On the Optimality of Greedy Policies in Dynamic Matching

Süleyman Kerimov, Itai Ashlagi, Itai Gurvich*

June 30, 2023

Abstract

We study centralized dynamic matching markets with finitely many agent types and heterogeneous match values. A network topology describes the pairs of agent types that can form a match and the value generated from each match.

A matching policy is hindsight optimal if the policy can (nearly) maximize the total value simultaneously at all times. We find that suitably designed greedy policies are hindsight optimal in two-way matching networks. This implies that there is essentially no positive externality from having agents waiting to form future matches.

We first show that the greedy longest-queue policy with a minor variation is hindsight optimal. Importantly, the policy is greedy relative to a residual network, which includes only non-redundant matches with respect to the static optimal matching rates. Moreover, when the residual network is acyclic (e.g., as in two-sided networks), we prescribe a greedy static priority policy that is also hindsight optimal. The priority order of this policy is robust to arrival rate perturbations that do not alter the residual network.

Hindsight optimality is closely related to the lengths of type-specific queues. Queue-lengths cannot be smaller (in expectation) than of the order of ϵ^{-1}, where ϵ is the *general position gap* that quantifies the stability in the network. The greedy longest-queue policy achieves this lower bound.

1 Introduction

We study centralized dynamic matching markets with finitely many agent types and heterogeneous match values. Delaying actions to accumulate "inventory" creates a positive externality from

*Kerimov: Jones Graduate School of Business, Rice University, kerimov@rice.edu. Ashlagi: Department of Management Science and Engineering, Stanford University, iashlagi@stanford.edu. Gurvich: Kellogg School of Management, Northwestern University, i-gurvich@kellogg.northwestern.edu. This work is supported by the Department of Defense grant STTR A18B-T007 and by NSF grant CMMI-2010940.

forming future matches that generate high value. This delay, however, inevitably compromises short-term value. The goal of this paper is to shed light on this tension within the family of two-way matching networks.

In our model, agents arrive sequentially to the market. The type of an agent is observed upon arrival and independently drawn from a given distribution over finitely many types; we associate each type with a queue that holds waiting agents of that type. A network topology describes which pairs of agent types can match. We assume that agents leave the market when they are matched. A matching policy determines when and which matches to form.

To evaluate a matching policy and the tension between short- and long-term value, we use the notion of *all-time* regret. The regret at a given time t is measured by the difference between the (expected) total value that can possibly be generated and the (expected) total value generated by the policy until time t. The existence of a policy that achieves a "small" regret *at all times* suggests that the tension between short- and long-term value is essentially moot. We refer to such a policy as *hindsight optimal*.

The networks considered in this paper are two-way (each match includes two agents) and satisfy a *general position condition*. General position is a weak (but necessary) condition that holds when the static-planning problem (a linear program that optimizes the first order matching rates) has a unique and non-degenerate optimal solution.

Optimality of greedy policies. Our main contribution is identifying that for the family of two-way matching networks, *suitably designed* greedy policies are hindsight optimal. This implies that two-way matching networks that satisfy general position are *simple*, in the sense that they can be managed locally (in time) without concern for long-term implications and using intuitive matching rules.

The greedy policies require a minimal preprocessing that includes the removal of all redundant matches from the network. Redundant matches are those that are not used by the static-planning problem. For hindsight optimality, any matching policy must mostly avoid performing redundant matches. Our policies operate on a *residual network* that is obtained from removing these matches from the original network.

An important group of agent types are those that are *under-demanded*. The static-planning problem—and hence any "reasonable" matching policy—matches only a fraction of under-demanded agent types. In every component of a residual network, we show that there is at most one such agent type; and there is exactly one in every acyclic component. These types anchor the policies.

The first greedy policy we prescribe is the *longest-queue* policy with a minor variation. When an agent of a given type arrives and enables possibly multiple feasible (and non-redundant) matches, it is matched to an agent from the longest neighboring queue. One exception is that ties are never broken in favor of an under-demanded agent type. Naturally, the longest-queue policy is a state-dependent policy.

Static (hence state-independent) priority policies are also appealing and common in practice. These policies are also greedy, and ties are broken according to an exogenous priority order. We construct a static priority policy that is hindsight optimal for two-way matching networks, whose residual networks (after removal of all redundant matches) is acyclic. Bipartite networks, which capture two-sided matching markets, fall into this family regardless of whether or not the network is cyclic.

The (static) priority orders that achieve hindsight optimality are easy to describe for each acyclic component, where we refer to the under-demanded agent type as the root. The design rule is that if two matches are on the same path from the root, a higher priority is assigned to the match that is farther away from the root.

Both policies, we show, are locally robust. That is, both policies that operate relative to a (mis)estimated arrival-rate vector remain hindsight optimal as long as this demand vector lies in the same (explicitly characterized) cone as the true arrival-rate vector.

Our findings do not extend to multi-way matching networks. In Kerimov et al. (2021), we studied multi-way matching networks; those where matches may include more than two agent types. In such networks, hindsight optimality is achievable under a periodic clearing policy with a carefully chosen period length. *Greedy* policies that do not wait to form matches are not hindsight optimal. This is because of complementarities that arise in multi-way matching networks. To perform a high-value match that requires multiple types, we must wait for arrivals to multiple queues to be simultaneously non-empty. Greedy policies rush to consume these agents by performing neighboring low-value matches instead; see Example 3.2 in Kerimov et al. (2021). It is important that for two-way matching networks, there exist greedy policies that achieve hindsight optimality: in these, waiting is not necessary for hindsight optimality. From a mathematical standpoint, two-way networks allow us to express the optimality gap explicitly in terms of the network parameters.

Our proofs are based on inferring bounds on regret from bounds on queue-lengths. This creates an intimate connection between the optimal scaling for regret and the optimal scaling for queue-lengths as a function of the network primitives.

In a classic single-server queueing system with utilization ρ, the stationary queue-length is, in expectation, proportional to $1/(1-\rho)$. It is generally true that one cannot achieve smaller stationary queue-length than $1/(1-\rho)$ as long as there is some stochasticity in the arrival rates or service times. In general networks, ρ is *the network utilization* and typically identified via a deterministic static-planning problem akin to the one we use in this paper.

Matching networks like the one we study in this paper are fundamentally different. In the single-server queue, capacity is "wasted" if there are no customers. In our matching networks, capacity can be "inventoried". An arriving agent that finds all queues empty will wait to be matched later. Nevertheless, we find such a fundamental lower bound on queue-length. In Kerimov et al. (2021), we proved that (except for trivial networks), the long-run average queue-length is at least of the order of ϵ^{-1}, where ϵ is the *general position gap*—a parameter that arises naturally from our matching version of the static-planning problem. In this paper, we establish that the greedy longest-queue policy achieves this lower bound at all times in two-way matching networks. Our narrower focus also facilitates crisper regret bounds. That is, we are able to identify the constant before the optimal scaling ϵ^{-1} and, in turn, reflect its dependence on the number of agent types in the network. From a technical/analysis perspective, there is some unavoidable overlap with our earlier paper. The key mathematical ingredient we import from our earlier paper is the aforementioned connection between queue-lengths and regret. Once that connection is made, however, the current paper—because of the algorithms/policies that did not appear in our earlier paper—requires its own separate analysis to bound the queue-lengths. This analysis draws on both new graph-related results and stochastic (Lyapunov function) results. The proof that the static priority policy achieves hindsight optimality requires recursively creating a Lyapunov function, which might be of independent interest.

Beyond anchoring the policy construction (through the under-demanded agent types) and revealing the optimal scaling for regret, the static-planning problem plays a central role in our analysis of the stochastic system. The Lyapunov functions that we construct use our explicit characterization of the optimal solution of the static-planning problem. Our Lyapunov function arguments for longest-queue optimality is simpler than analogous proofs for capacitated queueing networks, and stability of general class of max-weight policies (e.g., see Jonckheere et al. (2022)). As such, it serves to introduce methods from the queueing theory toolbox to a broader (non-queueing) community studying matching networks.

1.1 Related literature

Two streams of literature are closely related to this work. The first stream concerns matching in random graphs, and the other stream concerns matching in queueing systems.

Matching in random graphs. This literature studies random graphs, where agents arrive over time and form an edge with existing agents in the system with some exogenous probabilities. A large subset of this literature focus on matching, where all matches generate the same value in contrast to our heterogeneous setting. Several studies find greedy policies to be asymptotically optimal (either when the matching probability vanishes or when the arrival rates grow large) when the objective is to minimize waiting times (Anderson et al., 2017; Ashlagi et al., 2019a), or, in the presence of departures, when the objective is to maximize the number of matches (Akbarpour et al., 2020) (unless departure times are observable), or, under both measures (Ashlagi et al., 2019b).

In our paper, we consider heterogeneous match values so that greediness alone does not specify the policy completely. One must specify which match to perform when multiple matches are available. Still, we are able to show that—in two-way matching networks—being greedy with respect to a preoptimized network is hindsight optimal at all times (and truly optimal in the long-run average sense). The network being two-way is a necessary condition for hindsight optimality of greedy policies in our case, as greedy policies are suboptimal in general networks (Kerimov et al., 2021).

Several papers study dynamic matching in two-sided networks with heterogeneous match values and departures. Blanchet et al. (2020) consider a model, in which match values are generated from a continuous distribution (in contrast to our finite setting). The paper finds that greedy threshold policies, which assure that the market is sufficiently thick, are (nearly) asymptotically optimal as the market grows large. Collina et al. (2020) interpolate between immediate and delayed actions in order to achieve an approximation guarantee.

Matching in queues. Intuitively speaking, agents waiting in queues at a given time correspond to match values that have not yet been realized. Achieving the optimal scaling for regret, as a function of the general position gap ϵ, is thus intimately linked to the minimal achievable queue-length scaling. Within our analysis, we establish that ϵ^{-1} is the minimal scaling for queue-length and it is achievable.

This question of minimal stationary queue-length scaling has a long history in the capacitated queueing networks literature. In the simplest of these—the single station single server queue—the stationary delay is of the order of $1/(1-\rho)$, where ρ is the utilization. The single server is "perfectly efficient", since the server idles only when there is no work. In more general networks, in contrast, some servers might be idle and have nothing to work on although there is work (somewhere else) in the network. The natural question is then whether there is a centralized policy, under which a scaling of $1/(1-\rho)$ is achievable. In capacitated queueing networks, ρ is identified by a static planning problem; see e.g. Harrison and Lopez (1999). Shah et al. (2014); Maguluri and Srikant (2016) show that max-weight policies are those that achieve the optimal scaling. In matching networks, agents play the dual role of demand and capacity. Nevertheless, a static matching problem characterizes the general position gap and, in turn, the optimal stationary queue-length scaling. We prove that a suitably defined longest-queue policy (an instance of max-weight) achieves the optimal stationary queue-length scaling.

Stationarity is, itself, non-obvious in matching models. Networks of the type we study here—where arrivals are sequential—are generally unstable. Conditions on the model and primitives that guarantee stability are studied in, for example, Mairesse and Moyal (2016); Bušić et al. (2013). The control policy matters for stability. It is sometimes fixed—as in the growing body of work on the stability of First-Come-First-Serve two-way matching networks (e.g., Adan et al. (2018)) and sometimes chosen explicitly to stabilize the network Jonckheere et al. (2022). It is a by-product of our analysis that with sequential arrivals, the network is stabilizable as long as the static planning problem solution induces a residual network with odd cycles (which makes the network, in particular, non-bipartite).

The minimization of holding costs has been studied in the literature; see, e.g., Bušić and Meyn (2015); Gurvich and Ward (2014); Cadas et al. (2019) that focus on holding-cost minimization. Our focus is on match value maximization, similar to Nazari and Stolyar (2019). The goals, however, are different. Nazari and Stolyar (2019) develops a policy that maximizes the long-run average value while stabilizing the queues. They are able to do so *without* knowing the arrival rates in advance; see also Aveklouris et al. (2021). Instead, we assume that arrival rates are known, and we focus on all-time regret—a stronger notion than long-run average optimality—and its scaling.

This paper is a follow-up to Kerimov et al. (2021), where we studied multi-way matching networks and proposed a batching policy that achieves the minimal all-time regret scaling. In general, we show there, that acting greedily is suboptimal. In this paper we show that when restricting attention to two-way matching networks, there exist greedy policies that can achieve the optimal all-time regret scaling. The restriction to two-way networks allows for a more explicit characterization of the regret bound in terms of the network parameters. Finally, we show how the network structure can be used to define a static priority policy that achieves constant regret. Of independent interest may be our recursive construction of a Lyapunov function for the static priority policy.

In our paper, we prove that—in two-way matching networks—one can achieve constant regret while being truly greedy. Under our policies, matches are performed as long as there is at least one feasible match available. The only choice is which of the multiple feasible matches to perform. A subsequent work Gupta (2022) introduces a weaker version of greediness, where the policy commits an item to a match upon arrival even if that match is not available at that point in time. This policy might leave items in queues even when there are matches available, but the weaker definition supports near optimality beyond two-way networks. Importantly, the policies are structurally different.

Notation. For real numbers x and y, we use $x \wedge y = \min\{x, y\}$. We use $[n]$ to denote the set of integers $\{1, 2, \ldots, n\}$. We follow the accepted meaning of little o, big $\mathcal{O}$ and big Ω. For example, $a_t = \Omega(b_t)$ for all $t > 0$ (for non-negative a_t, b_t) means that $\liminf_{t\to\infty} a_t/b_t > 0$. Missing proofs in the body of the paper appear in the appendix.

2 Model

Matching network. There is a finite set of *agent types* $\mathcal{A} = \{1, 2, \ldots, n\}$, a finite set of *matches* $\mathcal{M} = \{1, \ldots, d\}$, and a *match value* $r_m > 0$ for each match $m \in \mathcal{M}$. Each match $m \in \mathcal{M}$ is characterized by *two* participating agent types, denoted by the set $\mathcal{A}(m)$. The *network topology* is specified by a *matching matrix* $M \in \{0, 1\}^{n\times d}$, where $M_{im} = 1$ if and only if $i \in \mathcal{A}(m)$. There is no harm in assuming that each agent type participates in at least one match. Each agent type $i \in \mathcal{A}$ is associated with an *arrival probability* $\lambda_i > 0$; $\sum_{i\in\mathcal{A}} \lambda_i = 1$. We refer to the tuple $\mathcal{G} = (M, \lambda, r)$ as the *matching network*.

The matching network induces a weighted undirected simple graph, where the set of vertices is $\mathcal{A}$ and the set of edges is $\mathcal{M}$: there is an edge between $i, j \in \mathcal{A}$ with weight r_m if and only if there exists $m \in \mathcal{M}$ such that $\mathcal{A}(m) = \{i, j\}$. We say that $i, j \in \mathcal{A}$ are *neighbors* if $\mathcal{A}(m) = \{i, j\}$ for some $m \in \mathcal{M}$. With slight abuse of notation, we denote this induced simple graph also by $\mathcal{G}$. We assume without loss of generality that $\mathcal{G}$ is connected.

Dynamics. Time is discrete, and there is a single agent arrival every period. The arriving agent is of type $i \in \mathcal{A}$ with probability λ_i. We maintain a separate queue for each agent type, and agents join their type-dedicated queues upon arrival. All queues are empty at time $t = 0$.

Match $m \in \mathcal{M}$ is *available* at time t if and only if the queues of both agent types in $\mathcal{A}(m)$ are non-empty at that time. Performing $m \in \mathcal{M}$ once requires one agent from each type in $\mathcal{A}(m)$ and generates a value of r_m. Matched agents leave the market immediately.

The process A_i^t counts the number of arrivals to queue $i \in \mathcal{A}$ until (and including) time t. The sequence of events in a time period is: an agent arrival is realized, then matches are performed, and queue-lengths are updated. The process Q_i^t tracks the number of agents waiting in queue $i \in \mathcal{A}$ at time t, *after* all matches for this period have been performed.

Matching policy. A matching *policy* is a mapping from histories of arrivals and performed matches to a (possibly empty) set of matches. Given the history, the matching policy determines how many times each match is performed at each time period. An *admissible* matching policy is an increasing non-anticipative process $D^t := (D_m^t : m \in \mathcal{M}, t \geq 0)$, where D_m^t is the number of times match $m \in \mathcal{M}$ is performed by time t; D^t must satisfy

$$Q^t = A^t - MD^t \text{ for all } t \geq 0. \tag{1}$$

We assume that D^t is right-continuous with left limits (RCLL). $\Delta D_m^t := D_m^t - D_m^{t-1}$ is then the number of times match $m \in \mathcal{M}$ is performed at time $t > 0$. We add the superscript D on expectations to make explicit the dependence on the policy, where the superscript is omitted when the context is clear. The family of all admissible matching policies is denoted by Π.

Greedy policies are a large family of admissible policies. These policies perform, whenever possible, a match among those available within a prespecified set.

Definition 2.1 (greedy policy)**.** *Given a matching network $\mathcal{G}$ and a subset $\mathcal{S} \subseteq \mathcal{M}$ (not necessarily strict), we say that a policy D is a* greedy policy *with respect to $\mathcal{S}$, if*

(i) a match is performed whenever at least one match becomes available to perform in $\mathcal{S}$, and

(ii) matches in $\mathcal{M} \backslash \mathcal{S}$ are never performed, i.e., $D_m^t = 0$ for all $m \in \mathcal{M} \backslash \mathcal{S}$ and for all $t \geq 0$.

Note that under any greedy policy, at most one match can be performed at any time period, since exactly one agent arrives every period. Note that when a match is performed, it must include the arriving agent. Definition 2.1 does not specify which match to perform when multiple matches are available, which can happen upon an agent arrival. Which available match to perform remains as a degree of freedom in the policy definition. This choice will differ between two greedy policies that we will introduce.

Defining a greedy policy relative to a set $\mathcal{S}$, that could be $\mathcal{M}$ or a strict subset thereof, gives us flexibility. Our proposed policies will be greedy relative to a strict subset of $\mathcal{M}$.

Optimality criterion. The expected *total value* generated by time t under a policy D is given by

$$\mathcal{R}^{D,t} := \mathbb{E}^D[r \cdot D^t].$$

For any *fixed* t, the *optimal value* $\mathcal{R}^{*,t} := \max_{D \in \Pi} \mathcal{R}^{D,t}$ is trivially attained by the policy, which takes no action until time t and follows an optimal (static) weighted matching at time t. That is,

$$\mathcal{R}^{*,t} := \mathbb{E} \left[\begin{array}{ll} \max & r \cdot y \\ \text{s.t.} & My \leq A^t \\ & y \in \mathbb{Z}^d_{\geq 0} \end{array} \right],$$

where the expectation is taken over all realizations of A^t.

The function $\mathcal{R}^{*,t}$ can be interpreted as the *hindsight upper bound* at time t, i.e., the decision maker is allowed to correct past decisions so that previously performed matches may be revoked to perform new ones at all times. A matching policy is *hindsight optimal* if it is, *at all times*, *almost* as good as the optimal value.

