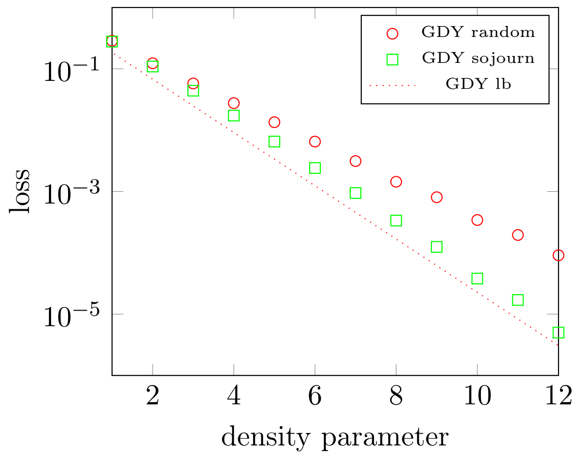
[1]The whiskers denote the minimum and maximum of the average proportion of unmatched agents, respectively, while the lower and upper bound of the box represent the first and third quartile. Finally, the middle line corresponds to the median.



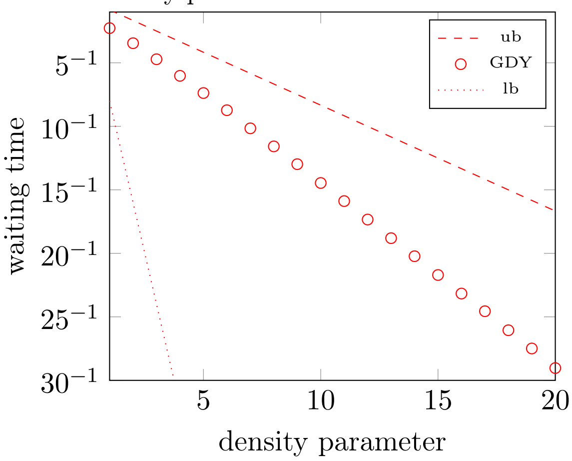
(a) Logarithmic plot of average proportion of unmatched agents for GDY and PAT.



(b) Box-and-whisker plots describing average proportion of unmatched agents for GDY under small density parameter.



(c) Logarithmic plot of average proportion of unmatched agents for GDY under different match selection.



(d) Inverse linear plot of average waiting time for GDY.

Figure 2: Visualization of results of simulations. Each datapoint in the first, third, and forth figure corresponds to 10 runs for $m = 1000$ and $T = 100$. Each box-and-whisker plot in the second figure represents 200 runs for $m = 500$ and $T = 100$. Note that the linear behavior in the logarithmic plots of the first and third figure mean that the decay is exponential.

individual agents, one natural matching choice is to select the compatible agent with the least remaining sojourn as a partner. This may avoid the perishing of an agent with small remaining sojourn time. In fact, this algorithm performs close to our guaranteed lower bound for the loss of greedy algorithms as can be seen in Figure 2c. Note that this lower bound does not depend on the choice of the match, but even holds for an arbitrary matching policy.

Finally, we investigate the average waiting time of agents for the greedy algorithm in Figure 2d. Note that the upper bound on the waiting time in Proposition 4.13 already holds for relatively small arrival rates, approximately at $m \approx 125$. Our simulations show that the aggregated waiting time is consistently below the expected upper bound, again demonstrating the superiority of the greedy algorithm over the patient algorithm with respect to waiting time.

7 Discussion and Conclusion

We have analyzed instantaneous matching as captured by the greedy algorithm and matching at the exact departure times as captured by the patient algorithm. We found out that the performance of the greedy algorithm strongly depends on the distribution of the departure times. In particular, if agents are guaranteed to stay in the market for some time, then the greedy algorithm achieves an exponentially small loss. This is a striking result because we operate in a thin market, whereas previous theoretical results proving the optimality of matching algorithms heavily relied on market thickness.[2] We thus offer a new type of justification to promote greedy algorithms in dynamic matching markets.

To conclude the paper, we elaborate on two important aspects of our work that give prospect to further investigation. First, we consider the loss equivalence of the greedy and patient algorithm under unit waiting times as observed in our simulations. Second, we discuss possible extensions of our work to the case of heterogeneous agents.

7.1 Equivalence of Greedy and Patient Algorithm

While our theoretical results only show an exponential loss in both cases, Figure 2a suggests that the loss of the two algorithms is (approximately) identical. Indeed, there is some intuition why the loss should be approximately $\frac{1}{2}e^{-\frac{d}{2\ln(2)}}$ for both the greedy and the patient algorithm. In this section, we provide two different, non-rigorous arguments which lead to exactly the loss observed in our simulations in both cases. Unfortunately, these arguments are heavily based on steady-state analysis and are therefore hard to be made precise for *non*-Markovian processes. Identifying a relationship of the two algorithms that establishes an identical loss is an intriguing problem for further research.

