The Annals of Applied Probability
2017, Vol. 27, No. 6, 3385–3434
https://doi.org/10.1214/17-AAP1283

ON THE INSTABILITY OF MATCHING QUEUES

BY PASCAL MOYAL AND OHAD PERRY[1]

Université de Technologie de Compiègne and Northwestern University

A matching queue is described via a graph, an arrival process and a matching policy. Specifically, to each node in the graph there is a corresponding arrival process of items, which can either be queued or matched with queued items in neighboring nodes. The matching policy specifies how items are matched whenever more than one matching is possible. Given the matching graph and the matching policy, the stability region of the system is the set of intensities of the arrival processes rendering the underlying Markov process positive recurrent. In a recent paper, a condition on the arrival intensities, which was named NCOND, was shown to be necessary for the stability of a matching queue. That condition can be thought of as an analogue to the usual traffic condition for traditional queueing networks, and it is thus natural to study whether it is also sufficient. In this paper, we show that this is not the case in general. Specifically, we prove that, except for a particular class of graphs, there always exists a matching policy rendering the stability region strictly smaller than the set of arrival intensities satisfying NCOND. Our proof combines graph- and queueing-theoretic techniques: After showing explicitly, via fluid-limit arguments that the stability regions of two basic models is strictly included in NCOND, we generalize this result to any graph inducing either one of those two basic graphs.

CONTENTS

Received June 2015; revised January 2017.
[1]Supported by NSF Grant CMMI 1436518.
MSC2010 subject classifications. Primary 60K25; secondary 60F17.
Key words and phrases. Matching queues, instability, fluid limits, graphs.

1. Introduction.

We consider a continuous-time matching queueing system in which items of different classes arrive one by one and depart in pairs. Specifically, we assume that any item is either matched with exactly one other item immediately upon arrival, and both items leave the system, or is stored in a buffer until it is matched. Since matchings are pairwise, such a matching model can be represented via an undirected graph, in which each node represents a class of arrivals, and an edge between two nodes represents that items of the two corresponding classes can be matched together. A *matching policy* describes the matching rule whenever more than one matching is possible for an incoming item.

The model just described is closely related to the discrete-time *stochastic matching model* introduced in [23], in which items enter the system at discrete time points, and their class is drawn upon arrival from a given probability distribution on the set of classes. We elaborate on the relation between the two models below.

Matching models in the literature. Matching queueing models arise directly in several applications, such as organ transplantation [12] and public-housing assignments [31]. They were also employed in the literature as relaxations for complex many-server queueing systems [1, 9], stochastic processing networks and assemble-to-order systems; see [17]. We refer to these references for comprehensive literature reviews of related models.

A more recent application for matching queues is in modeling *sharing-economy* (or *collaborative consumption*) platforms, with the most relevant examples being car-sharing platforms such as Uber and Lyft, lodging services such as Airbnb and virtual call centers (namely call centers with home-based agents), as considered in, for example, [16, 18]. Since a platform operating in a sharing-economy market must match supply and demand at every instance, possibly in a multiregion setting, matching queues can be used to model and optimize such platforms; see [30] for a recent application in the car-sharing setting.

The term *matching queues* was introduced in [17]. In that reference, items can be matched by groups of size two or more, and the goal is to minimize finite-horizon cumulative holding costs. Moreover, the controller can keep matchable items in storage for more "profitable" future matches. A myopic, discrete-review control is shown to be asymptotically optimal, as the arrival rate grows large. Thus, both the model and the objectives of [17] are different than ours here, since we consider pairwise matchings, and are concerned with stability properties.

The stability of matching models operating under the *first-come first-served* (FCFS) policy was studied for several particular graphs in [9], assuming that both the arrivals and the departures occur by pairs. (If items arrive one by one, the model can never be stable, as will become clear below.) A discrete-time Markov chain representation was employed in combination with Lyapunov techniques, to study the stability (ergodicity) of certain models having bipartite graphs. An important (and expected) observation from this latter reference is that characterizing the stability region, namely, the law of the arrival process under which the system is stable, is nontrivial even for models with relatively simple graphs. The matching model in [9] was further analyzed in [8]. An alternative Markov representation was introduced, leading to a more complete picture of the stability problem. General sufficient and necessary conditions for stability of the underlying Markov chain were given, together with properties of several matching policies. In particular, the stability of any model applying the "match the longest" policy (under which an item that has more than one matching option upon arrival is matched with an item from the longest queue) was proved, assuming the necessary condition for stability holds.