Definition 2.2 (hindsight optimality)**.** *A matching policy D is* hindsight optimal *if*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \mathcal{O}(1) \textit{ for all } t > 0,$$

which implies, in particular, $\mathcal{R}^{D,t}/\mathcal{R}^{*,t} = 1 - \mathcal{O}(1/t)$ for all $t > 0$.

The existence of a hindsight optimal matching policy means that the tension between short- and long-term objectives is essentially moot; a good performance at time t_0 does not necessitate a significant compromise at time $t_1 > t_0$. Observe that a hindsight optimal matching policy is also optimal in the long-run average sense:

$$\frac{\mathcal{R}^{*,T} - \mathcal{R}^{D,T}}{\mathcal{R}^{*,T}} = \mathcal{O}(1/T) \to 0 \text{ as } T \to \infty. \tag{2}$$

3 Main results

We identify two greedy policies that are hindsight optimal, one is state-dependent and the other one is state-independent. The policies (and their analyses) use properties of the optimal solution of a static (offline) linear matching problem.

3.1 Preliminaries

We begin with some preliminaries before presenting our main results.

Static-planning and general position. Relaxing the integrality constraints and applying Jensen's inequality gives the following upper bound on $\mathcal{R}^{*,t}$:

$$\mathcal{R}^{*,t} = \mathbb{E}\begin{bmatrix} \max & r\cdot y \\ \text{s.t.} & My \le A^t \\ & y \in \mathbb{Z}^d_{\ge 0} \end{bmatrix} \le \begin{array}{ll} \max & r\cdot x \\ \text{s.t.} & Mx \le \lambda t \\ & x \in \mathbb{R}^d_{\ge 0}. \end{array}$$

With the change of variables $z = x/t$, we can write the upper bound in standard form as follows:

$$\begin{array}{ll} \max & r\cdot z \\ \text{s.t.} & Mz + s = \lambda \\ & z \in \mathbb{R}^d_{\ge 0}, s \in \mathbb{R}^n_{\ge 0}. \end{array} \tag{SPP}$$

We refer to this formulation as the *static-planning problem* (SPP). The following definition introduces the notion of *general position* that captures the level of stability in a matching network and plays a crucial role in our main results. In fact, general position is a necessary condition to achieve hindsight optimality (Kerimov et al., 2021, Example 3.1).

Definition 3.1 (general position)**.** *A matching network $\mathcal{G}$ satisfies the* general position *condition* (**GP**) *if* (SPP) *has a unique non-degenerate optimal solution* (z^*, s^*), *i.e., all n basic variables in this solution are strictly positive. Define the sets*

$$\mathcal{M}_+ := \{m \in \mathcal{M} : z^*_m > 0\},\ \mathcal{M}_0 := \mathcal{M}\backslash\mathcal{M}_+,\ \mathcal{Q}_+ := \{j \in \mathcal{A} : s^*_j > 0\} \textit{ and } \mathcal{Q}_0 := \mathcal{A}\backslash\mathcal{Q}_+,$$

where $\mathcal{M}_+$ *is the set of* active *matches,* $\mathcal{M}_0$ *is the set of* redundant *matches,* $\mathcal{Q}_+$ *is the set of* under-demanded (non-empty) *queues, and* $\mathcal{Q}_0$ *is the set of* over-demanded (empty) *queues. The* general position gap *is defined as*

$$\epsilon := \min_{m \in \mathcal{M}_+} z_m^* \wedge \min_{j \in \mathcal{Q}_+} s_j^*.$$

Residual graph. To achieve hindsight optimality, any matching policy must mostly avoid performing redundant matches (Kerimov et al., 2021, Remark 3.1). Accordingly, the policies that we will propose are greedy with respect to the set $\mathcal{S} = \mathcal{M}_+ \subsetneq \mathcal{M}$. Let $\mathcal{G}' := \mathcal{G} - \mathcal{M}_0$ be the (SPP)-*residual graph*, which is obtained from $\mathcal{G}$ by removing all redundant matches (every $m \in \mathcal{M}$ with $z_m^* = 0$). The (SPP)-residual graph $\mathcal{G}'$ is then a union of (possibly) multiple components, and we write $\mathcal{G}' = \cup_{k \in [K]} \mathcal{C}_k$, where $\mathcal{C}_k$ is the k^{th} component of $\mathcal{G}'$. Since $\mathcal{G}$ is a simple graph, any edge (match) removal can increase the number of components at most by 1; $K \leq |\mathcal{M}_0| + 1$. Let $\mathcal{A}(\mathcal{C}_k)$ be the set of all vertices (queues) in $\mathcal{C}_k$, and let $\mathcal{M}(\mathcal{C}_k)$ be the set of all edges (matches) in $\mathcal{C}_k$ for all $k \in [K]$.

The (SPP)-residual graph $\mathcal{G}'$ has some useful properties, which will be crucial in the design and analysis of our policies.

Lemma 3.1. *Assume that* $\mathcal{G}$ *satisfies* **GP**. *Then each component* $\mathcal{C}_k$*,* $k \in [K]$*, of the* (SPP)*–residual graph* $\mathcal{G}'$ *satisfies the following properties: (i)* $\mathcal{C}_k$ *contains at most one cycle, (ii) if* $\mathcal{C}_k$ *does not contain a cycle, then* $\mathcal{C}_k$ *is a tree and* $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 1$*, and (iii) if* $\mathcal{C}_k$ *contains a cycle, then the cycle is of odd length and* $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 0$.

The proof of Lemma 3.1 appears in Appendix A and uses simple linear programming variable-counting arguments. This lemma will be crucial for the construction of our static priority policy. When each component of $\mathcal{G}'$ is a tree, the single under-demanded queue will serve as an anchor in determining the priority order over matches. Informally, the priority order of a match will be proportional to the distance from this queue.

3.2 Optimality of the longest-queue policy

Recall that that under a greedy policy, in the sense of Definition 2.1, a match can be performed only upon an arrival of an agent. The longest-queue policy is a greedy state-dependent policy defined as follows.

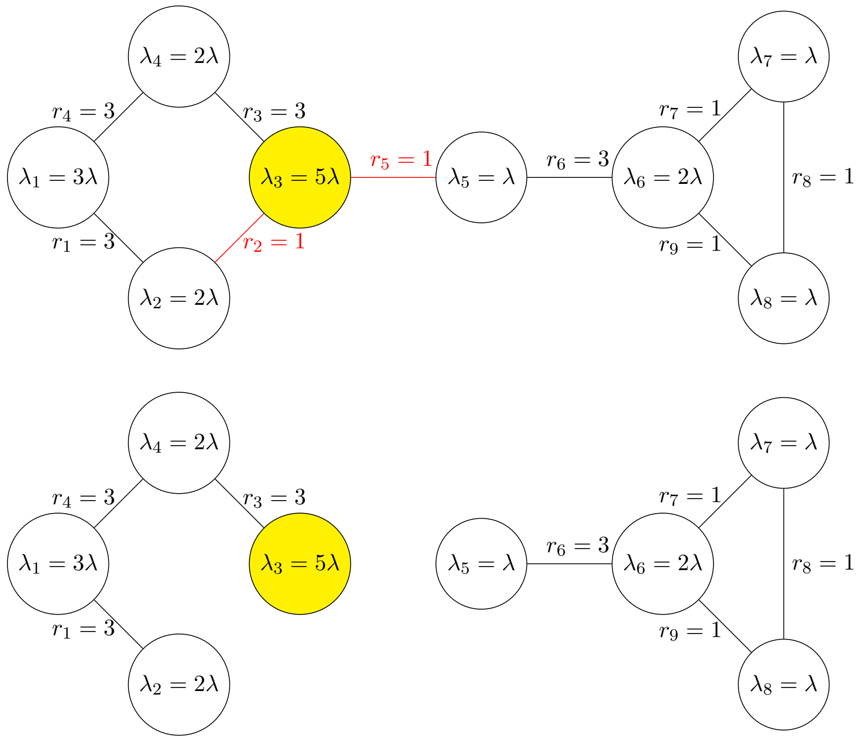

Figure 1: In all figures, arrival probabilities and match values are indicated on vertices and edges, respectively. (TOP) A matching network that satisfies **GP** with $\mathcal{M}_0 = \{2, 5\}$ (red edges) and $\mathcal{Q}_+ = \{3\}$ (yellow vertex). The scalar λ is chosen so that $\sum_{i \in \mathcal{A}} \lambda_i = 1$. (BOTTOM) The corresponding (SPP)–residual graph $\mathcal{G}'$.

Definition 3.2 (longest-queue policy)**.** *Given a matching network $\mathcal{G}$, the longest-queue policy, denoted by $LQ(\mathcal{M}_+)$, is a greedy policy with respect to $\mathcal{M}_+$ such that*

(i) At any time $t > 0$, upon arrival of an agent (say type$-i$), perform the available match $m \in \mathcal{M}_+$ such that $A(m) = \{i, j\}$ and $j \in \arg\max\{Q_k^t : A(m') = \{i, k\}$ for some $m' \in \mathcal{M}_+\}$, where ties are broken arbitrarily, and

(ii) at the end of each time period (after a match is performed), all agents of types $i \in \mathcal{Q}_+$ leave the market unmatched.

Upon arrival, the arriving agent is matched to an agent in a neighboring queue (given that there is a non-empty one) that contains the greatest number of agents. Consider, for example, the cyclic component in Figure 1(BOTTOM). Suppose that $Q_7^t > Q_5^t > 0$ at some time t, and there is an arrival to queue-6 next. Matches 6 and 7 can both be performed upon arrival but match-7 will be performed because $Q_7^t > Q_5^t > 0$. If $Q_7^t = Q_5^t > 0$, the choice between performing match-6 and match-7 is arbitrary.

In Definition 3.2, all agents of type i, $i \in \mathcal{Q}_+$, are "rejected" at the conclusion of a period. In particular, an agent of type $i \in \mathcal{Q}_+$ can only be matched upon arrival, which happens if one of its neighboring queues is non-empty. These rejections simplify our analysis because we do not have to keep track of the number of agents in queues in the set $\mathcal{Q}_+$. The process $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$ is itself a Markov chain and, we show, stable under our policies.

Our analysis reveals that such rejections do not sacrifice optimality, and it is practically reasonable to reject these agents, otherwise corresponding queues will grow without a bound; see Lemma 5.1.

Our first result is that $LQ(\mathcal{M}_+)$ is hindsight optimal with a ϵ^{-1} regret scaling. This, with the exception of trivial cases, is also a lower bound on regret scaling; see (Kerimov et al., 2021)[Theorem 3.1].

Theorem 3.1 (hindsight optimality for two-way matching networks)**.** *Assume that $\mathcal{G}$ satisfies* **GP***, and let ϵ be the* **GP** *gap. Then $LQ(\mathcal{M}_+)$ is hindsight optimal:*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \leq \frac{r_{\max} n}{\epsilon} \left(1 + \lambda_{\min}^{-1} \mathbb{1}\{t \leq \frac{n}{\epsilon \lambda_{\min}}\} \right),$$

where $r_{\max} = \max_{m \in \mathcal{M}_+} r_m$ and $\lambda_{\min} = \min_{i \in \mathcal{Q}_0} \lambda_i$.

3.3 Optimality of a static priority (state-independent) policy in bipartite matching networks

We are also interested in greedy policies that follow a static priority order over non-redundant matches, and in particular, make decisions independent of the state of the network. We will establish that there exists a hindsight optimal static priority policy, given that **GP** is satisfied and any component in the (SPP)-residual graph $\mathcal{G}'$ is a tree; an important family of matching networks satisfying this condition is *bipartite* matching networks.

Definition 3.3 (static priority policy)**.** *Given a matching network $\mathcal{G}$, the static priority policy, denoted by $SP(\mathcal{M}_+, p)$, is a greedy policy with respect to $\mathcal{M}_+$ such that*

(i) $p : \mathcal{M}_+ \to \{1, \dots, |\mathcal{M}_+|\}$ is a bijective static priority order. We say that $m \in \mathcal{M}_+$ has a higher priority than $m' \in \mathcal{M}_+$ if and only if $p(m) < p(m')$,

(ii) at any time $t > 0$, upon arrival of an agent (say type$-i$), perform the highest priority match $m \in \mathcal{M}_+$ among those available, where $m \in \arg\min\{p(m') : i \in A(m')\}$, and

(iii) at the end of each time period (after a match is performed), all agents of type−i, $i \in \mathcal{Q}_+$, leave the market unmatched.

Determining the static priority order $p(\cdot)$. Assume that any component $\mathcal{C}_k$, $k \in [K]$, in the (SPP)-residual graph $\mathcal{G}'$ is a tree. Fix some $\mathcal{C}_k$, where the following procedure is applied on each component separately. Per Lemma 3.1, there is a unique queue, denoted by k_+, such that $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+ = \{k_+\}$. We say that $p(\cdot)$ is a *topological order* if given any path starting from k_+ to any $i \in \mathcal{A}(\mathcal{C}_k) \backslash \{k_+\}$, for any two matches (edges) on this path $m, m' \in \mathcal{M}(\mathcal{C}_k)$, we have $p(m) < p(m')$ if and only if m is farther away from k_+ than m' when we traverse the path starting from k_+ to i. Note that there is at least one topological order $p(\cdot)$, as the path between k_+ to any $i \in \mathcal{A}(\mathcal{C}_k) \backslash \{k_+\}$ is unique, since $\mathcal{C}_k$ is an acyclic component.

Our second result shows that there exists a static priority policy, which is hindsight optimal.

Theorem 3.2 (hindsight optimality for two-way matching networks with acyclic residual graphs)**.** *Assume that $\mathcal{G}$ satisfies* **GP***, and any component in the* (SPP)*-residual graph $\mathcal{G}'$ is a tree. Then $SP(\mathcal{M}_+, p)$, where p is any topological order, is hindsight optimal:*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \leq \Gamma \text{ for all } t > 0,$$

where $\Gamma > 0$ is a constant that does not depend on t.

We conclude this section with some remarks regarding Theorems 3.1 and 3.2.

Remark 3.1.

(i) Theorem 3.2 holds for any bipartite matching network $\mathcal{G}$ that satisfies **GP**. This is because per Lemma 3.1, any component $\mathcal{C}_k$ in the (SPP)–*residual graph* $\mathcal{G}'$ is a tree, as $\mathcal{G}$ does not contain any odd cycles.

(ii) In contrast to Theorem 3.1, where we identified the dependence of the constant regret on the general position gap ϵ, in Theorem 3.2 we could only establish that the *regret is constant at all times*. However, this result still implies that $SP(\mathcal{M}_+, p)$ is optimal in the *long-run average sense*; see (2). We believe that $SP(\mathcal{M}_+, p)$ also achieves the optimal scaling of ϵ^{-1} for regret, as simulations in Example 6.1 suggest.

(iii) We do not know whether there exists a hindsight optimal static priority policy, when there is a component $\mathcal{C}_k$ in the (SPP)-residual graph $\mathcal{G}'$ that contains an odd cycle. Note that the tree structure is central to the design of a priority policy.

(iv) The topological order for the static priority policy is generally not unique (unless $\mathcal{G}$ is a path). For example in Figure 2, another possible topological order is $5 \succ 1 \succ 4 \succ 2 \succ 3 \succ 6 \succ 7$.

■

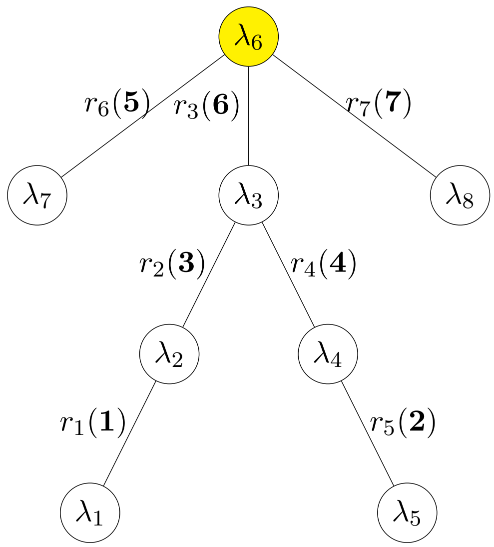

Figure 2: A tree component $\mathcal{C}_k$ with $k_+ = 6$. One possible topological order $p(\cdot)$ is indicated on the matches: $1 \succ 5 \succ 2 \succ 4 \succ 6 \succ 3 \succ 7$, where $m \succ m'$ means $p(m) < p(m')$.

The rest of the paper is organized as follows. In §4, we explicitly characterize the optimal solution of (SPP), which plays a key role in the design and analysis of our matching policies. In §5, we prove hindsight optimality of our matching policies. Finally in §6, we provide numerical examples to provide further insights about our matching policies.

4 Properties of the (deterministic) static-planning problem (SPP)

This section uncovers properties of the matching network and the static-planning problem (SPP), which are essential in the design and (stochastic) analysis of our matching policies.

The following theorem gives an explicit characterization of the optimal solution of (SPP). The characterization is instrumental as it captures permitted perturbation for λ to maintain the optimal basis in terms of the general position gap ϵ. This is a generalization of (Kerimov et al., 2021, Theorem 4.1) to matching networks with cyclic components.

Theorem 4.1 (explicit optimal solution of (SPP))**.** *Assume that* $\mathcal{G}$ *satisfies* **GP**. *Let* (z^*, s^*) *be the unique non-degenerate optimal solution of* (SPP) *with* $\mathcal{M}_+ = \{m \in \mathcal{M} : z^*_m > 0\}$ *and* $\mathcal{Q}_+ = \{j \in \mathcal{A} : s^*_j > 0\}$. *Then there exist* $|\mathcal{M}_+|$ *vectors* $y^m \in \{-1, -1/2, 0, 1/2, 1\}^n$ *and* $|\mathcal{Q}_+|$ *vectors* $y^j \in \{-1, 0, 1\}^n$ *such that*

$$z^*_m(\lambda) := y^m \lambda > 0 \textit{ for all } m \in \mathcal{M}_+ \textit{ and } s^*_j(\lambda) := y^j \lambda > 0 \textit{ for all } j \in \mathcal{Q}_+.$$

Surplus vectors. The *explicit construction* of the y vectors (surplus vectors) in Theorem 4.1 plays a key role in the design and analysis of our matching policies. The following procedure describes the construction for each of the components in the (SPP)-residual graph $\mathcal{G}'$. Fix a component $\mathcal{C}_k$, be it cyclic or a tree; see Lemma 3.1.