Greedy algorithm We start by reasoning about the pool size similar to the proof sketch in Section 4.1. Let z_{Eq} denote the pool size at equilibrium. Assuming that the loss is negligible, half of the agents instantaneously match at arrival. The probability of not forming an edge at arrival is about

$$\left(1 - \frac{d}{m}\right)^{z_{Eq}} \approx e^{-d\frac{z_{Eq}}{m}} \stackrel{!}{=} \frac{1}{2}$$

which suggests that the pool in equilibrium should be of size $\frac{\ln(2)}{d}m$. If an agent v does form an edge at arrival, but joins the pool, approximately m agents will arrive during her maximal sojourn time and approximately $\frac{m}{2}$ of these agents will match at arrival. The probability that a

[2] For instance, the market under the greedy algorithm is of size $\mathcal{O}(\frac{m}{d})$ for every distribution of the departure times (see the proof of Proposition 4.13), whereas the market size under the patient algorithm is of order $\Theta(m)$ most of the time under exponentially distributed departure times (Akbarpour et al., 2020).

specific agent, matching at arrival, matches with v equals $\frac{1}{z_{Eq}}$, by symmetry. So assuming that z_{Eq} stays approximately constant over time, the probability that v perishes can be estimated as

$$\left(1-\frac{1}{z_{Eq}}\right)^{\frac{m}{2}} \approx \left(1-\frac{d}{m\ln(2)}\right)^{\frac{m}{2}} \approx e^{-\frac{d}{2\ln(2)}}.$$

We conditioned on v not matching at arrival, but joining the pool instead. This happens for approximately half of all agents. Thus, the total loss equals approximately $\frac{1}{2}e^{-\frac{d}{2\ln(2)}}$.

Patient algorithm Again, we first want to control the pool size. Assuming a negligible loss, for a fixed agent v, approximately $\frac{m}{2}$ agents get critical in the interval $[t_v, t_v+1]$, where t_v is the arrival time of v. Under the assumption that the pool size stays constant over time, by symmetry, the probability that none of these agents matches with v should equal

$$\left(1-\frac{1}{z_{Eq}}\right)^{\frac{m}{2}} \approx e^{-\frac{m}{2\,z_{Eq}}} \stackrel{!}{=} \frac{1}{2}.$$

Solving this yields $z_{Eq} \approx \frac{m}{2\ln(2)}$. Now, the probability that an agent getting critical perishes equals approximately

$$\left(1-\frac{d}{m}\right)^{z_{Eq}} \approx \left(1-\frac{d}{m}\right)^{\frac{m}{2\ln(2)}} \approx e^{-\frac{d}{2\ln(2)}}.$$

Assuming the loss is small, approximately half of the agents get critical, i.e., they are not matched with an other agent before the end of their maximal sojourn time. Thus, the total loss among all agents equals approximately $\frac{1}{2}e^{-\frac{d}{2\ln(2)}}$. Hence, both arguments lead to an approximately equal loss.

7.2 Heterogeneous Agents

Many applications of dynamic matching markets comprise *heterogenous* agents differing in their market behavior, for instance with respect to their urgency to be matched or with respect to their capability to form matches with other agents. In this section, we give a brief overview of how to extend our results to markets with heterogeneous agents.

First, it is possible to extend our results to a setting with heterogeneous agents regarding their urgency to be matched, that is, agents differing with respect to their maximum sojourn times. As long as all agents *individually* satisfy the conditions on the maximum sojourn time in our results, they still work. To be more explicit, consider the following scenario. There are two types of agents arriving to the market at random. The first type consists of *urgent* agents, which have a maximum sojourn time of 1. The second type of agents consists of *patient* agents having a maximal sojourn time of $C > 1$. Assume that urgent agents arrive at the market at Poisson rate m_1, whereas patient agents independently from urgent agents arrive at Poisson rate m_2. Denote $m = m_1 + m_2$ and assume that the probability that two agents are compatible is $p = \frac{d}{m}$, where d is a constant. Then, the consequence of Theorem 4.6 still holds when m tends to infinity. Indeed, the sum of two Poisson processes is again a Poisson process, with the sum of the previous rate functions as a new rate function. Hence, the model with urgent and patient agents has the same distribution as the original model with incoming rate $m = m_1 + m_2$, where the maximal sojourn time of an agent is 1 with probability $\frac{m_1}{m_1+m_2}$ and C with probability $\frac{m_2}{m_1+m_2}$.

Second, a common type of heterogeneity considered in the literature differentiates easy-to-match and hard-to-match agents (Ashlagi et al., 2019b, 2022), differing regarding compatibility. In our setting, this can be modeled by having agents arrive with Poisson rate m where membership to a class is determined by a random variable. Then, the compatibility of agents is decided according to different density parameters d_{HH}, d_{EH}, and d_{EE} with $d_{HH} \leq d_{EH} \leq d_{EE}$ representing the density of edges between two hard-to-match, two different-type, and two easy-to-match agents, respectively. This leads to matching probabilities $p_{HH} = d_{HH}/m$, $p_{EH} = d_{EH}/m$, and $p_{EE} = d_{EE}/m$. It is possible to extend our results under certain assumptions on the relationship between the different density parameters. Still, a rigorous treatment of this model is beyond the scope of this paper, and we leave it as an intriguing direction for further research.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) under grants BR 2312/11-2, BR 2312/12-1, and 277991500/GRK2201. We thank Matthias Greger, Shunya Noda, and Peng Shi for valuable feedback.

References

I. Agarwal, R. Cole, and Y. Tao. Selecting a match: Exploration vs decision. Technical Report arXiv:2106.08033, arXiv.org, 2021.

M. Akbarpour, S. Li, and S. O. Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815, 2020.

R. Anderson, I. Ashlagi, D. Gamarnik, and Y. Kanoria. Efficient dynamic barter exchange. *Operations Research*, 65(6):1446–1459, 2017.

I. Ashlagi, A. Bingaman, M. Burq, V. H. Manshadi, D. Gamarnik, C. Murphey, A. E. Roth, M. L. Melcher, and M. A. Rees. Effect of match-run frequencies on the number of transplants and waiting times in kidney exchange. *American Journal of Transplantation*, 18(5):1177–1186, 2018.