The necessary condition for stability in [8] was employed in [1] to prove the existence of unique matching rates for models satisfying a certain "complete resource pooling" condition. The models considered in [1] are again bipartite and operate in the FCFS matching policy. Interestingly, the stationary distribution of the Markov chain in [1] is shown to have a product form.

In [2], a continuous-time model is considered for a bipartite matching system operating under the FCFS–ALIS (Assign Longest-Idle Server) policy. In this paper, Markovian service system (i.e., service times are considered) with skill-based routing, are modeled as matching queues, and the stationary distribution, when it exists, is shown to have a product form. Fluid limits are employed in the overloaded case to prove the existence of a local steady state.

The graphs in [1, 2, 8, 9] are all bipartite. Therefore, to make the question of stability nontrivial, items are assumed to arrive in pairs, as was mentioned above. This assumption is dropped in [23], which introduces the aforementioned (discrete-time) stochastic matching model with general graph topology. In [23], a thorough study of the structure of the stability region of the model is proposed, partially relying on the results in [8]. A natural necessary condition, named NCOND, for the stability of *any* such stochastic matching model is introduced, implying, in particular, that no model can be stable if the matching graph is a tree, or more generally, a bipartite graph. (This explains the assumption that arrivals occur in pairs in the papers dealing with bipartite graphs cited above. If items arrive one by one, the system cannot be stable.) In addition, a particular class of graphs is exhibited (the non-bipartite *separable* ones—see Definition 2 below), for which NCOND is also a sufficient condition for stability. However, the study of a particular model on a nonseparable graph (see [23], page 14) shows that NCOND is not sufficient in general for nonseparable graphs. This raises the question of whether the sufficiency of NCOND is true *only* for separable graphs. The Lyapunov-stability techniques that were employed for the particular model in [23], render the generalization of the arguments in [23] to a larger class of nonseparable graphs impractical in the discrete-time settings.

Since a matching queue As is easily seen, and will be shown in Theorem 2 below, the stability region of a discrete-time stochastic model can be studied by embedding it in an appropriate continuous-time model. Thus, the continuous-time counterparts of the results in [23] hold for our matching queues, and vice versa. The advantage of the continuous-time setting is that powerful fluid-limit techniques can be employed, which greatly facilitate the stability analysis.

Stability of stochastic networks via fluid limits. The necessary condition for stability NCOND, defined in (5) below, can be thought of as an analogue to the usual traffic condition of standard queueing networks, requiring that the long-term rate of arrivals to each service station be less than the long-run output rate at that station. Therefore, our work relates to the literature on (in)stability of subcritical stochastic networks, which we briefly review.

Consider a stochastic queueing network with $d \geq 1$ service stations and $K \geq 1$ classes. Let a_i^k denote the (long-run) arrival rate of class-k jobs into service station i, and let m_i^k denote the mean service time of class-k jobs in this station, $1 \leq k \leq K$. Then the system is subcritical if

$$\rho_i := \sum_k a_i^k m_i^k < 1 \qquad \text{for all } 1 \leq i \leq d. \tag{1}$$

It is well known that Condition (1) is not sufficient to ensure stability of stochastic networks in general. The first examples of this fact are the deterministic Lu–Kumar network [21] (and its stochastic counterpart; Section 3 in [7]), and the Rybko–Stolyar network [29], both of which consider static priority service policies. A subcritical, multiclass, two-station network, having a Poisson arrival process and exponential service times in both service stations for all classes, was shown to be unstable under the FIFO discipline in [6]. See [7] for an elaborate discussion, including a comprehensive literature review of the subject.

Fluid models are arguably the most effective tool to proving that a queueing network is stable, and can also be employed to prove instability of such networks. Specifically, following [29], Dai [10] showed that, under mild regularity conditions, if all the (subsequential) fluid limits of the queues, for all possible initial conditions, converge to 0 in finite time w.p.1, then the system is stable, in the sense that the underlying queue process is positive Harris recurrent. We rely on [10] to characterize the stability region of specific matching queues (see Proposition 2 and Corollary 3 below), but our main result is concerned with proving a general *instability* result. Regarding the use of fluid limits to prove instability of stochastic networks, we mention that partial converse results to [10] exist, for example, [11, 15, 24]. Here, we build on the theory in [28], Chpter 9, to prove our main result by characterizing fluid limits uniquely for appropriate initial conditions, and showing that those fluid limits do not decrease to the origin in finite time; see Lemma 1 below.