- **Tree components.** First assume that $\mathcal{C}_k$ is a tree. The corresponding surplus vectors are already constructed in (Kerimov et al., 2021, §4), but we repeat it here for completeness. Let k_+ be the unique queue in $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+$ per Lemma 3.1. Let $\mathcal{U}_0 := \{i \in \mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_0 : \sum_{m \in \mathcal{M}_+} M_{im} = 1\}$. This is the set of queues in $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_0$ participating in exactly one non-redundant match; $\mathcal{U}_0$ is a subset of the leaves of $\mathcal{C}_k$. Since $\mathcal{C}_k$ is a tree, $\mathcal{U}_0$ is clearly a non-empty set. For all $i \in \mathcal{U}_0$, we first traverse the unique path between k_+ and i in $\mathcal{C}_k$ starting from k_+. Any edge between $i' \in \mathcal{A}(\mathcal{C}_k)$ and $j' \in \mathcal{A}(\mathcal{C}_k)$ on this path is marked with the direction it is traversed, $i' \to j'$ or $j' \to i'$. Denote the resulting directed graph by $\overrightarrow{\mathcal{C}_k}$; e.g., see Figure 3. We refer to k_+ as the *root* of $\overrightarrow{\mathcal{C}_k}$. Since $\mathcal{C}_k$ is a tree and the root k_+ is unique, every edge in this component is marked with a unique direction.

  For each $i \in \mathcal{A}(\mathcal{C}_k)$, we let $\mathcal{T}_i$ be the subtree rooted at i, where $\mathcal{T}_i$ is the union of all directed paths from i to $j \in \mathcal{U}_0$. Note that $\mathcal{T}_{k_+}$ is $\overrightarrow{\mathcal{C}_k}$ itself. Let $\mathcal{A}(\mathcal{T}_i)$ be the set of queues in $\mathcal{T}_i$. Let $d(i,j)$ be the length of the directed path from $i \in \mathcal{A}(\mathcal{C}_k)$ to $j \in \mathcal{A}(\mathcal{C}_k)$ in $\overrightarrow{\mathcal{C}_k}$ with the convention $d(i,i) = 0$. For each $i \in \mathcal{A}(\mathcal{C}_k)$, we then define the surplus vector $y^i \in \{-1, 0, 1\}^n$ as follows:

$$(y^i)_j = \begin{cases} 0, & \text{if } j \in \mathcal{A} \backslash \mathcal{A}(\mathcal{T}_i), \\ 1, & \text{if } d(i,j) \equiv 0 \pmod 2, \\ -1, & \text{if } d(i,j) \equiv 1 \pmod 2. \end{cases} \tag{3}$$

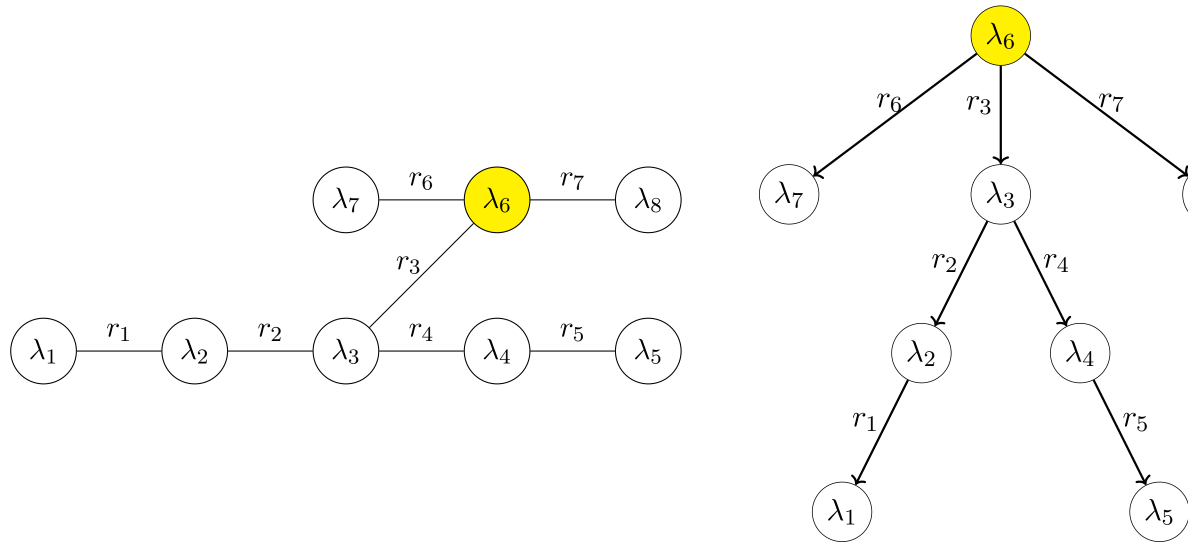

Figure 3: (LEFT) A tree component $\mathcal{C}_k$ with $k_+ = 6$ and $\mathcal{U}_0 = \{1, 5, 7, 8\} \subseteq \mathcal{A}(\mathcal{C}_k)$. (RIGHT) The corresponding directed graph $\overrightarrow{\mathcal{C}_k}$.

Note that since $d(k_+, k_+) = 0$, in particular, we have $(y^{k_+})_{k_+} = 1$. Finally, by the construction of $\overrightarrow{\mathcal{C}_k}$, for each $m \in \mathcal{M}(\mathcal{C}_k)$, there is a unique queue $i(m) \in \mathcal{A}(\mathcal{C}_k)$ such that the marked direction on m is incoming to $i(m)$. Then we define the surplus vector for each $m \in \mathcal{M}(\mathcal{C}_k)$ with the vector $y^{i(m)}$:

$$y^m := y^{i(m)} \text{ for all } m \in \mathcal{M}(\mathcal{C}_k).$$

For example in Figure 3, the surplus vector for queue 3 is $y^3 = [1, -1, 1, -1, 1, 0, 0, 0]$ and the surplus vector for match 2 is equal to the surplus vector for queue 2, which is $y^2 = [-1, 1, 0, 0, 0, 0, 0, 0]$.

- **Cyclic components.** Let us first consider the case when $\mathcal{C}_k$ is just a cycle of odd length. Let $\mathcal{A}(\mathcal{C}_k) = \{1, \ldots, 2n+1\}$ and $\mathcal{M}(\mathcal{C}_k) = \{1, \ldots, 2n+1\}$, where $\mathcal{A}(m) = \{m, m+1\}$ for all $m \in [2n]$ and $\mathcal{A}(2n+1) = \{1, 2n+1\}$. Since $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+ = \emptyset$ per Lemma 3.1, we must have that $\lambda_i = z^*_{i-1} + z^*_i$ for all $2 \leq i \leq 2n+1$ and $\lambda_1 = z^*_{2n+1} + z^*_1$. This yields $\sum_{i=1}^{2n+1} \lambda_i = 2 \sum_{m=1}^{2n+1} z^*_m$, and solving these equations we get

$$z^*_m = \frac{1}{2} \sum_{i=1}^{2n+1} \lambda_i - \Bigg(\sum_{\substack{j < m+1 \\ j \equiv m \pmod 2}} \lambda_j + \sum_{\substack{j > m \\ j \equiv m \pmod 2}} \lambda_j \Bigg) \text{ for all } m \in [2n], \tag{4}$$

$$z^*_{2n+1} = \frac{1}{2} \sum_{i=1}^{2n+1} \lambda_i - \Bigg(\sum_{j \equiv 0 \pmod 2} \lambda_j \Bigg). \tag{5}$$

Using (4) and (5), for any $m \in [2n]$, we define the surplus vector $y^m \in \{-1/2, 0, 1/2\}^{2n+1}$ as follows:

$$(y^m)_j = \begin{cases} -1/2, & \text{if } j < m+1 \text{ and } j \equiv m \pmod 2, \\ -1/2, & \text{if } j > m \text{ and } j \equiv m \pmod 2, \\ 1/2, & \text{otherwise.} \end{cases}$$

Similarly, we define the surplus vector for match $(2n+1)$ as follows:

$$(y^{2n+1})_j = \begin{cases} -1/2, & \text{if } j \equiv 0 \pmod 2, \\ 1/2, & \text{otherwise.} \end{cases}$$

By construction, observe that we have $z_m^* = y^m \lambda$ for all $m \in \mathcal{M}(\mathcal{C}_k)$.

Let us now expand and consider a component $\mathcal{C}_k$ that is cyclic. Per Lemma 3.1, this component contains exactly one cycle, and this cycle is of odd length. Denote this odd cycle by $\mathcal{C}_{\text{odd}}$, and let $\mathcal{A}(\mathcal{C}_{\text{odd}})$ be the set of queues included in this odd cycle. Define the set $\mathcal{U}_{\geq 3} := \{i \in \mathcal{A}(\mathcal{C}_{\text{odd}}) : \deg(i) \geq 3\}$, i.e., the set of queues in $\mathcal{A}(\mathcal{C}_{\text{odd}})$ participating in at least three non-redundant matches in $\mathcal{C}_k$. Fix some $i \in \mathcal{U}_{\geq 3}$. In other words, i participates in at least three non-redundant matches in $\mathcal{C}_k$, and in particular, i participates in at least one non-redundant match that is not a part of $\mathcal{C}_{\text{odd}}$.

Remove both edges (matches) that i participates in $\mathcal{C}_{\text{odd}}$. This decomposes $\mathcal{C}_k$ into two subgraphs and separates i from $\mathcal{C}_{\text{odd}}$. Consider the resulting subgraph that contains i. By construction, this component is a tree. Temporarily, we will assume that i plays the role of k_+ in the construction of the surplus vectors for tree components. Hence, we define the surplus vectors for all matches contained in this subgraph as in the previous construction for tree components, and we let $\bar{y}^i$ be the temporary surplus vector for i under this setting, where i plays the role of k_+. After applying this procedure for all $i \in \mathcal{U}_{\geq 3}$, we define $\bar{\lambda}_i := \bar{y}^i \lambda$, and we construct the remaining surplus vectors for all matches included in $\mathcal{C}_{\text{odd}}$ by replacing λ_i by $\bar{\lambda}^i$ in (4) and (5); e.g., see Figure 4.

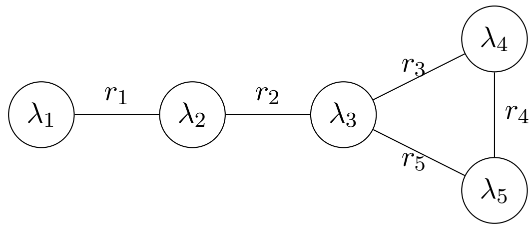

Figure 4: A cyclic component with $\mathcal{U}_{\geq 3} = \{3\}$. We remove match 3 and match 5, and we consider the resulting subgraph that contains queue 3, which is a tree. Our procedure first yields the vectors $y^1 = [1,0,0,0,0]$ and $y^2 = [-1,1,0,0,0]$, for match 1 and match 2, respectively. Then the temporary surplus vector for queue 3 is $\bar{y}^3 = [1,-1,1,0,0]$. Following the procedure, we set $\bar{\lambda}_3 = \bar{y}^3\lambda = \lambda_3 - \lambda_2 + \lambda_1$. Finally solving (4) and (5), where λ_3 is replaced by $\bar{\lambda}_3$, gives $z_3^* = \frac{1}{2}(-\lambda_5+\lambda_4+\lambda_3-\lambda_2+\lambda_1)$, $z_4^* = \frac{1}{2}(\lambda_5+\lambda_4-\lambda_3+\lambda_2-\lambda_1)$ and $z_5^* = \frac{1}{2}(\lambda_5-\lambda_4+\lambda_3-\lambda_2+\lambda_1)$. Hence, the surplus vectors for matches 3, 4 and 5 are $y^3 = [1/2,-1/2,1/2,1/2,-1/2]$, $y^4 = [-1/2,1/2,-1/2,1/2,1/2]$ and $y^5 = [1/2,-1/2,1/2,-1/2,1/2]$, respectively.

The proof of Theorem 4.1 is now immediate.

Proof of Theorem 4.1. Per (Kerimov et al., 2021, Theorem 4.1), for any tree component $\mathcal{C}_k$, we have $z_m^* = y^m\lambda$ for all $m \in \mathcal{M}(\mathcal{C}_k)$ and $s_{k_+}^* = y^{k_+}\lambda$, where k_+ is the unique queue in $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+$. For any cyclic component $\mathcal{C}_k$, we have $\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+ = \emptyset$, and the construction based on (4) and (5) immediately yields $z_m^* = y^m\lambda$ for all $m \in \mathcal{M}(\mathcal{C}_k)$. Finally, strict positivity follows from the assumed non-degeneracy under **GP**. □

We next characterize permitted perturbation for λ in terms of the general position gap ϵ, under which the optimal basis remains the same. By Theorem 4.1, for any other arrival-probability vector $\tilde{\lambda}$, where $\tilde{\lambda}$ is in the cone

$$\mathcal{Y} := \left\{\lambda' \in \mathbb{R}_+^n : y^m\lambda' > 0 \text{ for all } m \in \mathcal{M}_+ \text{ and } y^j\lambda' > 0 \text{ for all } j \in \mathcal{Q}_+\right\},$$

the optimal basis remains the same. Thus, the following is an immediate corollary.

Corollary 4.1 (right-hand side perturbations)**.** *Assume that $\mathcal{G}$ satisfies* **GP**. *Let (z^*, s^*) be the unique non-degenerate optimal solution of* (SPP) *with $\mathcal{M}_+ = \{m \in \mathcal{M} : z_m^* > 0\}$ and $\mathcal{Q}_+ = \{j \in \mathcal{M} : s_j^* > 0\}$. Then the same basis remains optimal for any $\tilde{\lambda} > 0$ such that $\tilde{\lambda} = \lambda + \zeta$, where $\zeta \in \mathbb{R}^n$ satisfies $y^m\zeta \geq -\epsilon$ for all $m \in \mathcal{M}_+$ and $y^j\zeta \geq -\epsilon$ for all $j \in \mathcal{Q}_+$.*

5 Analysis

5.1 Preliminary results

To establish the hindsight optimality of any matching policy, it suffices that all queues in $\mathcal{Q}_0$ remain bounded in expectation. This is shown in the following lemma, which is an analog of the optimality test in Kerimov et al. (2021)[Lemma 4.1].

Lemma 5.1 (optimality test). *Suppose that* **GP** *holds. Let* (z^*, s^*) *be the unique non-degenerate optimal solution of* (SPP). *Suppose the following conditions hold under a policy* D*:*

(i) no agent of type $i \in \mathcal{Q}_0$ *leaves the market unmatched,*

(ii) no matches in $\mathcal{M}_0$ *are performed, i.e.,* $D_m^t = 0$ *for all* $m \in \mathcal{M}_0$ *and for all* $t > 0$*, and*

(iii) $\sum_{i \in \mathcal{Q}_0} \mathbb{E}^D[Q_i^t] \le B$ *for all* $t > 0$*, where* $B > 0$ *is a constant.*

Then D *is hindsight optimal and* $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \le r_{\max} B$ *for all* $t > B\lambda_{\min}^{-1}$*, where* $r_{\max} := \max_{m \in \mathcal{M}_+} r_m$ *and* $\lambda_{\min} := \min_{i \in \mathcal{Q}_0} \lambda_i$*.*

Observe that, by construction, $LQ(\mathcal{M}_+)$ satisfies the first two conditions of Lemma 5.1. We will use Lyapunov function arguments to establish that condition (iii) of Lemma 5.1 holds under our policies. The following is a useful version of a standard tool.

Lemma 5.2 (Glynn and Zeevi (2008), Corollary 4). *Let* $X = (X^t : t \ge 0)$ *be a discrete-time* $\mathcal{S}$*-valued Markov chain with transition kernel* P*, and suppose* $f : \mathcal{S} \to \mathbb{R}$ *is non-negative. If there exists a non-negative function* $g : \mathcal{S} \to \mathbb{R}$ *and a constant* c *for which*

$$\int_S P(x, dy) g(y) - g(x) \le -f(x) + c \text{ for all } x \in \mathcal{S}, \tag{6}$$

then

$$\int_S \pi(dx) f(x) \le c, \tag{7}$$

for any stationary distribution π *of* X*.*

The function g in (6) is a so-called Lyapunov function. As is often the case, the key challenge is to identify suitable functions f and g.

Note that Lemma 5.2 gives a moment bound in stationary distribution. We will be interested in generating moment bounds on the expected size of queues in Q_0 for any time $t > 0$. The following lemma will be useful to establish this. The lemma couples two stochastic systems, one is initialized with $Q^0 = 0$ and the other one is initialized arbitrarily, and relates the total number of agents waiting in both systems at any time t.

The next lemma shows that greedy matching policies are non-expansive. Namely, the gap between two greedy-operated systems—that differs only in their initial queue-lengths—does not grow with time; see Moyal and Perry (2017) for a related result.

Lemma 5.3. *Let D be any greedy policy as in Definition 2.1 such that at the end of each time period (after a match is performed), all agents of type$-i$, $i \in \mathcal{Q}_+$, leave the market unmatched. Let $\mathcal{H}$ be a matching network that is identical to $\mathcal{G}$ except the initialization of the queue-length vector at $t = 0$. Let $(H^t : t \geq 0)$ be the corresponding queue-length vector to $\mathcal{H}$, and consider an arbitrary initialization for $\mathcal{H}$. Let $h := \sum_{i \in \mathcal{Q}_0} H_i^0$. Then under D, we have*

$$\Big| \sum_{i \in \mathcal{Q}_0} Q_i^t - \sum_{i \in \mathcal{Q}_0} H_i^t \Big| \leq h \text{ for all } t \geq 0. \tag{8}$$

Proof of Lemma 5.3. We refer to agents present at $t = 0$ in $\mathcal{H}$ as *labeled.* We also say that a performed match is *labeled* if it contains a labeled agent, and the match is *unlabeled* otherwise.

Let us first prove that $\sum_{i \in \mathcal{Q}_0} H_i^t \leq h + \sum_{i \in \mathcal{Q}_0} Q_i^t$ for all $t \geq 0$. Given the same arrival process, observe that the first ever match in both systems cannot be performed only in $\mathcal{G}$; it is possible that the first ever match will be performed in both systems at the same time. Consider all times when a match is performed in $\mathcal{G}$ but not in $\mathcal{H}$. Consider such a time, say t, and assume that the arriving agent at time t (that makes at least one match available to perform in $\mathcal{G}$) is of type$-i$ and matches to some agent of type$-j$. Since both systems are equipped with the same arrival process and j is not present in $\mathcal{H}$ at time t (otherwise there would be at least one available match to perform in $\mathcal{H}$), this implies that j was already matched in $\mathcal{H}$ at some time $t' < t$. Therefore, at any time $t > 0$, the total number of performed matches in $\mathcal{H}$ is greater than or equal to the total number of performed matches in $\mathcal{G}$. Note that there are only two types of matches, where either an arriving agent type is in $\mathcal{Q}_0$ and matches another agent type in $\mathcal{Q}_0$ that is present in the system, or an arriving agent type is in $\mathcal{Q}_+$ and matches an agent type in $\mathcal{Q}_0$. Thus, if a match is performed in $\mathcal{G}$ or $\mathcal{H}$, then $\sum_{i \in \mathcal{Q}_0} Q_i^t$ or $\sum_{i \in \mathcal{Q}_0} H_i^t$ decreases by 1, respectively. This proves that $\sum_{i \in \mathcal{Q}_0} H_i^t \leq h + \sum_{i \in \mathcal{Q}_0} Q_i^t$ for all $t \geq 0$.