I. Ashlagi, M. Burq, C. Dutta, P. Jaillet, A. Saberi, and C. Sholley. Edge weighted online windowed matching. In *Proceedings of the 20th ACM Conference on Economics and Computation (ACM-EC)*, pages 729–742, 2019a.

I. Ashlagi, M. Burq, P. Jaillet, and V. H. Manshadi. On matching and thickness in heterogeneous dynamic markets. *Operations Research*, 67(4):927–949, 2019b.

I. Ashlagi, N. Afshin, and P. Strack. Matching in dynamic imbalanced markets. *Review of Economic Studies*, 2022. Forthcoming.

A. Aveklouris, L. DeValve, and A. R. Ward. Matching impatient and heterogeneous demand and supply. Technical Report arXiv:2102.02710, arXiv.org, 2021.

M. Baccara, A. Collard-Wexler, L. Felli, and L. Yariv. Child-adoption matching: Preferences for gender and race. *American Economic Journal: Applied Economics*, 6(3):133–58, 2014.

M. Baccara, S. Lee, and L. Yariv. Optimal dynamic matching. *Theoretical Economics*, 15(3): 1221–1278, 2020.

J. N. Bearden. A new secretary problem with rank-based selection and cardinal payoffs. *Journal of Mathematical Psychology*, 50(1):58–59, 2006.

J. H. Blanchet, M. I. Reiman, V. Shah, L. M. Wein, and L. Wu. Asymptotically optimal control of a centralized dynamic matching market with general utilities. *Operations Research*, 2022. Forthcoming.

N. Collina, N. Immorlica, K. Leyton-Brown, B. Lucier, and N. Newman. Dynamic weighted matching with heterogeneous arrival and departure rates. In *Proceedings of the 16th International Conference on Web and Internet Economics (WINE)*, pages 17–30, 2020.

Y. Emek, S. Kutten, and R. Wattenhofer. Online matching: haste makes waste! In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 333–344, 2016.

P. Erdős and A. Rényi. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, 5(1):17–61, 1960.

J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *Proceedings of the 50th Symposium on Foundations of Computer Science (FOCS)*, pages 117–126, 2009.

T. S. Ferguson. Who solved the secretary problem? *Statistical Science*, 4(3):282–289, 1989.

J. Fong. Search, selectivity, and market thickness in two-sided markets: Evidence from online dating. Technical Report 3458373, SSRN, 2020.

T. Hagerup and C. Rüb. A guided tour of chernoff bounds. *Information Processing Letters*, 33 (6):305–308, 1990.

N. Kakimura and D. Zhu. Dynamic bipartite matching market with arrivals and departures. In *Proceedings of the 17th International Conference on Web and Internet Economics (WINE)*, page 544, 2021.

R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 352–358, 1990.

S. Kerimov, I. Ashlagi, and I. Gurvich. On the optimality of greedy policies in dynamic matching. In *Proceedings of the 23rd ACM Conference on Economics and Computation (ACM-EC)*, 2022. Forthcoming.

A. Klenke. *Probability theory: a comprehensive course.* Springer Science & Business Media, 2013.

G. Last and M. Penrose. *Lectures on the Poisson process*, volume 7. Cambridge University Press, 2017.

J. D. Leshno. Dynamic matching in overloaded waiting lists. *Available at SSRN 2967011*, 2021. Working Paper.

J. Li and S. Netessine. Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment. *Management Science*, 66(1):271–289, 2020.

S. Loertscher, E. V. Muir, and P. G. Taylor. Optimal market thickness. *Journal of Economic Theory*, 200:105383, 2022.

V. H. Manshadi, S. O. Gharan, and A. Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.

A. Mehta, A. Saberi, U. V. Vazirani, and V. V. Vazirani. Adwords and generalized online matching. *Journal of the ACM*, 54(5):22–es, 2007.

P. Mertikopoulos, H. H. Nax, and B. S. R. Pradelski. Quick or cheap? breaking points in dynamic markets. In *Proceedings of the 21st ACM Conference on Economics and Computation (ACM-EC)*, pages 877–878, 2020.

M. Pavone, A. Saberi, M. Schiffer, and M. Tsao. Online hypergraph matching with delays. Technical Report arXiv:2009.12022, arXiv.org, 2020.

P. P. Reese, N. Boudville, and A. X. Garg. Living kidney donation: outcomes, ethics, and uncertainty. *The Lancet*, 385(9981):2003–2013, 2015.

A. E. Roth. Marketplaces, markets, and market design. *American Economic Review*, 108(7): 1609–58, 2018.

A. E. Roth, T. Sönmez, and M. U. Ünver. Kidney exchange. *The Quarterly Journal of Economics*, 119(2):457–488, 2004.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151–188, 2005.

R. Shimer and L. Smith. Matching, search, and heterogeneity. *The BE Journal of Macroeconomics*, 1(1), 2001.

M. U. Ünver. Dynamic kidney exchange. *The Review of Economic Studies*, 77(1):372–414, 2010.

A Analysis of the Greedy Algorithm

In this section, we provide the omitted proofs of the analysis of the greedy algorithm.

We start with the technical lemma that allows us to compute the loss by integrating over the probability of perishing.