An interesting feature of the fluid limits we obtain is that their dynamics are determined by the stationary distribution of a "fast" CTMC. Specifically, if the fluid queue associated with one of the nodes is positive, then the relevant time scale for this queue is slower than the time scale for the fluid queues that are null. In the limit, the effect of the "fast" (i.e., null) queues on the evolution of the positive fluid queues is averaged-out instantaneously, a phenomena known as a stochastic *averaging principle* (AP) in the literature. See [26, 27] and the references therein, as well as [22, 36] for recent examples of fast averaging in queueing networks.

Organization of the paper. The rest of the paper is organized as follows. In Section 2, we elaborate on our model and introduce the main notation and terms that will be used. In Section 3, we present our main result, Theorem 3. Section 4 develops the fluid limit (Theorem 4) and Section 5 studies models that are key to the proof of Theorem 3. Theorem 3 is proved in Section 6, building on the results of Sections 4 and 5, and the FWLLN is proved in Section 7. In Section 8, we present a related instability result for an alternative matching policy. Summary of the main paper and directions for future research are presented in Section 9. In addition to the main paper, in the Appendix we demonstrate that our main results have implications to the construction of matchings on random graphs.

2. The model. In this section, we describe the matching queueing model in detail, after introducing the notation and key terms that we employ.

2.1. *Basic terms and notation.* We adopt the usual $\mathbb{R}$ and $\mathbb{Z}$ notation for the sets of real numbers and integers, respectively. We let $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote, respectively, the sets of nonnegative and strictly positive real numbers. Similarly, $\mathbb{Z}_+$ and $\mathbb{Z}_{++}$ denote the sets of nonnegative and strictly positive integers. For any two elements $a, b \in \mathbb{Z}_+$, we let $[\![a, b]\!] := \{a, a+1, \dots, b\}$. For a set A, we let $|A|$, denote the cardinality of A and for any $k \in \mathbb{Z}_{++}$, A^k denotes the set of k-dimensional vectors with components in A. For any $i \in [\![1, k]\!]$, the ith vector of the canonical basis of $\mathbb{R}^k$ is denoted by $\mathbf{e}_i$, namely, $\mathbf{e}_i$ has 1 in its ith coordinate and 0 elsewhere. For any subset $J \subset [\![1, k]\!]$ and $x \in \mathbb{R}^k$, we use the notation x_J for the restriction of x to its coordinates corresponding to the indices of J.

For an interval $I \subset [0, \infty)$, let $\mathbb{D}^d(I)$ denote the space of $\mathbb{R}^d$-valued functions on I that are right continuous and have limits from the left everywhere, endowed with the standard Skorohod J_1 topology [4]. To simplify notation, we write, for example, $\mathbb{D}^d(a, b)$ instead of $\mathbb{D}^d((a, b))$, and $\mathbb{D}(I)$ for $\mathbb{D}^1(I)$ (we remove the superscript when $d = 1$). We omit the interval from the notation whenever it can be taken to be an arbitrary compact interval, for example, convergence in $\mathbb{D}^d$ holds over compact subintervals of $[0, \infty)$. We write $\mathbb{C}^d(I)$ for the subspace of continuous functions on I, with $\mathbb{C} := \mathbb{C}^1$.

Random variables and processes. We work on the probability space $(\Omega, \mathcal{F}, P)$. We write $\stackrel{\text{d}}{=}$ to denote equality in distribution and $\Rightarrow$ to denote convergence in distribution. For a sequence of real-valued random variables $\{Y^n : n \in \mathbb{Z}_{++}\}$, we write $Y^n \Rightarrow \infty$ if $P(Y^n > M) \to 1$ as $n \to \infty$, for any $M > 0$. The fluid-scaled version of a sequence of stochastic processes $\{Y_n : n \in \mathbb{Z}_{++}\}$ is denoted by $\bar{Y}^n := Y^n/n$.