Next we show that $\sum_{i \in \mathcal{Q}_0} Q_i^t \leq h + \sum_{i \in \mathcal{Q}_0} H_i^t$ for all $t \geq 0$. Given the same arrival process, observe that the first ever unlabeled match in both systems cannot be performed only in $\mathcal{H}$. We claim that the total number of performed unlabeled matches in $\mathcal{G}$ is greater than or equal to the total number of performed unlabeled matches in $\mathcal{H}$. Consider a time t such that an unlabeled match is performed in $\mathcal{H}$ (say type-i arrives and matches to type$-j$), but not in $\mathcal{G}$. Similar to the previous arguments, this implies that j was already matched in $\mathcal{H}$ at some time $t' < t$. Since any performed match in $\mathcal{G}$ is unlabeled by definition, this proves the claim. Finally, since one can perform at most h many labeled matches in $\mathcal{H}$ under any arrival process, we must have that $\sum_{i\in\mathcal{Q}_0} Q_i^t \le h + \sum_{i\in\mathcal{Q}_0} H_i^t$. $\square$

5.2 Proof of Theorem 3.1

The proof will apply the optimality test Lemma 5.1. Recall that by construction, $LQ(\mathcal{M}_+)$ satisfies the first two conditions of Lemma 5.1. It remains to prove the third condition, which will be done by leveraging Lyapunov function arguments.

Recall that the (SPP)-residual graph $\mathcal{G}' = \cup_{k\in[K]}\mathcal{C}_k$ consists of components such that for any $k \in [K]$, $\mathcal{C}_k$ is either a tree with $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 1$, or cyclic, in which case $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 0$; see Lemma 3.1. Since our aim is to prove that all queues in $\mathcal{Q}_0$ remain bounded in expectation, we focus then on a single component in our analysis and treat $\mathcal{G}$ as the only component. The analysis is then the same for all the other components. The following is the main ingredient in the proof of Theorem 3.1.

Lemma 5.4. *Assume that $\mathcal{G}$ satisfies* **GP** *with $\mathcal{M}_0 = \emptyset$. Define $\mathcal{L}(Q^t) := \sum_{i\in\mathcal{Q}_0}(Q_i^t)^2,\ t \ge 0$. Then under $LQ(\mathcal{M}_+)$, the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \ge 0)$ is ergodic. Moreover, $\mathcal{L}(Q^t)$ decreases in expectation:*

$$\mathbb{E}[\mathcal{L}(Q^{t+1}) - \mathcal{L}(Q^t) \mid Q^t] \le -2\frac{\epsilon}{n}\|Q^t\|_1 + 1 \text{ for all } t \ge 0. \tag{9}$$

The proof of Lemma 5.4 is given in the next subsection. We first apply it to complete the proof of Theorem 3.1.

Proof of Theorem 3.1. We first prove the upper bound. The drift property (9) in Lemma 5.4 implies that the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \ge 0)$ is positive recurrent; e.g., see (Robert, 2003, Corollary 8.7). Given Lemma 5.4, moment bounds in the steady-state follow trivially from Lemma 5.2, where the functions f and g are $2\frac{\epsilon}{n}\|Q^t\|_1$ and $\mathcal{L}(Q^t)$, respectively. In particular, under the Markov chain's unique stationary distribution, which we denote by π, we have

$$\mathbb{E}_\pi[\|Q^0\|_1] \le \frac{n}{2\epsilon}, \tag{10}$$

where $Q^0 \sim \pi$. Note that we still need to establish a similar moment bound for any time $t > 0$ (not only for the steady-state). Per Lemma 5.3 and (10), we conclude that under $LQ(\mathcal{M}_+)$, we have

$\sum_{i \in \mathcal{Q}_0} \mathbb{E}[Q_i^t] \le n/\epsilon$ for all $t > 0$. Then by Lemma 5.1, we have

$$\mathcal{R}^{*,t} - \mathcal{R}^{LQ(\mathcal{M}_+),t} \le \frac{r_{\max} n}{\epsilon} \text{ for all } t > \frac{n}{\epsilon \lambda_{\min}}.$$

Note that regret is upper bounded by $r_{\max} t$ for any fixed time $t > 0$. Hence, $LQ(\mathcal{M}_+)$ is hindsight optimal as stated.

$\square$

Remark 5.1 (maximizing the number of matched agents in non-bipartite setting)**.** If each component in the (SPP)-residual graph $\mathcal{G}'$ is cyclic, then Theorem 3.1 has an immediate implication for the objective of maximizing the total number of matched agents in the long-run average sense.

A matching policy D matches $(MD^t)_i$ many agents of type i by time t. Recall that $MD^t = A^t - Q^t$ for all $t \ge 0$ per (1) so that the long-run average number of matched agents is given by

$$\liminf_{T\to\infty} \frac{1}{T} \mathbb{E}^D \left[\sum_{i \in \mathcal{A}} A_i^T - \sum_{i \in \mathcal{A}} Q_i^T \right] \ge \sum_{i \in \mathcal{A}} \lambda_i - \limsup_{T \to \infty} \frac{1}{T} \sum_{i \in \mathcal{A}} \mathbb{E}^D[Q_i^T].$$

Within the proof of Theorem 3.1, we showed that under $LQ(\mathcal{M}_+)$, we have $\sum_{i \in \mathcal{Q}_0} \mathbb{E}_\pi[Q_i^0] = \mathcal{O}(\epsilon^{-1})$, where π denotes the steady-state of the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \ge 0)$, and $Q^0 \sim \pi$. In particular, we have $\limsup_{T\to\infty} \frac{1}{T} \sum_{i \in \mathcal{Q}_0} \mathbb{E}[Q_i^T] = 0$ (note that it is sufficient to have $\sum_{i \in \mathcal{Q}_0} \mathbb{E}_\pi[Q_i^0] = \mathcal{O}(1)$ to have this limit).

Since each component of $\mathcal{G}'$ is cyclic, we have $\mathcal{Q}_0 = \mathcal{A}$ per Lemma 3.1 so that $\sum_{i \in \mathcal{A}} \mathbb{E}_\pi[Q_i^0] = \mathcal{O}(\epsilon^{-1})$, and in turn, we have

$$\liminf_{T\to\infty} \frac{1}{T} \mathbb{E} \left[\sum_{i \in \mathcal{A}} A_i^T - \sum_{i \in \mathcal{A}} Q_i^T \right] \ge \sum_{i \in \mathcal{A}} \lambda_i.$$

Since $\sum_{i \in \mathcal{A}} \lambda_i$ is an upper bound on the long-run average number of matched agents, $LQ(\mathcal{M}_+)$ is optimal for this objective. ∎

Remark 5.2 (stability of matching models)**.** If each component of the matching network has odd cycles, then $\mathcal{Q}_+$ is an empty set, and the process $Q(t)$ is (without any rejections) stationary with the expected queue-length bounded as in (10). This stability is consistent with the general stabilizability, proved in Mairesse and Moyal (2016), of dynamic matching in non-bipartite graphs.

∎

5.3 Proof of Lemma 5.4

Throughout this subsection, we simplify the notation by assuming that $q \in \mathbb{Z}^n_{\geq 0}$ is the initial state of the queues ($t = 0$), where there is no available matches to perform. Let Q^1 be the state of the queues at time $t = 1$ after all matches for this period have been performed. We also write the conditional expectation given a matching policy D and the initial state q as $\mathbb{E}^D_q[\cdot]$.

Without loss of generality, we assume that $\mathcal{G}$ satisfies **GP** with $\mathcal{M}_0 = \emptyset$ so that the proof focuses on a single component. We use the following quadratic Lyapunov function

$$\mathcal{L}(q) = \sum_{i \in \mathcal{Q}_0} q_i^2, \quad q \in \mathbb{Z}^n_{\geq 0}.$$

It will be convenient in this subsection to denote match $m \in \mathcal{M}$ sometimes by (i, j), where $\mathcal{A}(m) = \{i, j\}$. Also, it will be convenient to write with some abuse of notation $M_{i,(i,j)} := M_{ij}$.

Finally, we also define $\Delta A^1 := A^1 - A^0$ and $\Delta D^1 := D^1 - D^0$, i.e., the vector that tracks the number of arrivals at time $t = 1$ and the vector that tracks the number of matches performed at time $t = 1$, respectively.

Fix an arbitrary initial state of the queues $q \in \mathbb{Z}^n_{\geq 0}$. Note that under any greedy policy, q must satisfy that for any two queues i and j that are neighbors of each other, we cannot have $q_i, q_j > 0$. In other words, for any $m \in \mathcal{M}$ with $\mathcal{A}(m) = \{i, j\}$, at most one of these queues i or j can be non-empty in q.

The first simple result is generic and applies to any stationary policy. Define the sets

$$\mathcal{U}^+(q) := \{i \in \mathcal{A} : q_i > 0\} \text{ and } \mathcal{U}^0(q) := \mathcal{A} \backslash \mathcal{U}.$$

These are, respectively, the set of non-empty and empty queues in q. We also define

$$\mathcal{M}^+(q) := \{m \in \mathcal{M} : \mathcal{A}(m) \cap \mathcal{U}^+(q) \neq \emptyset\}.$$

This is the set of matches that have a participating non-empty queue in q.

The following proposition is the first step to analyze the one step transition of the quadratic Lyapunov function $\mathcal{L}(q)$. Define $x^D_m(q)$ to be the expected number of times match $m \in \mathcal{M}$ is performed in the first period under a greedy policy D. That is,

$$x^D_m(q) := \mathbb{E}^D_q[\Delta D^1_m].$$

Proposition 5.1. *Under any stationary greedy policy D, we have*

$$\mathbb{E}_q^D[\mathcal{L}(Q^1) - \mathcal{L}(q)] \leq 2q(\lambda - Mx^D(q)) + 1.$$

Moreover, $x_m^D(q)$ must satisfy that

$$x_m^D(q) \in \mathcal{Z}(q) := \left\{ z(q) \in \mathbb{R}_{\geq 0}^d : z_m(q) = 0 \textit{ for all } m \notin \mathcal{M}^+(q) \textit{ and } \sum_{\substack{m \in \mathcal{M}: \\ M_{im}=1}} z_m(q) \leq \lambda_i \textit{ for all } i \in \mathcal{U}^0(q) \right\}. \tag{11}$$

Proof of Proposition 5.1. Since $Q^1 = q + \Delta A^1 - M\Delta D^1$, we have

$$\begin{aligned}\mathcal{L}(Q^1) - \mathcal{L}(q) &= \sum_{i \in \mathcal{Q}_0} (Q_i^1 + q_i)(Q_i^1 - q_i) \\ &= 2q(\Delta A^1 - M\Delta D^1) + (\Delta A^1 - M\Delta D^1)'(\Delta A^1 - M\Delta D^1).\end{aligned}$$

Under any stationary greedy policy, we have $|\Delta D_m^1| \leq 1$ for all $m \in \mathcal{M}$, and since there is a single agent arrival every period, we have $(\Delta A^1 - M\Delta D^1)'(\Delta A^1 - M\Delta D^1) \leq 1$ with probability 1. This proves the first assertion of the proposition after taking expectations.

For the second assertion, first observe that if $m \notin \mathcal{M}^+(q)$, since there is a single agent arrival every period, m is unavailable to be performed after any agent arrival in the next period $t = 1$. Thus, we have $x_m(q) = \mathbb{E}_q^D[\Delta D_m^1] = 0$. Secondly, taking an empty queue $i \in \mathcal{U}^0(q)$, we have that

$$\sum_{\substack{m \in \mathcal{M}: \\ M_{im}=1}} \Delta D_m^1 \leq \mathbb{1}\{\Delta A_i^1 = 1\},$$

and, in turn, taking expectations yields

$$\sum_{\substack{m \in \mathcal{M}: \\ M_{im}=1}} \mathbb{E}_q^D[\Delta D_m^1] \leq \mathbb{E}_q^D[\mathbb{1}\{\Delta A_i^1 = 1\}] = \lambda_i$$

as stated. $\square$

Recall that under $LQ(\mathcal{M}_+)$, we break ties arbitrarily if any. Moreover, an arriving agent of type $i \in \mathcal{U}^0(q)$ at time $t=1$ matches to the longest-queue among those in the set

$$\mathcal{A}^+(i) := \{j \in \mathcal{U}^+(q) : M_{jm} = M_{im} = 1 \text{ for some } m \in \mathcal{M}_+\}.$$

The next proposition formalizes this fact within this subsection's framework.

Proposition 5.2. *For all $i \in \mathcal{U}^0(q)$ with $\mathcal{A}^+(i) \neq \emptyset$, we have*

$$x_{(i,j)}^{LQ(\mathcal{M}_+)}(q) = \lambda_i \text{ for } j = \arg\max_{k\in\mathcal{A}^+(i)}\{q_k\}, \tag{12}$$

and $x_m^{LQ(\mathcal{M}_+)}(q) = 0$ for all $m \in \mathcal{M}\backslash\{(i,j)\}$. If the arg max *set in* (12) *contains multiple queues, choose one arbitrarily.*

Proof of Proposition 5.2. Consider $i \in \mathcal{U}^0(q)$ such that $\mathcal{A}^+(i) \neq \emptyset$, i.e., the empty queue i has at least one non-empty neighbor at $t=0$. Let $j = \arg\max_{k\in\mathcal{A}^+(i)}\{q_k\}$ be the longest-queue neighbor of i. If there are multiple such queues, choose one arbitrarily. Then under $LQ(\mathcal{M}_+)$, an arriving agent of type i matches with queue j at time $t=1$: $\Delta D_m^1 = \mathbb{1}\{A_i^1 = 1\}$ for match m satisfying $M_{jm} = M_{im} = 1$. In turn, $x_m^{LQ(\mathcal{M}_+)}(q) = \mathbb{E}_q^{LQ(\mathcal{M}_+)}[\Delta D_m^1] = \lambda_i$.

To show that all other matches have $x_m^{LQ(\mathcal{M}_+)}(q) = 0$, consider first $m \notin \mathcal{M}^+(q)$. Because agent arrivals happen one at a time, we have that if $m \notin \mathcal{M}^+(q)$, then m is unavailable to perform at time $t=1$. Thus, we have $\mathbb{E}_q^{LQ(\mathcal{M}_+)}[\Delta D_m^1] = 0$ for such m's.

Consider now $(i,j') \in \mathcal{M}^+(q)$ with $i \in \mathcal{U}^0(q)$, $\mathcal{A}^+(i) \neq \emptyset$ and $j' \in \mathcal{U}^+(q)$, but such that $j' \notin \arg\max_{k\in\mathcal{A}^+(i)}\{q_k\}$. As before, we have $\sum_{m\in\mathcal{M}:M_{im}=1} \Delta D_m^1 \leq \mathbb{1}\{A_i^1 = 1\}$ so that $\sum_{\substack{m\in\mathcal{M}:\\ M_{im}=1}} x_m^{LQ(\mathcal{M}_+)} \leq \lambda_i$. Since, we already showed that $x_{(i,j)}^{LQ(\mathcal{M}_+)}(q) = \lambda_i$ for $j = \arg\max_{k\in\mathcal{A}^+(i)}\{q_k\}$, we must have that $x_{(i,j')}^{LQ(\mathcal{M}_+)}(q) = 0$ as stated. $\square$

Per Proposition 5.1, under $LQ(\mathcal{M}_+)$ we then have that

$$\mathbb{E}_q^{LQ(\mathcal{M}_+)}[\mathcal{L}(Q^1) - \mathcal{L}(q)] \leq 2q(\lambda - Mx^{LQ(\mathcal{M}_+)}(q)) + 1,$$

where $x^{LQ(\mathcal{M}_+)}(q)$ is as in Proposition 5.2. Then to establish Lemma 5.4, it remains to show that

$$2q(\lambda - Mx^{LQ(\mathcal{M}_+)}(q)) \leq -2\frac{\epsilon}{n}\|q\|_1.$$

This is proved in two steps. Proposition 5.3 below shows that $LQ(\mathcal{M}_+)$, specifically its immediate (expected) allocation $x^{LQ(\mathcal{M}_+)}(q)$, minimizes the instantaneous drift, which is already characterized in Proposition 5.1. Finally, Proposition 5.4 shows that the instantaneous drift of the quadratic Lyapunov function $\mathcal{L}(q)$ has the desired form, which will directly imply Lemma 5.4. Recall the definition of the set $\mathcal{Z}(q)$ in (11).