Lemma 4.5. *Let a matching algorithm ALG be given and assume that departure time is distributed according to an arbitrary probability μ. Then,*

$$\mathbf{L}_{ALG}(m,T) = \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes before time } T\right) dt,$$

and

$$\mathbf{L}_{ALG} = \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes}\right) dt.$$

Proof. In the proof, we omit the subscript μ, because the distribution of the sojourn times are fixed. Recall that X_i denotes the maximum sojourn time of agent v_i. In Section 3 we described that every agent v carries a random variable $U(v)$ that determines with which agents in the pool she forms an edge at arrival. We also assume that attached to agent v there is a random variable $T(v)$, which has the distribution Uniform $[0,1]^{\otimes\mathbb{N}}$. Whenever an agent v decides to match with another agent, but has several possibilities, it choose the agent v_i, where $T_i(v)$ is minimal. As the uniform distribution is absolutely continuous with respect to the Lebesgue measure, there is almost surely a unique minimizer. By symmetry, the agent that v matches with is chosen uniformly among all possibilities. So T helps with breaking ties, in case there are some. For the complete process we have four different sources of randomness: The incoming agents arrive random, according to a Poisson process, their departure times are random with distribution μ, the forming of edges which is determined by the vectors $U(v)$, and the random breaking of ties, if necessary, that is determined by the vectors $T(v)$. Let $Z = ((X_i, U(v_i), T(v_i)))_{i\geq 1}$ be the random variable containing all information about the random variables $X_i, U(v_i)$ and $T(v_i)$ for $i \geq 1$. Given Z, the evolution of an agent v_i depends only on the random variable $Y = (Y_t)_{t\in[0,T]}$ distributed according to a Poisson process with rate m, where Y_t is the number of agents arriving at time t. Thus, we can define the function

$$f_Z(t,Y) \coloneqq \mathbf{1}\left\{\text{an agent arrives at time } t \text{ and perishes before time } T\right\},$$

where $\mathbf{1}$ is the indicator function. The loss until time T can clearly be expressed as $\mathbf{L}_{\mathrm{ALG}}(m,T) = \frac{1}{mT}\mathbb{E}\left[\int_0^T f_Z(t,Y)Y(dt)\right]$. Note that Z and Y are independent, so the process Y is still the same Poisson process, given the information contained in Z. Furthermore, the intensity measure of this Poisson process is simply m times the Lebesgue measure on $[0,T]$. Thus, we can apply the tower property in the first and fourth equality and Mecke's equation (see, e.g., Last and Penrose, 2017, Theorem 4.1) in the second equality to compute

$$\begin{aligned}
\mathbb{E}\left[\int_0^T f_Z(t,Y)Y(dt)\right] &= \mathbb{E}\left[\mathbb{E}\left[\int_0^T f_Z(t,Y)Y(dt)\Big|\sigma(Z)\right]\right] = \mathbb{E}\left[\int_0^T \mathbb{E}\left[f_Z(t,Y+\delta_t)\Big|\sigma(Z)\right] m dt\right] \\
&= m\int_0^T \mathbb{E}\left[\mathbb{E}\left[f_Z(t,Y+\delta_t)\Big|\sigma(Z)\right]\right] dt = m\int_0^T \mathbb{E}\left[f_Z(t,Y+\delta_t)\right] dt \\
&= m\int_0^T \mathbb{P}\left(\text{an agent arriving at time } t \text{ perishes before time } T\right) dt.
\end{aligned}$$

There, $\sigma(Z)$ denotes the σ-algebra generated by Z, and δ_t is the Dirac delta function, i.e., $\delta_t(s) = \mathbf{1}_{\{s=t\}}$. Now, let us consider the second statement of the lemma. By the equality that we proved just before, we know that

$$\begin{aligned}\mathbf{L}_{\text{ALG}} &= \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes before time } T\right) dt \\ &= \limsup_{m,T\to\infty} \int_0^1 \mathbb{P}_\mu\left(\text{an agent arriving at time } sT \text{ perishes before time } T\right) dt. \qquad (12)\end{aligned}$$

where the last equality holds by integration by substitution. For $T > 0$, we define the functions $\tilde{f}_T, f_T : [0,1] \to [0,1]$ by

$$\begin{aligned}f_T(s) &:= \mathbb{P}_\mu\left(\text{an agent arriving at time } sT \text{ perishes before time } T\right), \\ \tilde{f}_T(s) &:= \mathbb{P}_\mu\left(\text{an agent arriving at time } sT \text{ perishes}\right).\end{aligned}$$

We clearly have $f_t(s) \le \tilde{f}_T(s)$ for all $s \in [0,1]$, and for fixed $s < 1$ and $m > 0$ we also have

$$\begin{aligned}0 \le \tilde{f}_T(s) - f_T(s) &= \mathbb{P}_\mu\left(\text{an agent arriving at time } sT \text{ perishes after time } T\right) \\ &\le \mathbb{P}_\mu\left(\text{the maximum sojourn of an agent is in } [(1-s)T, \infty)\right) \\ &= \mu\left([(1-s)T, \infty)\right)\end{aligned}$$

which converges to 0 as T tends to infinity. As the difference $f_T(s) - \tilde{f}_T(s)$ is bounded by 1 for all $s \in [0,1]$ we get that

$$\lim_{T\to\infty} \int_0^1 f_T(s) - \tilde{f}_T(s) ds = 0$$

and as this convergence holds uniformly for all m, we get

$$\begin{aligned}\mathbf{L}_{\text{ALG}} &= \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes before time } T\right) dt \\ &= \limsup_{m,T\to\infty} \frac{1}{T}\int_0^T \mathbb{P}_\mu\left(\text{an agent arriving at time } t \text{ perishes}\right) dt.\end{aligned}$$

□

Next, we provide the proof for our lemma about scaling invariance.