For two real-valued stochastic processes X and Y, we write $X \le_{\text{st}} Y$ if X is smaller than Y in *sample-path stochastic order*, namely, if it is possible to construct two processes $\tilde{X}$ and $\tilde{Y}$ on a common probability space, such that $\tilde{X} \stackrel{\text{d}}{=} X$, $\tilde{Y} \stackrel{\text{d}}{=} Y$, and the sample paths of $\tilde{X}$ lie below those of $\tilde{Y}$ with probability 1 (w.p.1 for short). When X and Y are $\mathbb{R}^d$-valued, $d > 1$, $X \le_{\text{st}} Y$ means that $X_i \le_{\text{st}} Y_i$, for all $1 \le i \le d$.

Graph-related terminology. In addition to the notation, we introduce basic terms of graph theory that will be used below. A graph G is denoted by $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are the set of nodes and edges, respectively. The nodes of G are labeled arbitrarily in $[\![1, |\mathcal{V}|]\!]$, and we often identify $\mathcal{V}$ with $[\![1, |\mathcal{V}|]\!]$. We write $i - j$ whenever $(i, j) \in \mathcal{E}$, namely, nodes i and j are connected by an edge, and $i \not- j$ otherwise. If $i - j$, then these two nodes are said to be *neighbors*. All the graphs considered in this paper are *nonoriented*, that is, $i - j$ if and only if $j - i$ [and the edges (i, j) and (j, i) are indistinguishable] and *simple*, *that is*, there is no edge connecting a node to itself ($i \not- i$ for any $i \in \mathcal{V}$), and two nodes are connected by at most one edge.

For any subset A of $\mathcal{V}$, we let $\mathcal{E}(A)$ be the set of neighbors of all the nodes in A, that is,

$$\mathcal{E}(A) := \{j \in \mathcal{V} : i - j, \text{ for some } i \in A\}.$$

We write $\mathcal{E}(i)$ for the neighbors of a single node i [instead of $\mathcal{E}(\{i\})$].

The graph $\breve{G} = (\breve{\mathcal{V}}, \breve{\mathcal{E}})$ is said to be a *subgraph* of $G = (\mathcal{V}, \mathcal{E})$ if $\breve{\mathcal{V}} \subset \mathcal{V}$ and $\breve{\mathcal{E}} \subset \mathcal{E}$. We say that G *induces* the subgraph $\breve{G}$, whenever $\breve{\mathcal{E}}$ equals the restriction of $\mathcal{E}$ to $\breve{\mathcal{V}}^2$ (recall that $A^2 = A \times A$ for any set A). In other words, if $(i, j) \in \mathcal{E}$, then $(i, j) \in \breve{\mathcal{E}}$, for all $i, j \in \breve{\mathcal{V}}$. In that case, $\breve{G}$ is said to be *induced by* $\breve{\mathcal{V}}$ *in* G.

For any $p \in \mathbb{Z}_{++}$, a *cycle of length* p (or p-*cycle*, for short) is a graph $G = (\mathcal{V}, \mathcal{E})$ such that $|\mathcal{V}| = p$ and any node in $\mathcal{V}$ has exactly two neighbors. In other words, we can label the nodes of G as $i_1, i_2, \dots, i_p$, such that

$$i_1 - i_2, \qquad i_2 - i_3, \qquad \dots, \qquad i_{p-1} - i_p \quad \text{and} \quad i_p - i_1.$$

We say that the p-cycle is *odd* if its length p is an odd number.

The *complement* graph of G is the graph $\bar{G} = (\bar{\mathcal{V}}, \bar{\mathcal{E}})$ such that $\bar{\mathcal{V}} = \mathcal{V}$ and $\bar{\mathcal{E}} = \mathcal{V}^2 \setminus (\mathcal{D} \cup \mathcal{E})$, where $\mathcal{D}$ is the diagonal of $\mathcal{V}^2$, namely, $\mathcal{D} := \{(i, i) : i \in \mathcal{V}\}$. For $q \geq 2$, the graph $G = (\mathcal{V}, \mathcal{E})$ is said to be q-partite, $q \in \mathbb{Z}_{++}$, if there exists a partition $\{\mathcal{V}_i : 1 \leq i \leq q\}$ of $\mathcal{V}$ such that

$$\mathcal{E} \subset \bigcup_{i,j \in [\![1,q]\!] : i \neq j} \mathcal{V}_i \times \mathcal{V}_j.$$

In other words, in a q-partite graph, every edge links two nodes in two distinct subsets of the partition. A 2-partite graph is called *bipartite*.