Proposition 5.3. *Under $LQ(\mathcal{M}_+)$, the expected one-period allocation $x^{LQ(\mathcal{M}_+)}(q)$ satisfies*

$$x_m^{LQ(\mathcal{M}_+)}(q) \in \arg\min_{\xi \in \mathcal{Z}(q)} \{q(\lambda - M\xi)\}.$$

Proof of Proposition 5.3. Consider the following linear optimization problem in the statement of this lemma:

$$\begin{aligned} &\min\ q(\lambda - M\xi) \\ &\text{s.t.}\quad \xi \in \mathcal{Z}(q), \end{aligned}$$

where $\mathcal{Z}(q)$ is the linear constraint set in (11). We will rewrite the objective function, which will make the claim straightforward. First, because $q\lambda$ is a constant in this problem, the problem is equivalent (in terms of optimizers) to the following problem

$$\begin{aligned} &\max\ qM\xi \\ &\text{s.t.}\quad \xi \in \mathcal{Z}(q). \end{aligned} \tag{13}$$

Under any stationary greedy policy, since each non-empty queue can only have empty neighboring queues at any time, we can rewrite the objective function as

$$qM\xi = \sum_{i \in \mathcal{U}^+(q)} q_i \sum_{j \in \mathcal{U}^0(q)} M_{ij}\xi_{(i,j)} = \sum_{j \in \mathcal{U}^0(q)} \sum_{i \in \mathcal{U}^+(q)} q_i \xi_{(i,j)},$$

where recall that we slightly abuse the notation by $M_{ij} = M_{i,(i,j)}$. Thus, solving the optimization problem (13) is the same (in terms of optimizers) as solving a family of independent problems, one for each $j \in \mathcal{U}^0(q)$:

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{U}^+(q)} q_i \xi_{(i,j)} \\ \text{s.t.} \quad & \sum_{i \in \mathcal{U}^+(q)} \xi_{(i,j)} \leq \lambda_j \\ & \xi_{(i,j)} \geq 0 \quad \text{for all} \quad i \in \mathcal{U}^+(q). \end{aligned}$$

This is a relaxation of the knapsack problem with a well-known simple optimal solution in the form

$$\xi^*_{(i,j)} = \lambda_j \text{ for } i \in \arg\max_{k \in \mathcal{A}^+(j)} \{q_k\},$$

and $\xi^*_{(i,j)} = 0$ otherwise; let ξ^* be the unified solution (across all individual problems for each $j \in \mathcal{U}^0(q)$). Per Proposition 5.2, $\xi^* = x^{LQ(\mathcal{M}_+)}(q)$ as stated. □

Proposition 5.4. *Assume that $\mathcal{G}$ satisfies* **GP**. *Then there exists $\xi \in \mathcal{Z}(q)$ such that*

$$q(\lambda - M\xi) \leq -\frac{\epsilon}{n}\|q\|_1.$$

Proof of Proposition 5.4. Per Theorem 4.1, $(Mz^*)_i = \lambda_i$ for all $i \in \mathcal{Q}_0$. Define $\tilde{\lambda}$ as

$$\tilde{\lambda}_i = \lambda_i \text{ for all } i \in \mathcal{U}^0(q), \text{ and } \tilde{\lambda}_i = \lambda_i + \frac{\epsilon}{n} \text{ for all } i \in \mathcal{U}^+(q),$$

where ϵ is the general position gap. Then by Corollary 4.1, there exists $\tilde{z}^*$ such that $(M\tilde{z}^*)_i = \tilde{\lambda}_i$ for all $i \in \mathcal{Q}_0$, where z^* and $\tilde{z}^*$ have the same optimal basis. This is because the perturbation satisfies the condition $y^l(\tilde{\lambda} - \lambda) \geq -\epsilon$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, as $y^l \in \{-1, -1/2, 0, 1/2, 1\}$ per Theorem 4.1. Note that $\tilde{z}^*$ also satisfies $\sum_{\substack{m \in \mathcal{M}: \\ M_{im}=1}} \tilde{z}^*_m \leq \lambda_i = \tilde{\lambda}_i$ for all $i \in \mathcal{U}^0(q)$. Now we construct ξ based on $\tilde{z}^*$. Let

$$\xi_m = \begin{cases} \tilde{z}^*_m, & \text{if } m \in \mathcal{M}^+(q), \\ 0, & \text{otherwise.} \end{cases}$$

Clearly we have $\xi \in \mathcal{Z}(q)$. Note that for all $i \in \mathcal{U}^+(q)$, by definition we have $m \in \mathcal{M}^+(q)$ for all

$m \in \mathcal{M}$ such that $M_{im} = 1$. This implies that for all $i \in \mathcal{U}^+(q)$, we have

$$(M\xi)_i = \tilde{\lambda}_i = \lambda_i + \frac{\epsilon}{n}.$$

Thus,

$$q(\lambda - M\xi) = \sum_{i \in \mathcal{U}^+(q)} q_i(\lambda_i - (M\xi)_i) = -\frac{\epsilon}{n} \sum_{i \in \mathcal{U}^+(q)} q_i = -\frac{\epsilon}{n} \sum_{i \in \mathcal{Q}_0} q_i$$

as stated. □

The proof of Lemma 5.4 is now immediate.

Proof of Lemma 5.4. Given Propositions 5.3 and 5.4, we have

$$q(\lambda - Mx^{LQ(\mathcal{M}_+)}(q)) \le -\frac{\epsilon}{n}\|q\|_1,$$

which implies that under $LQ(\mathcal{M}_+)$, per Proposition 5.1, we have

$$\mathbb{E}_q^{LQ(\mathcal{M}_+)}[\mathcal{L}(Q^1) - \mathcal{L}(q)] \le -2\frac{\epsilon}{n}\|q\|_1 + 1$$

as stated. □

5.4 Proof of Theorem 3.2

Throughout this subsection, we assume without loss of generality that $\mathcal{G}$ is a tree that satisfies **GP** with $\mathcal{M}_0 = \emptyset$, and we fix an arbitrary topological order $p(\cdot)$. Let j_+ be the unique queue such that $\mathcal{Q}_+ = \{j_+\}$ per Lemma 3.1.

The Lyapunov function. The quadratic Lyapunov function, which is used in our analysis of $LQ(\mathcal{M}_+)$, does not work for $SP(\mathcal{M}_+, p)$. To see this, consider the network in Figure 2. Note that for any topological order $p(\cdot)$, we have $p(1) < p(2)$, i.e., match 1 has a higher priority than match 2. Take t where all queues are empty except queue 1. Then under $SP(\mathcal{M}_+, p)$, $\mathcal{L}(Q^t) = \sum_{i \in \mathcal{Q}_0} (Q_i^t)^2$ does not necessarily decrease in expectation, since $\|Q^t\|_1$ decreases by 1 with probability λ_2, whereas $\|Q^t\|_1$ increases by 1 with probability $1 - \lambda_2 - \lambda_6$, and $\lambda_2 < (1/2)(1 - \lambda_6)$ does not violate the assumed **GP**.

Instead, we construct a Lyapunov function using the specific algebraic structure of the optimal solution of (SPP) given in Theorem 4.1. Before introducing the Lyapunov function, we introduce some useful definitions. Recall that $d(i, j)$ is the length of the directed path from $i \in \mathcal{A}$ to $j \in \mathcal{A}$

in $\overrightarrow{\mathcal{G}}$. We define the set $\mathcal{B}(i) := \{j \in \mathcal{A} : d(i,j) = 1\}$. An intuitive way to interpret the set $\mathcal{B}(i)$ is as follows. Consider "hanging" $\overrightarrow{\mathcal{G}}$ by the root j_+. Then $\mathcal{B}(i)$ contains all agent types that are *directly below* $i \in \mathcal{A}$ in $\overrightarrow{\mathcal{G}}$. For example in Figure 2, we have $\mathcal{B}(3) = \{2, 4\} \subseteq \mathcal{A}$.

Recalling that $Q^t = A^t - MD^t$ for all $t \geq 0$ per (1), consider the stochastic variant of (SPP) at a given time t:

$$\begin{aligned} \max \quad & r \cdot z \\ \text{s.t.} \quad & Mz + Q^t = A^t \\ & z \in \mathbb{R}^d_{\geq 0}, Q^t \in \mathbb{R}^n_{\geq 0}. \end{aligned}$$

It is a simple observation that by the construction of the surplus vectors, we have $y^m Mz = z_m$ for all $m \in \mathcal{M}_+$. Multiplying both sides of the linear constraint set of this stochastic variant with y^m, $m \in \mathcal{M}_+$, yields $z_m + y^m Q^t = y^m A^t$ for all $m \in \mathcal{M}_+$. Per Theorem 4.1, we should have $z_m \approx 0$ for all $m \in \mathcal{M}_0$ and $z^m \approx y^m A^t$ for all $m \in \mathcal{M}_+$ to achieve optimality for (SPP). This suggest that we should have $y^m Q^t \approx 0$ for all $m \in \mathcal{M}_+$.

It is then natural to construct a function $f(Q^t)$ such that when $f(Q^t) = 0$, then we have $y^m Q^t = 0$ for all $m \in \mathcal{M}_+$, or equivalently, $y^i Q^t = 0$ for all $i \in \mathcal{Q}_0$. To that end, define

$$Z_i^t := y^i Q^t \quad \text{for all} \quad i \in \mathcal{Q}_0.$$

The set $\mathcal{U}_0 = \{i \in \mathcal{Q}_0 : \sum_{m \in \mathcal{M}} M_{im} = 1\}$—the queues in $\mathcal{Q}_0$ that are leaves of $\mathcal{G}$—must be a non-empty set. Otherwise $\mathcal{G}$ must contain a cycle and here, recall, we are assuming that $\mathcal{G}$ is a tree. Trivially, $\mathcal{B}(i) = \emptyset$ for all $i \in \mathcal{U}_0$.

We take, for our Lyapunov function, the mapping

$$f(Q^t) := \sum_{i \in \mathcal{A} \setminus \mathcal{U}_0} \alpha_i \Big(\sum_{j \in \mathcal{B}(i)} Z_i^t \Big)^2, \tag{14}$$

where $\alpha_i > 0$ for all $i \in \mathcal{A} \setminus \mathcal{U}_0$. For example, the corresponding Lyapunov function to the matching network in Figure 2 is $f(Q^t) = \alpha_2 (Q_1^t)^2 + \alpha_4 (Q_5^t)^2 + \alpha_3 (Q_2^t - Q_1^t + Q_4^t - Q_5^t)^2 + \alpha_6 (Q_7^t + (Q_3^t - Q_2^t - Q_4^t + Q_1^t + Q_5^t) + Q_8^t)^2$.

The following is the main ingredient in the proof of Theorem 3.2.

Lemma 5.5. *Assume that $\mathcal{G}$ is a tree that satisfies* **GP** *with $\mathcal{M}_0 = \emptyset$. Then under $SP(\mathcal{M}_+, p)$, the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$ is ergodic. Moreover, there exist strictly positive coefficients*

$\alpha_i,\ i \in \mathcal{Q}_0$, *and constants* $\theta, \gamma > 0$ *such that* $f(Q^t)$ *in* (14) *decreases in expectation:*

$$\mathbb{E}[f(Q^{t+1}) - f(Q^t) \mid Q^t] \leq -\gamma\sqrt{f(Q^t)} + \theta \text{ for all } t \geq 0. \tag{15}$$

The proof of Lemma 5.5 is given in the appendix. Next we apply it to complete the proof of Theorem 3.2.

Proof of Theorem 3.2. The drift property (15) in Lemma 5.5 implies that the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$ is positive recurrent; e.g., see (Robert, 2003, Corollary 8.7). Given Lemma 5.5, moment bounds in the steady-state follow trivially from Theorem 5.2. In particular, under the Markov chain's unique stationary distribution, which we denote by π, we have

$$\mathbb{E}_\pi[|\sum_{j\in\mathcal{B}(i)} Z_i^0|] \leq \sqrt{\frac{\theta}{\alpha_i\gamma}} \text{ for all } i \in \mathcal{A}\backslash\mathcal{U}_0,$$

where $Q^0 \sim \pi$. This implies that for all $i \in \mathcal{A}\backslash\mathcal{U}_0$, we have $\mathbb{E}[|\sum_{j\in\mathcal{B}(i)} Z_i^t|] = \mathcal{O}(1)$ for all $t > 0$. Note that by the construction of the surplus vectors, we have

$$\sum_{j\in\mathcal{B}(i)} Z_i^t = \sum_{j\in\mathcal{B}(i)} y^i Q^t = \sum_{j\in\mathcal{B}(i)} Q_j^t - \sum_{j\in\mathcal{B}(i)}\sum_{k\in\mathcal{B}(j)} y^j Q^t = \sum_{j\in\mathcal{B}(i)} Q_j^t - \sum_{j\in\mathcal{B}(i)}\sum_{k\in\mathcal{B}(j)} Z_k^t,$$

for all $i \in \mathcal{A}\backslash\mathcal{U}_0$ and for all $t > 0$. Therefore, we conclude that under $SP(\mathcal{M}_+, p)$, we have

$$\sum_{i\in\mathcal{Q}_0} \mathbb{E}[Q_i^t] = \mathcal{O}(1) \text{ for all } t > 0.$$

Then per Lemma 5.1, $SP(\mathcal{M}_+, p)$ is hindsight optimal as stated. $\square$

The proof of Lemma 5.5 reveals that α_i's depend on γ's, and θ depends on α_i's in a complicated way (in contrast to (9)), which is the reason why we cannot establish the optimal scaling ϵ^{-1} for regret under $SP(\mathcal{M}_+, p)$.

6 Numerical examples

In this section, we present some simulation results to provide further insights about our greedy policies. All simulations are based on 10000 replications.

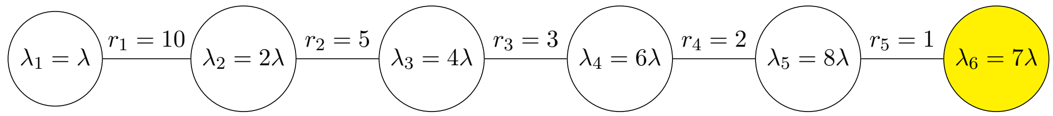

Figure 5: A matching network that satisfies **GP** with $\mathcal{M}_0 = \emptyset$ and $\mathcal{Q}_+ = \{6\}$ (yellow vertex), where the scalar λ is chosen so that $\sum_{i\in\mathcal{A}} \lambda_i = 1$ ($\lambda \approx 0.03$). The optimal solution of (SPP) has $z^* = (\lambda, \lambda, 3\lambda, 3\lambda, 5\lambda)$ and $s^* = (0, 0, 0, 0, 0, 2\lambda)$, and the general position gap is $\epsilon = \lambda$. In general, for any $\lambda_1 \in [\lambda, 2\lambda)$, we have $\epsilon = z_2^* = \lambda_2 - \lambda_1 = 2\lambda - \lambda_1$.

Example 6.1 (regret scaling of $SP(\mathcal{M}_+, p)$)**.** Consider the network in Figure 5. The priority order under $SP(\mathcal{M}_+, p)$ is $m_1 \succ m_2 \succ m_3 \succ m_4 \succ m_5$, where $m \succ m'$ means that $p(m) < p(m')$ (m has a higher priority than m'). We consider 10 separate scenarios, where λ_1 takes values sequentially in the set $\{\lambda, 1.1\lambda, \ldots, 1.9\lambda\}$. For any given scenario, the optimal basis remains unchanged, and the general position gap is $\epsilon = \lambda_2 - \lambda_1 = 2\lambda - \lambda_1$. Figure 6 suggests that the scaling for the achieved constant regret under $SP(\mathcal{M}_+, p)$ in Theorem 3.2 is ϵ^{-1}, as in Theorem 3.1. ∎

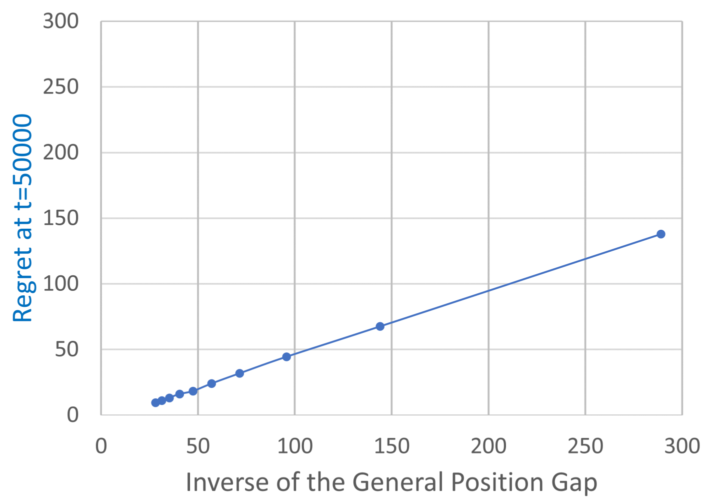

Figure 6: We consider the network in Figure 5. The x-axis corresponds to the inverse of the general position gap for each scenario, $\epsilon^{-1} = 1/(2\lambda - \lambda_1)$, where λ_1 takes values sequentially in the set $\{\lambda, 1.1\lambda, \ldots, 1.9\lambda\}$, and ϵ^{-1} increases as λ_1 increases. For the first scenario ($\lambda_1 = \lambda$) we have $\epsilon^{-1} \approx 28$, and for the last scenario ($\lambda_1 = 1.9\lambda$) we have $\epsilon^{-1} \approx 289$. The y-axis corresponds to the regret under $SP(\mathcal{M}_+, p)$ at time $t = 5 \cdot 10^4$, where the time horizon is $T = 10^5$.

Remark 6.1 (dependence of $p(\cdot)$ on λ)**.** The construction of the static priority order $p(\cdot)$ reflects the arrival probabilities only through their implication on the sets $\mathcal{M}_+$ and $\mathcal{Q}_+$. Given two different arrival-probability vectors, as long as they result in the same optimal basis for (SPP), the set of all possible topological orders $p(\cdot)$'s are the same.

Given λ, consider the optimal basis of (SPP), and the induced sets $\mathcal{M}_+$ and $\mathcal{Q}_+$ under **GP**.

Then per Theorem 4.1, we know that any other arrival-probability vector, say $\tilde{\lambda}$, such that $\tilde{\lambda}$ is in the cone

$$\mathcal{Y} := \left\{ \lambda' \in \mathbb{R}^n_+ : y^m \lambda' > 0 \text{ for all } m \in \mathcal{M}_+ \text{ and } y^j \lambda' > 0 \text{ for all } j \in \mathcal{Q}_+ \right\},$$

results in the same set of all topological orders. ∎

Example 6.2 (the robustness of $SP(\mathcal{M}_+, p)$)**.** Consider the network in Figure 5. Our priority order is $m_1 \succ m_2 \succ m_3 \succ m_4 \succ m_5$, where $m \succ m'$ means that m has a higher priority than m', i.e., $p(m) < p(m')$. Consider an alternative static priority policy with the priority order $m_2 \succ m_1 \succ m_3 \succ m_4 \succ m_5$, i.e., the priority assignment between m_1 and m_2 is altered. Note that this alternative priority order is not a topological order. Figure 7 shows that both policies achieve constant regret, i.e., one that does not grow with time.

Next let us change a bit the arrival probabilities, and consider $\lambda_1 = 1.9\lambda$ (instead of $\lambda_1 = \lambda$) with all else remaining the same. This perturbation on λ does not change the optimal basis—the sets $\mathcal{M}_+$ and $\mathcal{Q}_+$—and neither does it change our prescribed priority order. As Figure 8 illustrates, while our $SP(\mathcal{M}_+, p)$ still achieves constant regret, the alternative policy has a regret that grows with time. As argued in Remark 6.1, if the perturbation on λ does not affect the optimal basis, $SP(\mathcal{M}_+, p)$ remains hindsight optimal, whereas a deviation in the priority order may result in suboptimality. ∎

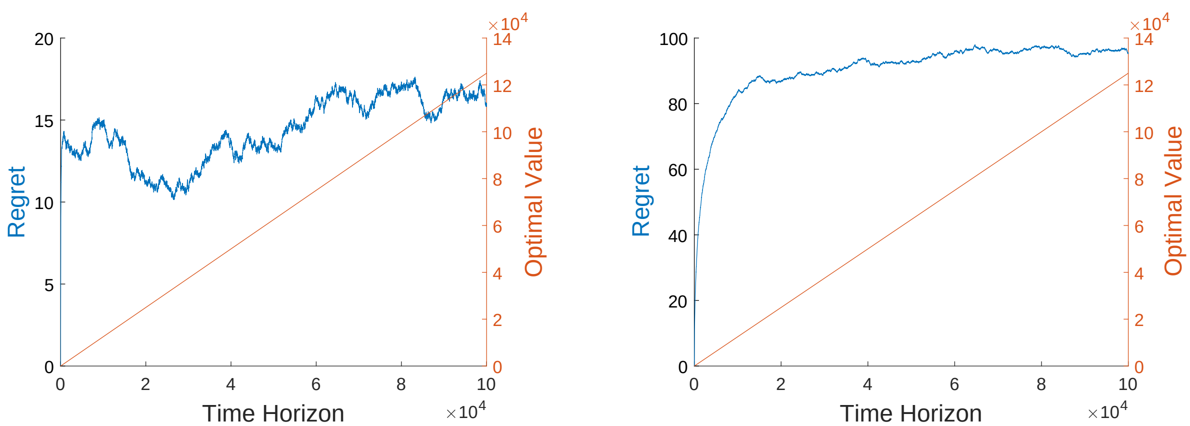

Figure 7: We consider the network in Figure 5, where $\epsilon \approx 0.03$. (LEFT) $SP(\mathcal{M}_+, p)$ is hindsight optimal. (RIGHT) The alternative greedy static priority policy achieves a higher regret than $SP(\mathcal{M}_+, p)$, but it is still hindsight optimal.