Lemma 4.7. *Let μ be a probability measure on $[0, +\infty]$, let $c > 0$, and let ν be the probability measure defined by $\nu([a,b]) = \mu([ca, cb])$ for all $a, b \in [0, +\infty]$. Then,*

$$\mathbf{L}_{GDY}(\nu, d) = \mathbf{L}_{GDY}(\mu, cd).$$

Proof. The claim follows by a change of time. If we consider the model with departure times distributed according to ν and speed up time by a factor of c, then agents arrive at Poisson rate cm and the departure time X of an agent satisfies

$$\mathbb{P}(X \le a) = \nu([0, a/c]) = \mu([0, a]).$$

Thus, the departure times are distributed according to μ. However, two agents are still compatible with a probability of $\frac{d}{m} = \frac{cd}{cm}$. So, effectively we see a model with departure times distributed according to μ and density parameter cd. □

Now, we provide the proof for the auxiliary lemma used to find lower bounds for the loss of the greedy algorithm.

Lemma 4.9. *Let $\varepsilon, \delta > 0$ and assume that the departure time is distributed according to a probability measure μ with $\mu([0,\varepsilon]) \geq \delta$. Then,*

$$\mathbf{L}_{GDY} \geq \frac{\delta}{2} e^{-\varepsilon d}.$$

Proof. Consider an agent v_i arriving at time t which is not matched upon arrival. Note that compatible agents come at Poisson rate $m \cdot \frac{d}{m} = d$. Recall that X_i is the maximum sojourn time of v_i. As the maximum sojourn time and the arrival of compatible agents are independent, we get

$$\begin{aligned}
&\mathbb{P}_\mu(v \text{ perishes} \mid v \text{ not matched upon arrival}) \\
&\geq \mathbb{P}_\mu(\text{no compatible agent arrives in the interval } [t, t+X_i] \mid v \text{ not matched upon arrival}) \\
&\geq \mathbb{P}_\mu(X_i \leq \varepsilon)\, \mathbb{P}_\mu(\text{no compatible agent arrives in the interval } [t, t+\varepsilon]) \geq \delta e^{-\varepsilon d}.
\end{aligned}$$

Additionally, $\mathbb{P}_\mu(v \text{ not matched upon arrival}) \geq \frac{1}{2}$, because at least half of the agents v are not matched upon arrival. Together,

$$\begin{aligned}
&\mathbb{P}_\mu(v \text{ perishes}) \\
&= \mathbb{P}_\mu(v \text{ perishes} \mid v \text{ not matched upon arrival})\, \mathbb{P}_\mu(v \text{ not matched upon arrival}) \geq \frac{1}{2}\delta e^{-\varepsilon d}.
\end{aligned}$$

□

Finally, we provide the proof for the lower bound on the waiting time in the greedy algorithm. As discussed above, it uses a similar technique as the proof of Lemma 4.5.

Proposition 4.14. *Let $c > 0$ and assume that the departure time is distributed according to a probability measure μ with $\mu([c,\infty]) > \frac{9}{10}$. Then, for any $d > \frac{1}{c}$, T, and m large enough, it holds that*

$$\mathbb{E}_\mu[W] \geq \frac{mT}{8d}.$$

Proof. In the proof, we omit the subscript μ, because the distribution of the sojourn times are fixed. We use the same notation as in the proof of Lemma 4.5. Let $Z = ((X_i, U(v_i), T(v_i)))_{i\geq 1}$ be the random variable containing all information about the random variables $X_i, U(v_i)$, and $T(v_i)$ for $i \geq 1$. Given Z, the evolution of an agent v_i depends only on the random variable $Y = (Y_t)_{t\in[0,T]}$ distributed according to a Poisson process with rate m, where Y_t is the number of agents arriving at time t. Thus, we can define the function

$$g_Z(t,Y) = \begin{cases} \text{waiting time up to time } T \text{ of the agent arriving at time } t & \text{if } Y_t = 1, \\ 0 & \text{else.} \end{cases}$$

For this function we have $W = \int_0^T g_Z(t,Y)Y(dt)$. Given the random variable Z, the evolution of the pool depends only on the Poisson process Y. Thus Mecke's equation (see, e.g., Last and Penrose, 2017, Theorem 4.1) and the tower property yield that

$$\begin{aligned}
\mathbb{E}_\mu[W] &= \mathbb{E}_\mu\left[\int_0^T g_Z(t,Y)Y(dt)\right] = \mathbb{E}_\mu\left[\mathbb{E}_\mu\left[\int_0^T g_Z(t,Y)Y(dt)\middle|\sigma(Z)\right]\right] \\
&= \mathbb{E}_\mu\left[\int_0^T \mathbb{E}_\mu\left[g_Z(t,Y+\delta_t)\middle|\sigma(Z)\right] mdt\right] = \int_0^T \mathbb{E}_\mu\left[\mathbb{E}_\mu\left[g_Z(t,Y+\delta_t)\middle|\sigma(Z)\right]\right] mdt \\
&= m\int_0^T \mathbb{E}_\mu\left[g_Z(t,Y+\delta_t)\right] dt. \qquad (13)
\end{aligned}$$