The *complete graph* $G = (\mathcal{V}, \mathcal{E})$ is such that $\mathcal{E} = \mathcal{V}^2 \setminus \mathcal{D}$. A *clique* of a graph G is a complete subgraph of G. The graph $G = (\mathcal{V}, \mathcal{E})$ is said to be *connected* if for any $i, j \in \mathcal{V}$, there exists a path from i to j, that is, a subset $\{i = i_1, i_2, \dots, i_q = j\} \subset \mathcal{V}$ such that $i_\ell - i_{\ell+1}$ for any $\ell \in [\![1, q-1]\!]$.

An *independent set* of a graph G is a nonempty subset $\mathcal{I} \subset \mathcal{V}$ such that $i \not- j$, for all $i, j \in \mathcal{I}$. We let $\mathbb{I}(G)$ denote the set of all independent sets of G. Notice that when G is simple, any node is an independent set, so that $\mathbb{I}(G)$ is nonempty. We say that the independent set $\mathcal{I}$ of G is *maximal* if, for any $j \in \mathcal{V} \setminus \mathcal{I}$, we have $i_j - j$ for some i_j in $\mathcal{I}$. In other words, $\mathcal{I} \cup \{j\}$ is not an independent set, for any $j \in \mathcal{V} \setminus \mathcal{I}$.

2.2. *Matching queues.* The *matching queue* associated with a graph $G = (\mathcal{V}, \mathcal{E})$, an arrival-rate vector $\lambda := (\lambda_1, \dots, \lambda_{|\mathcal{V}|})$ and a matching policy Φ, is defined as follows. Each node of the simple graph G (which we call *matching graph*) is associated with a class of items, and items of each class $i \in \mathcal{V}$ arrive to the system in accordance with a Poisson process N_i having intensity $\lambda_i > 0$. We also write

(2) $$\bar{\lambda} := \sum_{i \in \mathcal{V}} \lambda_i \quad \text{and} \quad \bar{\lambda}_A := \sum_{i \in A} \lambda_i, \qquad A \subset \mathcal{V}.$$

We assume that all $|\mathcal{V}|$ Poisson arrival processes are independent. Class-i items can be matched with class-j items if and only if $i - j$, that is, there is an edge between the two nodes $i, j \in \mathcal{V}$. We emphasize that the matching graphs G we consider are simple, so that items from the same class cannot be matched together.

Upon arrival, a class-i item is either matched with exactly one item from a class j such that $i - j$, if any such item is available, or is placed in an infinite buffer. Matched items leave the system immediately. We refer to the buffer content associated with each class i as the class-i queue, and denote the associated class-i queue process by $Q_i := \{Q_i(t) : t \geq 0\}$. Specifically, for all $t \geq 0$, $Q_i(t)$ is the number of the class-i items in queue at time t. Let

$$Q = (Q_1, \dots, Q_{|\mathcal{V}|}) \tag{3}$$

denote the $|\mathcal{V}|$-dimensional queue process of the system. For $t \geq 0$ and $A \subset \mathcal{V}$, we let $Q_A(t)$ be the restriction of $Q(t)$ to its coordinates in A.

Upon arrival to the system, a class-i item may find several possible matches, whenever more than one neighboring class has items queued. A *matching policy* is the rule specifying how to execute matchings in such cases. We say that a matching policy Φ is *admissible* if matchings always occur when possible, and decisions are made solely on the value of the queue process Q at arrival epochs. (We note that a larger class of policies was considered in [23].) Consequently, under an admissible matching policy, the queue process Q is a CTMC and $Q_i(t)Q_j(t) = 0$ for all $i, j \in \mathcal{V}$ such that $i - j$ and all $t \geq 0$.

An admissible matching policy is of *priority* type if for any node i, the set $\mathcal{E}(i)$ is a priori ordered: $\mathcal{E}(i) = (i_1, i_2, \dots, i_{|\mathcal{E}(i)|})$, so that, at any time t in which a class-i item enters the system, matching occurs with a class-i_m item, where $m = \min\{\ell \in [\![1, |\mathcal{E}(i)|]\!] : Q_{i_\ell}(t) > 0\}$. An important example of a nonpriority admissible matching policy is *Match the Longest*, denoted by ML, which was introduced in [8] for the bipartite matching queue, and in [23] for the general matching queue in discrete time. According to ML, an arriving item of class i is matched with an item of the class in $\mathcal{E}(i)$ that has the longest queue at that time, where ties are broken according to a uniform draw.