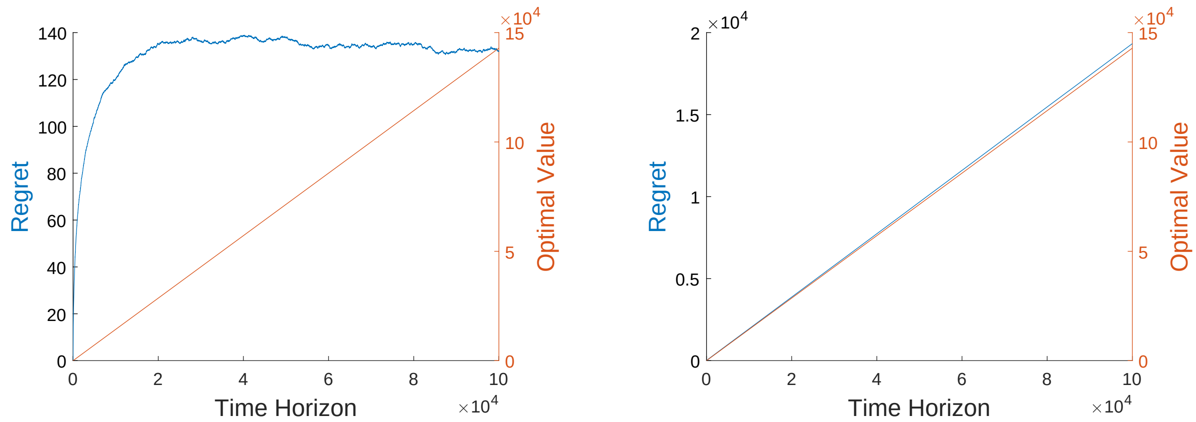

Figure 8: We consider the network in Figure 5, where $\lambda_1 = 1.9\lambda$ instead, and $\epsilon \approx 0.003$ now. (LEFT) $SP(\mathcal{M}_+, p)$ is still hindsight optimal given the perturbation on the arrival-probability vector. (RIGHT) Under the proposed perturbation, the alternative policy is no longer hindsight optimal, and its regret grows with t.

The next example illustrates that there exist matching networks, where $LQ(\mathcal{M}_+)$ generates a higher total value than $SP(\mathcal{M}_+, p)$, and vice versa.

Example 6.3 (Comparison of $LQ(\mathcal{M}_+)$ and $SP(\mathcal{M}_+, p)$)**.** Consider the network in Figure 5 again. Per Theorems 3.1 and 3.2, both $LQ(\mathcal{M}_+)$ and $SP(\mathcal{M}_+, p)$ are hindsight optimal. Intuitively, the priority order under $SP(\mathcal{M}_+, p)$ coincides with the order of matches with respect to their values, and one can expect that $SP(\mathcal{M}_+, p)$ may result in a smaller regret than $LQ(\mathcal{M}_+)$, and Figure 9 supports this intuition.

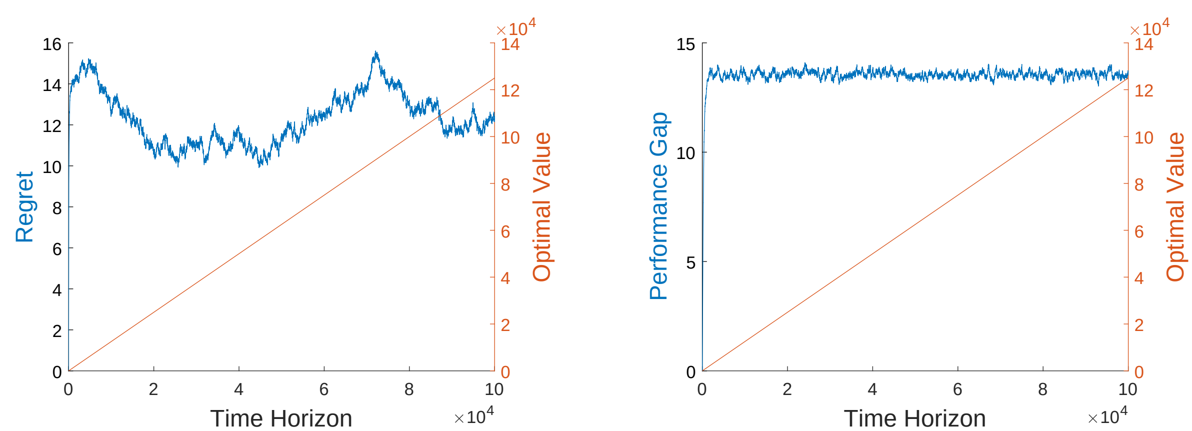

Figure 9: We consider the network in Figure 5, where $\epsilon \approx 0.03$. (LEFT) $SP(\mathcal{M}_+, p)$ is hindsight optimal. (RIGHT) The difference in the generated total value (performance gap) between $SP(\mathcal{M}_+, p)$ and $LQ(\mathcal{M}_+)$, where the former "dominates" the latter.

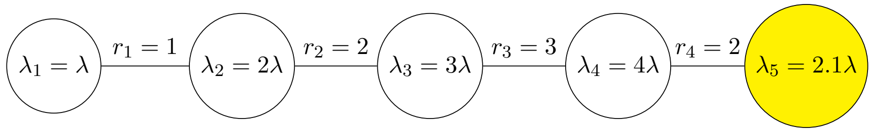

Figure 10: A matching network that satisfies **GP** with $\mathcal{M}_0 = \emptyset$ and $\mathcal{Q}_+ = \{5\}$ (yellow vertex), where the scalar λ is chosen so that $\sum_{i \in \mathcal{A}} \lambda_i = 1$ ($\lambda \approx 0.08$). The optimal solution of (SPP) has $z^* = (\lambda, \lambda, 2\lambda, 2\lambda)$ and $s^* = (0, 0, 0, 0, 0.1\lambda)$, and the general position gap is $\epsilon = s_5^* = 0.1\lambda$.

Now consider the network in Figure 10. Figure 11 illustrates that $LQ(\mathcal{M}_+)$ results in a smaller regret than $SP(\mathcal{M}_+, p)$. Hence, both simulations suggest that there exist matching networks and network primitives so that $LQ(\mathcal{M}_+)$ achieves smaller regret than $SP(\mathcal{M}_+, p)$, and vice versa. ∎

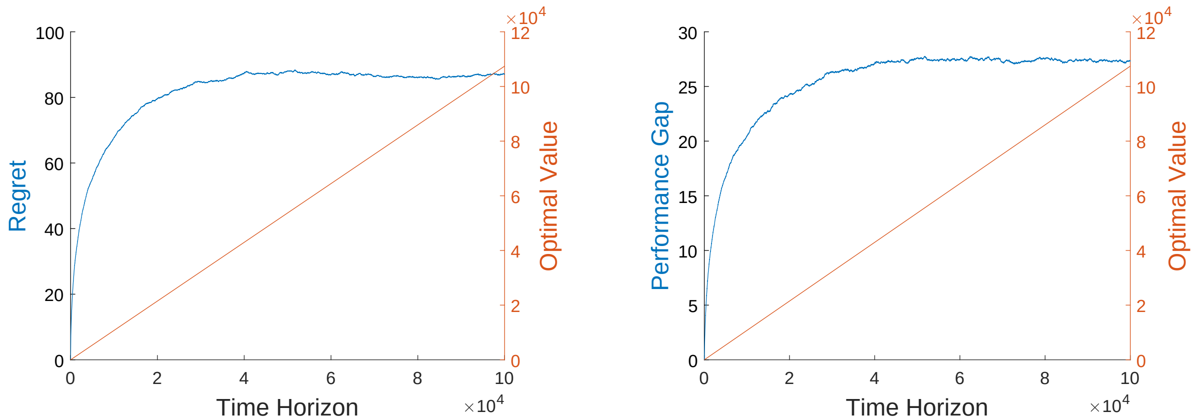

Figure 11: We consider the network in Figure 10, where $\epsilon \approx 0.008$. (LEFT) $LQ(\mathcal{M}_+)$ is hindsight optimal. (RIGHT) The difference in the generated total value (performance gap) between $LQ(\mathcal{M}_+)$ and $SP(\mathcal{M}_+, p)$, where the former "dominates" the latter.

Remark 6.2 (scaling of the lower bound on regret with the number of agent types)**.** Our current work, together with Kerimov et al. (2021) that precedes it, concern scaling of regret as a function of ϵ. The general lower bound in Kerimov et al. (2021) stipulates that no policy can do better than $\Omega(\epsilon^{-1})$; this lower bound is not explicit as to dependence on the network structure or, more specifically, the number of agent types. Our Theorem 3.1 shows that $LQ(\mathcal{M}_+)$'s regret is at most $r_{max}n/\epsilon$, a bound that grows with the number of agent types n. Whether or not—or more specifically under which conditions—the best achievable regret grows with n remains an open question. ∎

7 Concluding remarks

We found that in the general class of two-way matching networks that satisfy a general position condition, greedy policies (whose design is) based on static optimal matching rates achieve constant regret at all times; they are hindsight optimal. In these networks, in particular, there is no positive externality from waiting to form future matches. Moreover, greediness offers local and simple matching rules that, other than identifying static optimal matching rates a priori, does not require any additional optimization.

The greedy policies we prescribe, longest-queue and static priority, differ in whether they depend on the state of the network or not. Therefore, these policies may be appealing in different contexts.

These results complement those in our previous paper Kerimov et al. (2021), where we found that in multi-way matching networks, greedy policies are not hindsight optimal, but carefully designed periodic clearing matching policies do achieve hindsight optimality.

General position is a weak but necessary condition for hindsight optimality (Kerimov et al., 2021). Moreover, the optimal scaling for constant regret is given by ϵ^{-1}, where ϵ is a simple quantity that arises from the static-planning problem (a deterministic counterpart) that also provides the optimal matching rates. This quantity is intimately linked with stability; if queues of types that are not under-demanded (queues in $\mathcal{Q}_0$) are bounded by ϵ^{-1} (in expectation) at all times, the policy is hindsight optimal and the scaling for constant regret is the same as the moment bound on the queue-lengths.

We hope that what we learned in this paper can be leveraged to expand the results to include several practical, yet challenging considerations.

The simplest is the inclusion of holding costs. In matching networks, like the ones we consider in this paper, there is an intimate connection between value maximization and holding-cost minimization; we refer the reader to Kerimov et al. (2021) for a detailed discussion of this correspondence. Less immediate are expansions of the models to capture agent departures and decentralized matching networks. When agents abandon the market without matching after some agent-specific (possibly random) time, it is no longer clear—even in networks with two-way matches—that there exists a greedy hindsight optimal policy. In such networks, it might be important to build an inventory in anticipation of "short-fuse" agents that participate in high value matches and are highly impatient. In decentralized dynamic matching markets, agents wish to maximize their own payoffs (Leshno, 2011; Baccara et al., 2020), and agents might act in a way that compromise global optimality. Combining the queueing modeling in this paper with mechanism design tools might help to shed further light on how to regulate such settings.

References

I. Adan, A. Bušić, J. Mairesse, and G. Weiss. Reversibility and further properties of FCFS infinite bipartite matching. *Mathematics of Operations Research*, 43(2):598–621, 2018.

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

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

A. Aveklouris, L. DeValve, A.R. Ward, and X. Wu. Matching impatient and heterogeneous demand and supply. *arXiv preprint arXiv:2102.02710*, 2021.

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

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. *preprint arXiv:2002.03205*, 2020.

A. Bušić and S. Meyn. Approximate optimality with bounded regret in dynamic matching models. *ACM SIGMETRICS Performance Evaluation Review*, 43(2):75–77, 2015.

A. Bušić, V. Gupta, and J. Mairesse. Stability of the bipartite matching model. *Advances in Applied Probability*, 45(2):351–378, 2013.

A. Cadas, A. Bušić, and J. Doncel. Optimal control of dynamic bipartite matching models. In *Proceedings of the 12th EAI International Conference on Performance Evaluation Methodologies and Tools*, pages 39–46, 2019.

N. Collina, N. Immorlica, K. Leyton-Brown, B. Lucier, and N. Newman. Dynamic weighted matching with heterogeneous arrival and departure rates. In *International Conference on Web and Internet Economics*, pages 17–30. Springer, 2020.

P.W. Glynn and A. Zeevi. Bounding stationary expectations of markov processes. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, pages 195–214. Institute of Mathematical Statistics, 2008.

V. Gupta. Greedy algorithm for multiway matching with bounded regret. *Operations Research*, 2022.

I. Gurvich and A.R. Ward. On the dynamic control of matching queues. *Stochastic Systems*, 4(2): 479–523, 2014.

J. M. Harrison and M. J. Lopez. Heavy traffic resource pooling in parallel-server systems. *Queueing Systems*, 33:339–368, 1999.

M. Jonckheere, P. Moyal, C. Ramírez, and N. Soprano-Loto. Generalized max-weight policies in stochastic matching. *Stochastic Systems*, 2022.

S. Kerimov, I. Ashlagi, and I. Gurvich. Dynamic matching: Characterizing and achieving constant regret. *Available at SSRN 3824407*, 2021.

J.D. Leshno. Dynamic matching in overloaded systems. 2011.

S.T. Maguluri and R. Srikant. Heavy traffic queue length behavior in a switch under the maxweight algorithm. *Stochastic Systems*, 6(1):211–250, 2016.

J. Mairesse and P. Moyal. Stability of the stochastic matching model. *Journal of Applied Probability*, 53(4):1064–1077, 2016.

P. Moyal and O. Perry. On the instability of matching queues. *The Annals of Applied Probability*, 27(6):3385–3434, 2017.

M. Nazari and A.L. Stolyar. Reward maximization in general dynamic matching systems. *Queueing Systems*, 91(1):143–170, 2019.

P. Robert. *Stochastic networks and queues.* Springer-Verlag Berlin Heidelberg, 2003.

D. Shah, N.S. Walton, and Y. Zhong. Optimal queue-size scaling in switched networks. *The Annals of Applied Probability*, 24(6):2207–2245, 2014.

A Proofs of auxiliary lemmas

Proof of Lemma 3.1. Let (z^*, s^*) be the unique non-degenerate optimal solution of (SPP) under **GP**. Note that the projection of (z^*, s^*) remains as a non-degenerate optimal basic feasible solution when restricted to each component $\mathcal{C}_k$, $k \in [K]$. This immediately follows from the construction of the (SPP)-residual graph $\mathcal{G}'$, as any component $\mathcal{C}_k$ is "disconnected" from $\mathcal{G}$ by removing all redundant matches in $\mathcal{M}_0$ ($m \in \mathcal{M}$ with $z_m^* = 0$). Assume that $\mathcal{C}_k$ contains $n_k \geq 1$ vertices (queues) and $m_k \geq 0$ edges (matches) for all $k \in [K]$.

(i) Since $\mathcal{C}_k$ does not contain any redundant matches, non-degeneracy implies that $n_k \geq m_k$, since there are n_k many basic variables in the projection of (z^*, s^*), and all m_k variables corresponding to active matches in $\mathcal{C}_k$ are basic. If $\mathcal{C}_k$ contains at least two cycles, then we must have $m_k > n_k$, which is a contradiction. Thus, $\mathcal{C}_k$ contains at most one cycle.

(ii) Since $\mathcal{C}_k$ is a component, it is connected, and if it does not contain a cycle, then it must be a tree. Thus, we have $n_k = m_k + 1$. Then non-degeneracy implies that $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 1$.

(iii) Per (i), $\mathcal{C}_k$ contains exactly one cycle. Then we must have $n_k = m_k$. Assume to the contrary that this cycle is of even length. Consider the projection of (z^*, s^*) when restricted to $\mathcal{C}_k$, which remains as a non-degenerate optimal basic feasible solution as argued above. Since $n_k = m_k$, all slack variables in the projection are non-basic. Consider all the matches that are included in this even cycle and the corresponding column vectors of these matches in the matching matrix M. Since the cycle is of even length, these column vectors are not linearly independent, which is a contradiction to the fact that the optimal solution is a basic feasible solution. Thus, $\mathcal{C}_k$ can only contain an odd cycle, where we note that the corresponding columns to the matches included in the cycle are linearly independent. Finally, non-degeneracy and $n_k = m_k$ implies that $|\mathcal{A}(\mathcal{C}_k) \cap \mathcal{Q}_+| = 0$. □

Proof of Lemma 5.1. The proof follows immediately as in (Kerimov et al., 2021, Lemma 4.1) with the following modifications. Since we must have $\tilde{\lambda}_i > 0$ for all $i \in \mathcal{Q}_0$, fix t such that $\lambda_{\min} > Bt^{-1}$. Per Theorem 4.1, we have $\omega = 1$. Therefore, we conclude that $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \leq tr_{\max}\omega\|\lambda - \tilde{\lambda}\| \leq tr_{\max}Bt^{-1} = r_{\max}B$.

□

B Proof of Lemma 5.5

Before proving Lemma 5.5, we begin with some preliminaries. We define a family of functions $g_i,\ i \in \mathcal{A}$, as follows. Let $g_i^t := 0$ for all $i \in \mathcal{U}_0$ and for all $t \geq 0$, and sequentially define

$$g_i^t := \Big(\sum_{j\in\mathcal{B}(i)} Z_j^t\Big)^2 + \alpha_i \sum_{j\in\mathcal{B}(i)} g_j^t \text{ for all } i \in \mathcal{A}\backslash\mathcal{U}_0 \text{ and for all } t \geq 0,$$

where the coefficients $\alpha_i > 0,\ i \in \mathcal{A}\backslash\mathcal{U}_0$, will be determined later in the section. To prove Lemma 5.5, we prove the following slightly more general result.