Thus, it suffices to bound $\mathbb{E}_\mu[g_Z(t, Y+\delta_t)]$, which is the expected waiting time up to time T, of an agent v_i arriving at time t. Let $t < T - \frac{1}{d}$ be fixed. We know that $z_t \le \frac{4m}{5d}$ with very high probability, for m large, by Corollary 4.3 (note that $\frac{4}{5} > \ln(2)$). So an agent arriving at time t has a chance of at least $\left(1-\frac{d}{m}\right)^{\frac{4m}{5d}} - m^{-9} \ge \frac{2}{5}$ of making it to the pool, for m large enough. Compatible agents arrive at rate $m \cdot \frac{d}{m} = d$. Hence, there is a chance of at least e^{-1} that no compatible agent arrives during the time interval $\left[t, t+\frac{1}{d}\right]$. Furthermore, there is a chance of at least $\frac{9}{10}$ that the agent did not perish up to time $t+\frac{1}{d}$, as $\mu\left(\left[\frac{1}{d}, +\infty\right]\right) \ge \frac{9}{10}$. As all these events are independent, we get

$$\mathbb{E}_\mu[g_Z(t, Y+\delta_t)] \ge \frac{2}{5}\mathbb{P}(X_i > 1/d)\,\mathbb{P}\left(\text{no compatible agent arrives until time } t+\frac{1}{d}\right)\frac{1}{d} \ge \frac{2}{5}\frac{9}{10e}\frac{1}{d},$$

and inserting this into (13) yields

$$\mathbb{E}_\mu[W] \ge m\left(T-\frac{1}{d}\right)\frac{18}{50ed} \ge \frac{mT}{8d}$$

for T large enough. □

B Analysis of the Patient Algorithm

In this section, we provide the proof for the two auxiliary lemmas in the analysis of the patient algorithm with a maximum sojourn time of exactly one time unit.

Lemma B.1. *In the notation of the proof of Theorem 5.1, $\tilde{K}_1$ stochastically dominates K_1.*

Proof. We first describe three different experiments: In the first one, we have a pool $Z_{t-\frac{4}{3}}$ of some size. We call the agents in the pool at time $t-\frac{4}{3}$ the uncolored *black agents*. These agents get critical exactly one time step after their arrival, so they will not be in the pool at time $t-\frac{1}{3}$ anymore. The k_1 many agents arriving in the interval $\left[t-\frac{4}{3}, t-1\right)$ are called the *red agents* and the $n = k_2 + k_3$ many agents arriving in the interval $\left[t-1, t-\frac{1}{3}\right)$ are called the *blue agents*. Then agents can depart from the pool by perishing or by matching, so that l colored, i.e., red or blue, agents are in the pool at time $t-\frac{1}{3}$. We are interested in the number K_1, which is the number of red agents in the pool at time $\left[t-\frac{1}{3}\right]$.
In the second experiment, we have an urn with k_1 red balls and n blue balls and we take k_1+n-l balls out of the urn, uniformly at random without replacement. We are interested in the number of red balls K_1' that are left in the urn after this procedure.
In the third experiment, we have an urn with k_1 red balls and n blue balls and we take l balls out of the urn, uniformly at random without replacement. We are interested in the number of red balls $\tilde{K}_1$ that are taken out of the urn.

Leaving l balls in an urn or taking l balls out of an urn gives rise to the exact same distribution. Thus it suffices to show that K_1' stochastically dominates K_1. For this, we couple the urn to the pool in the following way: At time $t-\frac{4}{3}$, we start with the initial urn, i.e., the one with k_1 red balls and n blue balls in it. Let $(R_s)_{s\in\left[t-\frac{4}{3}, t-\frac{1}{3}\right]}$, respectively $(B_s)_{s\in\left[t-\frac{4}{3}, t-\frac{1}{3}\right]}$, be the number of red, respectively blue, agents in the pool at time s. Analogously, let $(R_s')_{s\in\left[t-\frac{4}{3}, t-\frac{1}{3}\right]}$, respectively $(B_s')_{s\in\left[t-\frac{4}{3}, t-\frac{1}{3}\right]}$, be the number of red, respectively blue, balls in the urn at time s. The underlying algorithm of the matching process in the first experiment does not distinguish between red and blue agents. So whenever an agent gets critical and matches with a colored agent, only the proportion of red/blue agents matters. So we can think of the process of the pool as follows: When an agent gets critical, say at time t_0, and matches with a colored agent, there is an underlying random variable $U_{t_0} \sim \text{Uniform}[0,1]$ that determines whether the agent matches with a blue or with a red agent. When $U_{t_0} \leq \frac{R_{t_0}}{R_{t_0}+B_{t_0}}$, then we choose uniformly at random one of the R_{t_0} many red agents. Otherwise we choose one of the B_{t_0} many blue agents uniformly at random. Whenever there is a matching with a colored agent in the pool, we also remove a ball from the urn. We pick a red ball if $U_{t_0} \leq \frac{R'_{t_0}}{R'_{t_0}+B'_{t_0}}$, and otherwise we pick a blue ball. As U_{t_0} is the uniform distribution, this also gives rise to the correct probabilities. Note that the proportion of red agents among the colored agents in the pool can change either when agents depart by matching or when a new agent arrives. Contrary, the proportion of red balls in the urn changes only at times when colored agents match in the pool. Let S_s, respectively S'_s be the number of red agents, respectively balls, that are removed from the pool, respectively urn, by time s. We claim that

$$S_s \geq S'_s \text{ for all } s \in \left[t - \frac{4}{3}, t - \frac{1}{3}\right] . \tag{14}$$