Clearly, the initial queue length $Q(0)$, together with G, λ and the matching policy Φ fully determine the distribution of Q. We thus characterize the system by the triple $(G, \lambda, \Phi)_{\text{C}}$ (where we append the subscript C to denote a *continuous-time* model, as opposed to the one in discrete time, which will be denoted with a subscript D).

Stability of a matching queue. The matching queue $(G, \lambda, \Phi)_{\text{C}}$ is said to be *stable* if the corresponding CTMC Q is positive recurrent, and unstable otherwise.

DEFINITION 1. The *stability region* corresponding to the connected graph G and the matching policy Φ is the set

$$\{\lambda \in \mathbb{R}_{++}^{|\mathcal{V}|} : (G, \lambda, \Phi)_{\text{C}} \text{ is stable}\}.$$

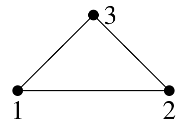

FIG. 1. *Triangle.*

We also say that node $i \in \mathcal{V}$ is unstable if, for some initial condition, the mean time for its associated queue to empty is infinite. Otherwise, the node is stable.

2.3. *The necessary condition for stability* NCOND. It is natural to ask what is the analogue of (1) in the context of matching queues. Clearly, it must hold that, for each node i,

$$\lambda_i < \bar{\lambda}_{\mathcal{E}(i)}. \tag{4}$$

[Recall the notational convention in (2).] However, it is easy to see that (4) can hold for matching queues having bipartite matching graphs, for example, although such models are never stable. (See the discussion following the proof of Theorem 1 below.) Thus, a necessary condition for stability should be stronger than (4).

To gain intuition, we contrast two simple examples for matching graphs, the triangle and the simplest graph including a triangle, which we shall call the "pendant graph", depicted in Figures 1 and 2, respectively. For the triangle, it is straightforward that the corresponding matching queue is stable under (4) for any admissible matching policy. Indeed, at most one of the three queues is positive at any given time, and the drift of any positive queue is necessarily negative under (4). Now consider the pendant graph with the priority matching policy in Figure 2, under which node 3 gives strict priority to nodes 1 and 2, that is, a class-3 arrival who finds items in node 4 and in one of the remaining two nodes, say node 1, will be matched with the class-1 item. This priority policy is depicted by the arrows in Figure 2. Under this policy, node 4 may be unstable despite the fact that $\lambda_4 < \lambda_3$. Indeed, for node 4 to be stable, we must have an adequate number of class-3 arrivals so that, even though $\lambda_3 < \lambda_1 + \lambda_2 + \lambda_4$, sufficiently many class-3 items are left to be matched with all the class-4 items in the long run. Since many class 1

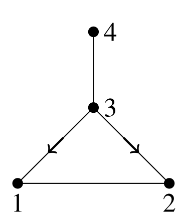

FIG. 2. *Pendant graph.*

and class 2 items will be matched with each other, it is intuitively clear that, in addition to (4), we must require that

$$\lambda_4 + \lambda_1 < \lambda_3 + \lambda_2 \quad \text{and} \quad \lambda_4 + \lambda_2 < \lambda_3 + \lambda_1.$$

This suggests that one needs to consider the arrival rates to subsets of nonneighboring nodes, and require that those rates are smaller than the arrival rates to the neighborhoods of those subsets. Therefore, for any matching graph G, we define

$$\textsc{Ncond}_{\mathrm{C}}(G) := \{\lambda \in \mathbb{R}_{++}^{|\mathcal{V}|} : \bar{\lambda}_{\mathcal{I}} < \bar{\lambda}_{\mathcal{E}(\mathcal{I})} \text{ for all } \mathcal{I} \in \mathbb{I}(G)\},$$

where we recall that $\mathbb{I}(G)$ is the set of independent sets of G. We say that NCOND holds for G and λ if

$$\lambda \in \textsc{Ncond}_{\mathrm{C}}(G). \tag{5}$$

THEOREM 1 (Necessary condition for stability of matching queues). *Let G be a connected graph and Φ an admissible matching policy. Then the stability region corresponding to G and Φ is included in* $\textsc{Ncond}_{\mathrm{C}}(G)$.