Lemma B.1. *Assume that $\mathcal{G}$ is a tree that satisfies* **GP** *with $\mathcal{M}_0 = \emptyset$. Then under $SP(\mathcal{M}_+, p)$, the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$ is ergodic. Moreover, there exist strictly positive coefficients $\alpha_i,\ i \in \mathcal{Q}_0$, and constants $\gamma_i, \theta_i > 0,\ i \in \mathcal{Q}_0$, such that*

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t] \leq -\gamma_i \sqrt{g_i^t} + \theta_i \textit{ for all } t \geq 0. \tag{16}$$

Observe that Lemma 5.5 follows immediately from Lemma B.1, noting that $f(Q^t)$ is $g_{j_+}^t$ itself with coefficients redefined, where recall that j_+ is the root of $\overrightarrow{\mathcal{G}}$.

Recall that $d(i,j)$ is the length of the directed path from $i \in \mathcal{A}$ to $j \in \mathcal{A}$ in $\overrightarrow{\mathcal{G}}$, and that $\mathcal{B}(i) = \{j \in \mathcal{A} : d(i,j) = 1\}$, which contains the agent types that are *directly below* $i \in \mathcal{A}$ in $\overrightarrow{\mathcal{G}}$. Define $\hat{\mathcal{B}}(i) := \{m \in \mathcal{M} : \mathcal{A}(m) = \{i,k\} \text{ and } k \in \mathcal{B}(i)\}$, i.e., the set that contains all matches that are *directly below* $i \in \mathcal{A}$ in $\overrightarrow{\mathcal{G}}$. For example in Figure 2, we have $\hat{\mathcal{B}}(6) = \{3,6,7\} \subseteq \mathcal{M}$.

For any $i \in \mathcal{Q}_0$, let $i^\uparrow$ be the unique queue in $\overrightarrow{\mathcal{G}}$ such that $d(i^\uparrow, i) = 1$. In words, $i^\uparrow$ is the queue that is *directly above* $i \in \mathcal{Q}_0$ in $\overrightarrow{\mathcal{G}}$, i.e., the direction on m such that $A(m) = \{i, i^\uparrow\}$ is incoming to i. Note that such a queue does not exist for the root $j_+ \in \mathcal{Q}_+$.

The following remark is crucial in the proof of Lemma 5.5, and it follows directly from the definition of a topological order $p(\cdot)$.

Remark B.1. *Fix some $i \in \mathcal{Q}_0$, and let $m^\uparrow \in \mathcal{M}$ be the match that i and $i^\uparrow$ participates, i.e., $A(m^\uparrow) = \{i, i^\uparrow\}$. Given any topological order $p(\cdot)$, we have $p(m) < p(m^\uparrow)$ for all $m \in \hat{\mathcal{B}}(i)$. In words, any match that is directly below i has a higher priority than the match that is directly above i.*

Finally, define $d_i := \max_{j\in\mathcal{U}_0\cap\mathcal{A}(\mathcal{T}_i)} d(i,j)$ for all $i \in \mathcal{A}$. One can intuitively view d_i as the parameter indicating the position of queue i relative to the root j_+ of $\overrightarrow{\mathcal{G}}$; the larger the d_i, the closer queue i is to the root j_+. For example in Figure 2, we have $\mathcal{U}_0 = \{1,5,7,8\} \subseteq \mathcal{A}$, $d_2 = 1$, $d_3 = 2$ and $d_6 = 3$. Now we prove Lemma B.1.

Proof of Lemma B.1. We will first establish (16), and we then prove the ergodicity of the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$. We will also use the following three aids in the proof. Propositions B.1 and B.2 are proven in the end of this section, and Proposition B.3 is a known standard result that appears, for example, in (Robert, 2003, Corollary 8.7). Throughout the proof, for ease of exposition, for all $i \in \mathcal{A}\backslash\mathcal{U}_0$ and for all $t \geq 0$, we let

$$Z^t_{\mathcal{B}(i)} := \sum_{j\in\mathcal{B}(i)} Z^t_j,\ g^t_{\mathcal{B}(i)} := \sum_{j\in\mathcal{B}(i)} g^t_j,\ \text{and } Q^t_{\mathcal{B}(i)} := \sum_{j\in\mathcal{B}(i)} Q^t_j.$$

Proposition B.1 (bounded jumps)**.** *Under the assumptions of Theorem B.1, for all $i \in \mathcal{A}\backslash\mathcal{U}_0$, for all $t \geq 0$, and for any constant $B_i > 0$, we have*

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, g_i^t \leq B_i] \leq \theta_i$$

for some constant $\theta_i > 0$, which depends only on n and B_i.

Proposition B.2. *Under the assumptions of Theorem B.1, for all $i \in \mathcal{A}\backslash\mathcal{U}_0$ and for all $t \geq 0$, we have*

$$\mathbb{E}[(Z^{t+1}_{\mathcal{B}(i)})^2 - (Z^t_{\mathcal{B}(i)})^2 \mid Q^t, Q^t_{\mathcal{B}(i)} > 0] = -\eta_i Z^t_{\mathcal{B}(i)} + \delta_i,$$

where $\eta_i := 2y^i\lambda$ and $\delta_i := \sum_{j\in\mathcal{A}(\mathcal{T}_i)} \lambda_j$.

Proposition B.3 (Robert (2003), Corollary 8.7)**.** *Let $(M_t : t \geq 0)$ be a discrete-time, homogeneous, irreducible and aperiodic Markov chain with values in a countable state space $\mathcal{X}$. If there exist a function $f : \mathcal{X} \to \mathbb{R}_+$ and constants $K, \eta > 0$ such that*

(i) $\mathbb{E}_x[f(M_1) - f(x)] \leq -\eta$ *when* $f(x) > K$,

(ii) $\mathbb{E}_x[f(M_1)] < \infty$ *when* $f(x) \leq K$, *and*

(iii) *the set* $\{x \in \mathcal{X} : f(x) \leq K\}$ *is finite,*

then the Markov chain $(M_t : t \geq 0)$ is ergodic.

We use strong induction on d_i, $i \in \mathcal{A}\backslash\mathcal{U}_0$, where recall that $d_i = \max_{j\in\mathcal{U}_0\cap\mathcal{A}(\mathcal{T}_i)} d(i,j)$. The following simple observation is used repeatedly in the analysis, which follows from the definition

$g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t$:

$$\sqrt{g_i^t} \geq |Z_{\mathcal{B}(i)}^t|. \tag{17}$$

Basis. Consider $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_i = 1$. This implies $\mathcal{B}(i) \subseteq \mathcal{U}_0$ so that $y^j Q^t = Q_j^t$ for all $j \in \mathcal{B}(i)$. By definition, since $g_j^t = 0$ for all $j \in \mathcal{B}(i) \subseteq \mathcal{U}_0$, we have

$$g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t = \left(\sum_{j\in\mathcal{B}(i)} y^j Q^t\right)^2 = (Q_{\mathcal{B}(i)}^t)^2.$$

Fix $B_i > 0$ (its specific value will be determined later), and consider the following two cases:

- **1: ($g_i^t > B_i$).** Note that $g_i^t = (Q_{\mathcal{B}(i)}^t)^2$ and $Q_{\mathcal{B}(i)}^t > \sqrt{B_i} > 0$. Per Lemma B.2, we have

$$\begin{aligned}\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, g_i^t > B_i] &= \mathbb{E}[(Z_{\mathcal{B}(i)}^{t+1})^2 - (Z_{\mathcal{B}(i)}^t)^2 \mid Q^t, g_i^t > B_i] \\ &= -\eta_i Z_{\mathcal{B}(i)}^t + \delta_i \\ &= -\eta_i \sqrt{g_i^t} + \delta_i.\end{aligned}$$

  Choose any $\gamma_i \in (0, \eta_i)$ and update B_i to a sufficiently large constant such that $\sqrt{g_i^t} > \sqrt{B_i} \geq \dfrac{\delta_i}{\eta_i - \gamma_i}$. This implies

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, g_i^t > B_i] = -\eta_i \sqrt{g_i^t} + \delta_i \leq -\gamma_i \sqrt{g_i^t}.$$

- **2: ($g_i^t \leq B_i$, where B_i is chosen as in case 1).** Per Lemma B.1, we have

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, g_i^t \leq B_i] \leq \theta_i,$$

  for some constant $\theta_i > 0$, which depends only on n and B_i.

Combining both cases above, we have

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t] \leq -\gamma_i \sqrt{g_i^t} + \theta_i,$$

for all $t \geq 0$, where θ_i is a redefined constant. This concludes the basis of the induction.

Inductive step. Assume that the induction hypothesis holds for all $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_i \leq d$, $d \geq 1$. Consider $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_i = d + 1$. By the induction hypothesis, for all $j \in \mathcal{B}(i)$, there exist constants $\alpha_j, \gamma_j, B_j, \theta_j > 0$ such that

$$\mathbb{E}[g_j^{t+1} - g_j^t \mid Q^t] \leq -\gamma_j \sqrt{g_j^t} + \theta_j,$$

for all $t \geq 0$, since $d_j \leq d$.

For now, fix some $\alpha_i > 0$ and $B_i > 0$ at the beginning of each of the following cases. These constants are place holders, and their values will be determined at the end of each case analysis. It might be helpful here to point out that α_i will be a function of γ_i, B_i will be a function of α_i, and θ_i will be a function of B_i. We divide the analysis into three cases: (1) $g_i^t > B_i$ and $|Z_{\mathcal{B}(i)}^t| \leq \sum_{j\in\mathcal{B}(i)} \sqrt{g_j^t}$, (2) $g_i^t > B_i$ and $|Z_{\mathcal{B}(i)}^t| > \sum_{j\in\mathcal{B}(i)} \sqrt{g_j^t}$, and (3) $g_i^t \leq B_i$.

- **1: ($g_i^t > B_i$ and $|Z_{\mathcal{B}(i)}^t| \leq \sum_{j\in\mathcal{B}(i)} \sqrt{g_j^t}$).** Denote this case by $\mathcal{E}_1$. Since

  $$g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t > B_i,$$

  and $|Z_{\mathcal{B}(i)}^t| \leq \sum_{j\in\mathcal{B}(i)} \sqrt{g_j^t}$, we can choose B_i sufficiently large so that if $g_i^t > B_i$, then also $g_j^t > B_j$ for at least one $j \in \mathcal{B}(i)$. Define the sets

  $$\mathcal{J}_> := \left\{ j \in \mathcal{B}(i) : g_j^t > B_j \right\} \text{ and } \mathcal{J}_\leq := \mathcal{B}(i) \backslash \mathcal{J}_> = \left\{ j \in \mathcal{B}(i) : g_j^t \leq B_j \right\}.$$

  Since $|\mathcal{B}(i)| \leq |\mathcal{A}| = n$ and $|Q_j^{t+1} - Q_j^t| \leq 1$ for all $j \in \mathcal{B}(i)$, we have $|Z_{\mathcal{B}(i)}^{t+1} - Z_{\mathcal{B}(i)}^t| \leq n$ for all $t \geq 0$. By the definition of $\mathcal{J}_\leq$, $\sum_{j\in\mathcal{J}_\leq} \sqrt{g_j^t} \leq \sum_{j\in\mathcal{B}(i)} \sqrt{B_j} =: U_1$ and

  $$|Z_{\mathcal{B}(i)}^t| \leq \sum_{j\in\mathcal{B}(i)} \sqrt{g_j^t} \leq U_1 + \sum_{j\in\mathcal{J}_>} \sqrt{g_j^t}. \tag{18}$$

  Combining these altogether, we have

  $$\begin{aligned}
  \mathbb{E}[(Z_{\mathcal{B}(i)}^{t+1})^2 - (Z_{\mathcal{B}(i)}^t)^2 \mid Q^t, \mathcal{E}_1] &= \mathbb{E}[(Z_{\mathcal{B}(i)}^{t+1} - Z_{\mathcal{B}(i)}^t)(Z_{\mathcal{B}(i)}^{t+1} + Z_{\mathcal{B}(i)}^t) \mid Q^t, \mathcal{E}_1] \\
  &\leq n\left(n + 2U_1 + 2\sum_{j\in\mathcal{J}_>} \sqrt{g_j^t}\right) \\
  &\leq U_2\left(U_1 + \sum_{j\in\mathcal{J}_>} \sqrt{g_j^t}\right)
  \end{aligned} \tag{19}$$

  for some constant $U_2 > 0$, which depends only on n and U_1. For all $j \in \mathcal{J}_>$, we have by the induction hypothesis that

$$\mathbb{E}[g_j^{t+1} - g_j^t \mid Q^t, \mathcal{E}_1] \leq -\gamma_j \sqrt{g_j^t}, \tag{20}$$

and per Lemma B.1, for all $j \in \mathcal{J}_{\leq}$ we have that

$$\mathbb{E}[g_j^{t+1} - g_j^t \mid Q^t, \mathcal{E}_1] \leq U_3 \tag{21}$$

for some constant $U_3 > 0$, which depends only on n and B_j. We want to show that there exist constants $\alpha_i, \gamma_i, B_i > 0$ such that

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, \mathcal{E}_1] \leq U_2\left(U_1 + \sum_{j \in \mathcal{J}_>} \sqrt{g_j^t}\right) - \alpha_i\left(\sum_{j \in \mathcal{J}_>} \gamma_j \sqrt{g_j^t}\right) + |\mathcal{J}_{\leq}|\alpha_i U_3 \tag{22}$$

$$\leq U_2\left(U_1 + \sum_{j \in \mathcal{J}_>} \sqrt{g_j^t}\right) - \alpha_i\left(\sum_{j \in \mathcal{J}_>} \gamma_j \sqrt{g_j^t}\right) + n\alpha_i U_3 \tag{23}$$

$$\leq -\gamma_i \sqrt{g_i^t} \tag{24}$$

holds, where (22) follows from (19), (20) and (21), and (23) follows from the fact that $|\mathcal{J}_{\leq}| \leq |\mathcal{A}| = n$. First, we have that

$$\begin{aligned}\sqrt{g_i^t} &\leq |Z_{\mathcal{B}(i)}^t| + \sqrt{\alpha_i}\sqrt{g_{\mathcal{B}(i)}^t} \leq \left(\sum_{j \in \mathcal{B}(i)} \sqrt{g_j^t}\right) + \sqrt{\alpha_i}\left(\sum_{j \in \mathcal{B}(i)} \sqrt{g_j^t}\right) \\ &\leq \left(1 + \sqrt{\alpha_i}\right)\left(U_1 + \sum_{j \in \mathcal{J}_>} \sqrt{g_j^t}\right),\end{aligned}$$

where the first inequality follows from the definition $g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t$, the second inequality follows from the requirement of this case (case 1), and the last inequality follows from (18). Thus, for any $\gamma_i > 0$, we have

$$-\gamma_i\left(1 + \sqrt{\alpha_i}\right)\left(U_1 + \sum_{j \in \mathcal{J}_>} \sqrt{g_j^t}\right) \leq -\gamma_i \sqrt{g_i^t}. \tag{25}$$

Fixing an arbitrary $\gamma_i^{\text{Case 1}} > 0$, take $\alpha_i^{\text{Case 1}} > 0$ sufficiently large such that $\alpha_i^{\text{Case 1}}(\min_{j \in \mathcal{J}_>} \gamma_j) > \alpha_i^{\text{Case 1}}(\min_{j \in \mathcal{B}(i)} \gamma_j) > \left(U_2 + \gamma_i^{\text{Case 1}}\left(1 + \sqrt{\alpha_i^{\text{Case 1}}}\right)\right)$. Setting B_i to a sufficiently large constant $B_i^{\text{Case 1}}$ (recalling that $g_j^t > B_j$ for all $j \in \mathcal{J}_>$), we have the inequality

$$\alpha_i^{\text{Case 1}}\Big(\sum_{j\in\mathcal{J}_>}\gamma_j\sqrt{g_j^t}\Big) - \Big(U_2 + \gamma_i^{\text{Case 1}}\Big(1+\sqrt{\alpha_i^{\text{Case 1}}}\Big)\Big)\Big(U_1 + \sum_{j\in\mathcal{J}_>}\sqrt{g_j^t}\Big) \\ - n\alpha_i^{\text{Case 1}}U_3 \geq 0, \tag{26}$$

since $\sum_{j\in\mathcal{J}_>}\sqrt{g_j^t}$ can be made sufficiently large.

Thus, if we update the previously fixed constants α_i and B_i to $\alpha_i^{\text{Case 1}}$ and $B_i^{\text{Case 1}}$, respectively, then (26) implies that the left hand side of (25) is greater than or equal to the right hand side of (22). Therefore, (24) holds for $\gamma_i^{\text{Case 1}}$, $\alpha_i^{\text{Case 1}}$ and $B_i^{\text{Case 1}}$.

- **2: ($g_i^t > B_i$ and $|Z_{\mathcal{B}(i)}^t| > \sum_{j\in\mathcal{B}(i)}\sqrt{g_j^t}$).** Denote this case by $\mathcal{E}_2$. We claim that in this case we have

$$(i)\ Q_{\mathcal{B}(i)}^t > 0,\ \text{and}\ (ii)\ Z_{\mathcal{B}(i)}^t > 0, \tag{27}$$

where both claims are proven in the end of this case. We have $Q_{\mathcal{B}(i)}^t > 0$ by (27)(i), and clearly $\delta_i = \sum_{j\in\mathcal{A}(\mathcal{T}_i)}\lambda_j \leq \sum_{j\in\mathcal{A}}\lambda_j = 1$. Thus, Lemma B.2 yields

$$\mathbb{E}[(Z_{\mathcal{B}(i)}^{t+1})^2 - (Z_{\mathcal{B}(i)}^t)^2 \mid Q^t, \mathcal{E}_2] \leq -\eta_i Z_{\mathcal{B}(i)}^t + 1, \tag{28}$$

where recall that $\eta_i = 2y^i\lambda$. By the induction hypothesis, for all $j \in \mathcal{J}_>$, we have

$$\mathbb{E}[g_j^{t+1} - g_j^t \mid Q^t, \mathcal{E}_2] \leq -\gamma_j\sqrt{g_j^t}, \tag{29}$$

and per (21), for all $j \in \mathcal{J}_\leq$ we have that

$$\mathbb{E}[g_j^{t+1} - g_j^t \mid Q^t, \mathcal{E}_2] \leq U_3. \tag{30}$$

Similar to the previous case, we want to show that there exists constants $\alpha_i, \gamma_i, B_i > 0$ such that

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, \mathcal{E}_2] \leq -\eta_i Z^t_{\mathcal{B}(i)} + 1 - \alpha_i\bigg(\sum_{j\in\mathcal{J}_>} \gamma_j\sqrt{g_j^t}\bigg) + |\mathcal{J}_\leq|\alpha_i U_3 \tag{31}$$

$$\leq -\eta_i Z^t_{\mathcal{B}(i)} + 1 - \alpha_i\bigg(\sum_{j\in\mathcal{J}_>} \gamma_j\sqrt{g_j^t}\bigg) + n\alpha_i U_3 \tag{32}$$

$$\leq -\gamma_i\sqrt{g_i^t} \tag{33}$$

holds, where (31) follows from (28), (29) and (30), and (32) follows from the fact that $|\mathcal{J}_\leq| \leq |\mathcal{A}| = n$.