This is clear for $s < t - 1$, as S_s is simply the number of removed colored agents, as no blue agent entered the pool yet. Thus we can assume $s \geq t - 1$ from here on. Whenever $S_s > S'_s$ just before some time t_0, at which an agent gets critical and matches with a colored agent, then $S_s \geq S'_s$ still holds after this match. If $S_s = S'_s$ and $s \geq t-1$, then all red agents already arrived, which gives $R_s = k_1 - S_s = R'_s$. As the number of colored balls in the urn is at least as high as the number of colored agents in the pool, we have

$$\frac{R'_s}{R'_s + B'_s} \leq \frac{R_s}{R_s + B_s}.$$

Assume that $S_s = S'_s$ just before a match with a colored agent in the pool occurs. If a red ball is drawn from the urn, i.e., $U_s \leq \frac{R'_s}{R'_s+B'_s}$, then also a red agent leaves the pool in a matching. Thus $S_s \geq S'_s$ still holds directly after the match. So we also get that $S_{t-\frac{1}{3}} \geq S'_{t-\frac{1}{3}}$ which directly gives

$$K'_1 = R'_{t-\frac{1}{3}} = k_1 - S'_{t-\frac{1}{3}} \geq k_1 - S_{t-\frac{1}{3}} = R_{t-\frac{1}{3}}$$

and thus finishes the proof.

□

Lemma B.2. *In the notation of the proof of Theorem 5.1,* $\mathbb{P}\left(\tilde{K_1} \geq \frac{l}{2}\right) \leq 2m\, 0.98^{m/8}$.

Proof. Consider a set $U \subseteq \{1, \ldots, l\}$. Then, the probability of drawing a red ball at the i-th draw if and only if $i \in U$ is given by

$$\prod_{i=0}^{|U|-1} \frac{k_1 - i}{N - i} \prod_{i=0}^{l-|U|-1} \frac{n - i}{N - |U| - i} = \frac{1}{\prod_{i=0}^{l-1} N - i} \prod_{i=0}^{|U|-1} (k_1 - i) \prod_{i=0}^{l-|U|-1} (n - i).$$

Among all $U \subseteq \{1, \ldots, l\}$ with $|U| \geq \frac{l}{2}$, this is clearly maximized when $|U| = \lceil \frac{l}{2} \rceil$, as $n \geq k_1$. Let us assume that l is even from here on. The proof for l odd works completely similar, just with one additional term. So we have for all l even and all $U \subseteq \{1, \ldots, l\}$ with $|U| \geq \frac{l}{2}$ the bound

$$\begin{aligned}
&\frac{1}{\prod_{i=0}^{l-1} N - i} \prod_{i=0}^{|U|-1} (k_1 - i) \prod_{i=0}^{l-|U|-1} (n - i) \leq \frac{1}{\prod_{i=0}^{l-1} N - i} \prod_{i=0}^{l/2-1} (k_1 - i) \prod_{i=0}^{l/2-1} (n - i) \\
&= \prod_{i=0}^{l/2-1} \frac{(k_1 - i)(n - i)}{(N - 2i)(N - 2i - 1)} = \prod_{i=0}^{l/2-1} \frac{(k_1 - i)(n - i)}{(N - 2i)(N - 2i)} \prod_{i=0}^{l/2-1} \frac{N - 2i}{N - 2i - 1}.
\end{aligned} \tag{15}$$

The second factor in (15) can be bounded by

$$\prod_{i=0}^{l/2-1} \frac{N-2i}{N-2i-1} \leq \prod_{i=0}^{l/2-1} \frac{N-2i+1}{N-2i-1} = \frac{N+1}{N-l+1} \leq N+1 \; .$$

In order to bound the components of the first factor, first notice that

$$\frac{k_1-i}{N-2i} + \frac{n-i}{N-2i} = \frac{N-2i}{N-2i} = 1,$$

so we multiply two numbers that sum to one. Furthermore, $\frac{k_1-i}{N-2i} \leq \frac{2N/5-i}{N-2i} = \frac{2}{5}\frac{N-5i/2}{N-2i} \leq \frac{2}{5}$. Recall that we assumed that $\frac{k_1}{N} \leq \frac{2}{5}$, which is used in the first inequality. The product of $z(1-z)$ with $z \in \left[0, \frac{2}{5}\right]$ is maximized for $z = \frac{2}{5}$. Thus we already have

$$\prod_{i=0}^{l/2-i} \frac{(k_1-i)(n-i)}{(N-2i)(N-2i)} \leq \left(\frac{2}{5}\cdot\frac{3}{5}\right)^{l/2} \leq 0.24^{l/2}.$$

Inserting this into (15) we get that for all $U \subseteq \{1,\dots,l\}$ with $|U| \geq l/2$

$$\mathbb{P}\left(\text{The } i\text{-th draw is red if and only if } i \in U\right) \leq (N+1)0.24^{l/2}.$$

Recall that we have $l \geq m/8$ and, due $k_i \leq m/3 + m^{2/3}$ for $i \in \{1,2,3\}$, $N+1 = k_1+k_2+k_3+1 \leq 2m$ for m large enough. By a union bound, we thus get

$$\begin{aligned}\mathbb{P}\left(\tilde{K}_1 \geq \frac{l}{2}\right) &= \sum_{\substack{U\subseteq\{1,\dots,l\}:\\ |U|\geq l/2}} \mathbb{P}\left(\text{The } i\text{-th draw is red if and only if } i \in U\right) \leq 2^l(N+1)0.24^{l/2} \\ &\leq 2m\left(2\sqrt{0.24}\right)^l \leq 2m\,0.98^l \leq 2m\,0.98^{m/8}.\end{aligned}$$

□

C Auxiliary Statements

In this section we give several auxiliary statements that we used above.