Recall from the previous case that $\sum_{j\in\mathcal{J}_\leq}\sqrt{g_j^t} \leq U_1$. By (27)(ii), we have $Z^t_{\mathcal{B}(i)} > 0$ and

$$\sqrt{g_i^t} \leq |Z^t_{\mathcal{B}(i)}| + \sqrt{\alpha_i}\sqrt{\sum_{j\in\mathcal{B}(i)} g_j^t} = Z^t_{\mathcal{B}(i)} + \sqrt{\alpha_i}\sqrt{\sum_{j\in\mathcal{B}(i)} g_j^t}$$
$$\leq Z^t_{\mathcal{B}(i)} + \sqrt{\alpha_i}\bigg(\sum_{j\in\mathcal{B}(i)}\sqrt{g_j^t}\bigg) \leq Z^t_{\mathcal{B}(i)} + \sqrt{\alpha_i}\bigg(U_1 + \sum_{j\in\mathcal{J}_>}\sqrt{g_j^t}\bigg).$$

Thus, for any $\gamma_i > 0$ we have that

$$-\gamma_i\bigg(Z^t_{\mathcal{B}(i)} + \sqrt{\alpha_i}\bigg(U_1 + \sum_{j\in\mathcal{J}_>}\sqrt{g_j^t}\bigg)\bigg) \leq -\gamma_i\sqrt{g_i^t}. \tag{34}$$

Pick an arbitrary $\gamma_i^{\text{Case 2}} \in (0, \eta_i)$ and $\alpha_i^{\text{Case 2}} > 0$ such that $\alpha_i^{\text{Case 2}}(\min_{j\in\mathcal{J}_>}\gamma_j) > \alpha_i^{\text{Case 2}}(\min_{j\in\mathcal{B}(i)}\gamma_j) > \gamma_i^{\text{Case 2}}\sqrt{\alpha_i^{\text{Case 2}}}$. Note that $Z^t_{\mathcal{B}(i)}$ can be made arbitrarily large by updating B_i to a sufficiently large constant $B_i^{\text{Case 2}}$, since $g_i^t = (Z^t_{\mathcal{B}(i)})^2 + \alpha_i(g^t_{\mathcal{B}(i)}) > B_i$ and $|Z^t_{\mathcal{B}(i)}| = Z^t_{\mathcal{B}(i)} > \sum_{j\in\mathcal{B}(i)}\sqrt{g_j^t} \geq \sqrt{g^t_{\mathcal{B}(i)}}$ by the assumption of this case (case 2). Hence, the following inequality holds:

$$\begin{aligned}(\eta_i - \gamma_i^{\text{Case 2}})Z^t_{\mathcal{B}(i)} &+ \alpha_i^{\text{Case 2}}\bigg(\sum_{j\in\mathcal{J}_>}\gamma_j\sqrt{g_j^t}\bigg)\\ &- \gamma_i^{\text{Case 2}}\sqrt{\alpha_i^{\text{Case 2}}}\bigg(U_1 + \sum_{j\in\mathcal{J}_>}\sqrt{g_j^t}\bigg) - 1 - n\alpha_i^{\text{Case 2}}U_3 \geq 0.\end{aligned} \tag{35}$$

Thus, if we update the previously fixed constants α_i and B_i to $\alpha_i^{\text{Case 2}}$ and $B_i^{\text{Case 2}}$, respectively, then (35) implies that the left hand side of (34) is greater than or equal to the right hand side of (31). Therefore, (33) holds for $\gamma_i^{\text{Case 2}}, \alpha_i^{\text{Case 2}}$ and $B_i^{\text{Case 2}}$.

We now prove the two claims in (27).

- **Claim (i).** Note that by the construction of the surplus vectors $y^i,\ i \in \mathcal{A}$, we have

$$Z^t_{\mathcal{B}(i)} = \sum_{j\in\mathcal{B}(i)} Z^t_j = \sum_{j\in\mathcal{B}(i)} y^j Q^t = \sum_{j\in\mathcal{B}(i)} Q^t_j - \sum_{j\in\mathcal{B}(i)} \sum_{k\in\mathcal{B}(j)} y^k Q^t = Q^t_{\mathcal{B}(i)} - \sum_{j\in\mathcal{B}(i)} Z^t_{\mathcal{B}(j)},$$

  which implies

$$|Z^t_{\mathcal{B}(i)}| \leq Q^t_{\mathcal{B}(i)} + \sum_{j\in\mathcal{B}(i)} |Z^t_{\mathcal{B}(j)}|.$$

  By our simple observation (17), we have $\sqrt{g^t_j} \geq |Z^t_{\mathcal{B}(j)}|$ for all $j \in \mathcal{B}(i)$, which yields $\sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j} \geq \sum_{j\in\mathcal{B}(i)} |Z^t_{\mathcal{B}(j)}|$. Since $|Z^t_{\mathcal{B}(i)}| > \sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j}$ by the assumption of this case (case 2), we have

$$0 \leq \sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j} < |Z^t_{\mathcal{B}(i)}| \leq Q^t_{\mathcal{B}(i)} + \sum_{j\in\mathcal{B}(i)} |Z^t_{\mathcal{B}(j)}| \leq Q^t_{\mathcal{B}(i)} + \sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j},$$

  which implies $Q^t_{\mathcal{B}(i)} > 0$ as stated.

- **Claim (ii).** Note that we cannot have $Z^t_{\mathcal{B}(i)} = 0$ by the assumption of this case (case 2). Assume to the contrary that $Z^t_{\mathcal{B}(i)} < 0$. Then the assumption of this case yields

$$Z^t_{\mathcal{B}(i)} < -\sum_{j\in\mathcal{B}(i)} \sqrt{g^t_i}. \tag{36}$$

  Since $Q^t_{\mathcal{B}(i)} > 0$ per (27)(i), and $\sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j} \geq \sum_{j\in\mathcal{B}(i)} |Z_{\mathcal{B}(j)}(t)|$ per (17), we have

$$Z^t_{\mathcal{B}(i)} = Q^t_{\mathcal{B}(i)} - \sum_{j\in\mathcal{B}(i)} Z^t_{\mathcal{B}(j)} \geq Q^t_{\mathcal{B}(i)} - \sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j} \geq -\sum_{j\in\mathcal{B}(i)} \sqrt{g^t_j},$$

  which contradicts (36). Therefore, we have $Z^t_{\mathcal{B}(i)} > 0$ as stated.

- **3: ($g^t_i \leq B_i$).** Denote this event by $\mathcal{E}_3$. Per Lemma B.1, we have

$$\mathbb{E}[g^{t+1}_i - g^t_i \mid Q^t, \mathcal{E}_3] \leq \theta_i,$$

  for some constant $\theta_i > 0$, which depends only on n and B_i.

Combining cases 1-3. Let $\gamma_i = \min\left\{\gamma_i^{\text{Case 1}}, \gamma_i^{\text{Case 2}}\right\}$, $\alpha_i = \max\left\{\alpha_i^{\text{Case 1}}, \alpha_i^{\text{Case 2}}\right\}$ and $B_i =$

$\max\left\{B_i^{\text{Case 1}}, B_i^{\text{Case 2}}\right\}$. We can then write

$$\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t] \leq -\gamma_i\sqrt{g_i^t} + \theta_i,$$

for all $t \geq 0$, where θ_i is a redefined constant. Hence, the induction hypothesis also holds for $i \in \mathcal{A}\backslash\mathcal{U}_0$ with $d_i = d+1$. This concludes the proof for the drift condition (16).

For the ergodicity result, note that the Markov chain $(Q_i^t : i \in \mathcal{Q}_0, t \geq 0)$ is clearly homogeneous, irreducible and aperiodic. Since we already established the drift condition (16), the first two conditions of Proposition B.3 hold. Since $|\mathcal{A}| = n < \infty$, the third condition of Lemma B.3 is also satisfied. Thus, the Markov chain is ergodic. $\square$

We conclude this section with the proofs of Propositions B.1 and B.2.

Proof of Proposition B.1. We use strong induction on d_i, $i \in \mathcal{A}\backslash\mathcal{U}_0$, where recall that $d_i = \max_{j\in\mathcal{U}_0\cap\mathcal{A}(\mathcal{T}_i)} d(i,j)$.

Basis. Consider $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_i = 1$. This implies $\mathcal{B}(i) \subseteq \mathcal{U}_0$ so that $y^jQ^t = Q_j^t$ for all $j \in \mathcal{B}(i)$. By definition, since $g_j^t = 0$ for all $j \in \mathcal{B}(i) \subseteq \mathcal{U}_0$, we have

$$g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t = \left(\sum_{j\in\mathcal{B}(i)} y^jQ^t\right)^2 = (Q_{\mathcal{B}(i)}^t)^2.$$

Since $|\mathcal{B}(i)| \leq |\mathcal{A}| = n$ and $|Q_j^{t+1} - Q_j^t| \leq 1$ for all $j \in \mathcal{B}(i)$, we have $|Q_{\mathcal{B}(i)}^{t+1} - Q_{\mathcal{B}(i)}^t| \leq n$ for all $t \geq 0$. By the assumption of this proposition, we have $\sqrt{g_i^t} = Q_{\mathcal{B}(i)}^t \leq \sqrt{B_i}$. Combining these altogether, we have

$$\begin{aligned}
\mathbb{E}[g_i^{t+1} - g_i^t \mid Q^t, g_i^t \leq B_i] &= \mathbb{E}[(Q_{\mathcal{B}(i)}^{t+1})^2 - (Q_{\mathcal{B}(i)}^t)^2 \mid Q^t, g_i^t \leq B_i] \\
&= \mathbb{E}[(Q_{\mathcal{B}(i)}^{t+1} - Q_{\mathcal{B}(i)}^t)(Q_{\mathcal{B}(i)}^{t+1} + Q_{\mathcal{B}(i)}^t) \mid Q^t, g_i^t \leq B_i] \\
&\leq n\left(\sqrt{B_i} + \sqrt{B_i} + n\right) =: \theta_i,
\end{aligned}$$

which concludes the basis of the induction.

Inductive step. Assume that the induction hypothesis holds for all $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_j \leq d$, $d \geq 1$. Consider $i \in \mathcal{A}\backslash\mathcal{U}_0$ such that $d_i = d+1$. Recall that $g_i^t = (Z_{\mathcal{B}(i)}^t)^2 + \alpha_i g_{\mathcal{B}(i)}^t \leq B_i$. By the induction hypothesis, since $d_j \leq d$ for all $j \in \mathcal{B}(i)$, we have

$$\alpha_i\mathbb{E}[g_{\mathcal{B}(i)}^{t+1} - g_{\mathcal{B}(i)}^t \mid Q^t, g_i^t \leq B_i] \leq U$$

for some constant $U > 0$, which depends only on α_i and B_i. Since $|Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)}| \leq n$ (see case 1 in the proof of Lemma B.1) and $|Z^t_{\mathcal{B}(i)}| \leq \sqrt{B_i}$, we have

$$\begin{aligned}\mathbb{E}[(Z^{t+1}_{\mathcal{B}(i)})^2 - (Z^t_{\mathcal{B}(i)})^2 \mid Q^t, g^t_i \leq B_i] &= \mathbb{E}[(Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)})(Z^{t+1}_{\mathcal{B}(i)} + Z^t_{\mathcal{B}(i)}) \mid Q^t, g^t_i \leq B_i] \\ &\leq n\left(\sqrt{B_i} + \sqrt{B_i} + n\right),\end{aligned}$$

which implies

$$\mathbb{E}[g^{t+1}_i - g^t_i \mid Q^t, g^t_i \leq B_i] \leq n\left(\sqrt{B_i} + \sqrt{B_i} + n\right) + U =: \theta_i.$$

Hence, the induction hypothesis holds for $i \in \mathcal{A}\backslash\mathcal{U}_0$ with $d_i = d + 1$. □

Proof of Proposition B.2. For each match $m \in \mathcal{M}$ with $\mathcal{A}(m) = \{i, j\}$, let $D^t_{i,j}$ be the number of times match m is performed under $SP(\mathcal{M}_+, p)$ by time t. We first claim that for all $i \in \mathcal{A}$ and $t \geq 0$, we have

$$Z^t_i = y^i A^t - D^t_{i,i^\uparrow}. \tag{37}$$

The proof of this claim is given in the end of the current proof. (37) implies that $Z^t_{\mathcal{B}(i)} = \sum_{j\in\mathcal{B}(i)} Z^t_j = \sum_{j\in\mathcal{B}(i)}(y^j A^t - D^t_{j,i})$. Thus, for all $t \geq 0$, we have

$$Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)} = \sum_{j\in\mathcal{B}(i)} (y^j \Delta A^t - \Delta D^t_{j,i}), \tag{38}$$

where $\Delta A^t = A^{t+1} - A^t$ and $\Delta D^t_{j,i} = D^{t+1}_{j,i} - D^t_{j,i}$. Since $Q^t_{\mathcal{B}(i)} > 0$ by the assumption of this lemma, we must have $Q^t_i = 0$ (otherwise a match would be executed between i and some $j \in \mathcal{B}(i)$ due to the greedy nature of the policy). Per (38), $Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)} \neq 0$ if and only the arriving agent at time $t+1$ is of type k, $k \in \mathcal{A}(\mathcal{T}_i)$. Consider such an arrival:

- If $d(i,k)$ is odd, then by the construction of the surplus vectors, ΔA^t_k has a positive sign in the summation (38). Since there is a single agent arrival per period, $\Delta A^t_l = 0$ for all $l \in \mathcal{A}(\mathcal{T}_i)\backslash\{k\}$. Since $Q^t_i = 0$, no matches with agent type j can be performed so that also $\Delta D^t_{j,k} = 0$ for all $j \in \mathcal{B}(i)$. Overall, we have $Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)} = \Delta A^t_k = 1$.

- If $d(i,k)$ is even and $k \neq i$, then ΔA^t_k has a negative sign in the summation (38). Following the same argument as above, we conclude that $Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)} = \Delta A^t_k = -1$.

- If $k = i$, then $\Delta A^t_j = 0$ for all $j \in \mathcal{B}(i)$. Since $Q^t_{\mathcal{B}(i)} > 0$, agent type i will be matched with some $l \in \mathcal{B}(i)$ upon arrival; see Remark B.1. Thus, $\Delta D^t_{l,i} = 1$ and $\Delta D^t_{j,i} = 0$ for all

$j \in \mathcal{B}(i) \backslash \{l\}$. Plugging into (38), we have that $Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)} = -1$.

An agent of type k arrives at time $t+1$ with probability λ_k so that we can conclude the following from the three cases above:

$$\begin{aligned}
&\mathbb{E}[(Z^{t+1}_{\mathcal{B}(i)})^2 - (Z^t_{\mathcal{B}(i)})^2 \mid Q^t, Q^t_{\mathcal{B}(i)} > 0] = \mathbb{E}[(Z^{t+1}_{\mathcal{B}(i)} - Z^t_{\mathcal{B}(i)})(Z^{t+1}_{\mathcal{B}(i)} + Z^t_{\mathcal{B}(i)}) \mid Q^t, Q^t_{\mathcal{B}(i)} > 0] \\
&= -\lambda_i(2Z^t_{\mathcal{B}(i)} - 1) + \sum_{\substack{k \in \mathcal{A}(\mathcal{T}_i)\backslash\{i\}: \\ d(i,k) \text{ is odd}}} \lambda_k (2Z^t_{\mathcal{B}(i)} + 1) - \sum_{\substack{k \in \mathcal{A}(\mathcal{T}_i)\backslash\{i\}: \\ d(i,k) \text{ is even}}} \lambda_k (2Z^t_{\mathcal{B}(i)} - 1) \\
&= -2\Bigg(\lambda_i - \sum_{\substack{k \in \mathcal{A}(\mathcal{T}_i)\backslash\{i\}: \\ d(i,k) \text{ is odd}}} \lambda_k + \sum_{\substack{k \in \mathcal{A}(\mathcal{T}_i)\backslash\{i\}: \\ d(i,k) \text{ is even}}} \lambda_k\Bigg) Z^t_{\mathcal{B}(i)} + \sum_{k \in \mathcal{A}(\mathcal{T}_i)} \lambda_k \\
&= (-2y^i\lambda) Z^t_{\mathcal{B}(i)} + \sum_{k \in \mathcal{A}(\mathcal{T}_i)} \lambda_k \\
&= -\eta_i Z^t_{\mathcal{B}(i)} + \delta_i.
\end{aligned}$$

as stated. Now we prove the claim (37). We use strong induction on d_i, $i \in \mathcal{A}$, where recall that $d_i = \max_{j \in \mathcal{U}_0 \cap V(\mathcal{T}_i)} d(i,j)$.

Basis. Consider $i \in \mathcal{A}$ such that $d_i = 0$. This implies $i \in \mathcal{U}_0$ so that $Z^t_i = y^i Q^t = Q^t_i = (A^t_i - D^t_{i,i^\uparrow}) = y^i A^t - D^t_{i,i^\uparrow}$ as required. This concludes the basis of the induction.

Inductive step. Assume that the induction hypothesis holds for all $i \in \mathcal{A}$ such that $d_i \le d$, $d \ge 0$. Consider $i \in \mathcal{A}$ such that $d_i = d+1$. By the induction hypothesis, for all $j \in \mathcal{B}(i)$, we have

$$Z^t_j = y^j A^t - D^t_{j,i}.$$

By the construction of the surplus vectors, we have $Z^t_i = y^i Q^t = Q_i(t) - \sum_{j \in \mathcal{B}(i)} y^j Q^t = Q^t_i - \sum_{j \in \mathcal{B}(i)} Z^t_j$. Since $Q^t_i = A^t_i - D^t_{i,i^\uparrow} - \sum_{j \in \mathcal{B}(i)} D^t_{j,i}$, we have

$$\begin{aligned}
Z_i(t) &= Q^t_i - \sum_{j \in \mathcal{B}(i)} Z^t_j \\
&= \Bigg(A^t_i - D^t_{i,i^\uparrow} - \sum_{j \in \mathcal{B}(i)} D^t_{j,i}\Bigg) - \sum_{j \in \mathcal{B}(i)} (y^j A^t - D^t_{j,i}) \\
&= \Bigg(A^t_i - \sum_{j \in \mathcal{B}(i)} y^j A^t\Bigg) - \Bigg(\sum_{j \in \mathcal{B}(i)} D^t_{j,i} - \sum_{j \in \mathcal{B}(i)} D^t_{j,i}\Bigg) - D^t_{i,i^\uparrow} \\
&= y^i A^t - D^t_{i,i^\uparrow}.
\end{aligned}$$

Hence, the induction hypothesis holds for $i \in \mathcal{A}$ with $d_i = d+1$. □