The first lemma is an auxiliary inequality that follows from the inequality $1 + x \leq e^x$ with $x = -\frac{c}{m}$.

Lemma C.1. *Let $c \geq 0$ and $m > c$. Then,*

$$\left(1 - \frac{c}{m}\right)^m \leq e^{-c}. \tag{16}$$

An application of Chernoff's inequality (see, e.g., Hagerup and Rüb, 1990) yields the next result.

Lemma C.2. *Let X be the sum of independent Bernoulli random variables or a Poisson random variable with expectation value μ and let $0 < \delta \leq 1$. Then,*

$$\mathbb{P}\left(X \geq (1+\delta)\mu\right) \leq e^{-\mu\delta^2/3}, \tag{17}$$

$$\mathbb{P}\left(X \leq (1-\delta)\mu\right) \leq e^{-\mu\delta^2/3}. \tag{18}$$

The next lemma is a typical application of Markov chains to analyze random walks with drift, and its proof uses a classical technique.

Lemma C.3. *Let $0 < \varepsilon < \frac{1}{2}$ and let $(X_n)_{n\in\mathbb{N}}$ be an irreducible Markov chain with state space $\mathbb{N}$, that makes jumps between nearest neighbours only and with transition probabilities satisfying $p(z, z+1) \geq \frac{1}{2} - \varepsilon$ for all $z \geq M$. When the Markov chain starts at $X_0 = M$, the probability of hitting some integer $N > M$ before going to $M - 1$ is bounded by $(1-2\varepsilon)^{N-M}$.*

Proof. We provide a proof for the case $M = 1$. The proof for general M works completely analogous.

We define the function

$$f : \mathbb{N} \to \mathbb{R} \ , \ n \mapsto \sum_{i=0}^{n-1} \prod_{j=1}^{i} \frac{p(j, j-1)}{p(j, j+1)},$$

i.e.,

$$f(0) = 0, f(1) = 1, \text{ and } f(n+1) = f(n) + \prod_{j=1}^{n} \frac{p(j, j-1)}{p(j, j+1)} \text{ for } n \geq 1.$$

The function f is harmonic on $\mathbb{N}_{>0}$ for the transition probabilities $(p(k,l))_{k,l\in\mathbb{N}}$, as

$$\begin{aligned}
&p(n, n+1)f(n+1) + p(n, n-1)f(n-1) \\
&= p(n, n+1) \left(f(n) + \prod_{j=1}^{n} \frac{p(j, j-1)}{p(j, j+1)} \right) + p(n, n-1) \left(f(n) - \prod_{j=1}^{n-1} \frac{p(j, j-1)}{p(j, j+1)} \right) \\
&= f(n) + p(n, n+1) \prod_{j=1}^{n} \frac{p(j, j-1)}{p(j, j+1)} - p(n, n-1) \prod_{j=1}^{n-1} \frac{p(j, j-1)}{p(j, j+1)} = f(n)
\end{aligned}$$

for all $n \geq 1$. For an integer $t \in \mathbb{Z}$, define the stopping time $\tau_t = \min\{n : X_n = t\}$, and define $\tau = \tau_0 \wedge \tau_N = \min\{n : X_n \in \{0, N\}\}$. Assume that the Markov chain starts at $X_0 = 1$. Then the process $f(X_{n\wedge\tau})$ is a martingale with respect to the natural filtration. As the state space is finite, this martingale is also uniformly integrable. By the optional sampling theorem (Klenke, 2013, Theorem 10.21), we have

$$\begin{aligned}
1 &= \mathbb{E}\left[f(X_\tau)|X_0 = 1\right] = f(0) \cdot \mathbb{P}\left(X_\tau = 0|X_0 = 1\right) + f(N) \cdot \mathbb{P}\left(X_\tau = N|X_0 = 1\right) \\
&= f(N) \cdot \mathbb{P}\left(\tau_N < \tau_0|X_0 = 1\right).
\end{aligned} \tag{19}$$

Next, we bound $f(N)$ by

$$\begin{aligned}
f(N) &= \sum_{i=0}^{n-1} \prod_{j=1}^{i} \frac{p(j, j-1)}{p(j, j+1)} \geq \sum_{i=0}^{n-1} \prod_{j=1}^{i} \frac{\frac{1}{2} + \varepsilon}{\frac{1}{2} - \varepsilon} = \sum_{i=0}^{n-1} \left(\frac{1+2\varepsilon}{1-2\varepsilon} \right)^i = \frac{\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^N - 1}{\frac{1+2\varepsilon}{1-2\varepsilon} - 1} = \frac{\left(\frac{1+2\varepsilon}{1-2\varepsilon}\right)^N - 1}{\frac{4\varepsilon}{1-2\varepsilon}} \\
&= \frac{(1+2\varepsilon)^N - (1-2\varepsilon)^N}{4\varepsilon(1-2\varepsilon)^{N-1}} \geq \frac{4\varepsilon}{4\varepsilon(1-2\varepsilon)^{N-1}} = \frac{1}{(1-2\varepsilon)^{N-1}}.
\end{aligned}$$

Inserting this inequality into (19) and solving for $\mathbb{P}\left(\tau_N < \tau_0|X_0 = 1\right)$ gives that

$$\mathbb{P}\left(\tau_N < \tau_0|X_0 = 1\right) \leq (1-2\varepsilon)^{N-1}.$$

□