Before proving Theorem 1, let us briefly describe the (discrete-time) stochastic matching model introduced in [23]. Given a graph G and an admissible matching policy Φ, the discrete-time stochastic matching model $(G, \mu, \Phi)_{\mathrm{D}}$ is defined similarly to the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$, except that items enter the system one by one, at any discrete time $n \in \mathbb{Z}_{++}$. Assuming that the sequence of classes of the items entering the system is independent and identically distributed with a common probability measure μ on $\mathcal{V}$, $(G, \mu, \Phi)_{\mathrm{D}}$ is represented by the $\mathbb{Z}_+^{|\mathcal{V}|}$-valued Discrete-Time Markov Chain (DTMC) $U := \{U(n) : n \geq 1\}$, where for any $i \in \mathcal{V}$ and any $n \in \mathbb{Z}_{++}$, $U_i(n)$ counts the number of items of class i in the buffer at time n. Then, letting $\{N(t) : t \geq 0\}$ denote the superposition of of the Poisson arrival processes $N_1, \dots, N_{|\mathcal{V}|}$, we have $Q(t) = U(N(t))$, $t \geq 0$, due to uniformization, implying that Q is positive recurrent if and only if U is.

PROOF OF THEOREM 1. Define the following set of probability measures μ on $\mathcal{V}$:

$$\textsc{Ncond}_{\mathrm{D}}(G) := \{\mu \text{ with support } \mathcal{V} : \mu(\mathcal{I}) < \mu(\mathcal{E}(\mathcal{I})) \text{ for all } \mathcal{I} \in \mathbb{I}(G)\},$$

and for any λ, define the probability measure

$$\mu_\lambda(i) := \lambda_i / \bar{\lambda}, \qquad i \in \mathcal{V}. \tag{6}$$

Then for any graph G it holds that, if $\lambda \in \textsc{Ncond}_{\mathrm{C}}(G)$, then $\mu_\lambda \in \textsc{Ncond}_{\mathrm{D}}(G)$. The statement of the theorem thus follows from Proposition 2 in [23], and the fact that Q is positive recurrent if and only if the DTMC U corresponding to $(G, \mu, \Phi)_{\mathrm{D}}$ is. $\square$

An immediate consequence of Theorem 1 is that matching queues $(G, \lambda, \Phi)_C$ having a connected bipartite graph G are never stable. Indeed, if $\mathcal{I}_1 \cup \mathcal{I}_2$ denotes the bipartition of $\mathcal{V}$ into maximal independent sets, then (5) implies that $\bar{\lambda}_{\mathcal{I}_1} < \bar{\lambda}_{\mathcal{I}_2}$ and $\bar{\lambda}_{\mathcal{I}_1} > \bar{\lambda}_{\mathcal{I}_2}$, so that $\text{NCOND}_C(G)$ is empty. In (i) of Theorem 2 below, we show that the converse of this result also holds.

In ending, we remark that (5) is equivalent to (4) for the triangle in Figure 1 or, more generally, for any complete graph. Aside from this case, condition (5) is always strictly stronger (and harder to verify) than (4). It is therefore significant that it can be verified in $O(|\mathcal{V}|^3)$ time; see Proposition 1 in [23].

3. The main result.

3.1. *Separable graphs*. The notion of a *separable graph*, introduced in [23], will play a crucial role in what follows.

DEFINITION 2. A graph $G = (\mathcal{V}, \mathcal{E})$ is said to be *separable of order* q, $q \geq 2$, if there exists a partition of $\mathcal{V}$ into maximal independent sets $\mathcal{I}_1, \dots, \mathcal{I}_q$, such that $u - v$ for all $u \in \mathcal{I}_i$ and $v \in \mathcal{I}_j$, for all $i \neq j$.

Equivalently, G is separable of order q if its complement graph can be partitioned into q disjoint cliques. Notice that a separable graph of order 2 is bipartite, whereas a separable graph or order 3 or more is non-bipartite.

As we now demonstrate, separable and complete graphs are closely related. First observe that, for any $q > 0$, the complete graph of size q is separable of order q. Conversely, any separable graph of order q can be related to the complete graph of size q in the following way. Let $G = (\mathcal{V}, \mathcal{E})$ be a separable graph of order q, and $\mathcal{I}_1, \dots, \mathcal{I}_q$ be its maximal independent sets. Observe that, as G is separable, the binary relation "$\not-$" is an equivalence relation in $\mathcal{V}$. If we "contract" G by "merging" all the nodes in each equivalence class (i.e., maximal independent set), so that each node in the contracted graph represents an independent set in G, and merge all the edges emanating from merged nodes that point to the same nodes, we obtain the complete graph $\tilde{G}$ of size q; see Figure 3.

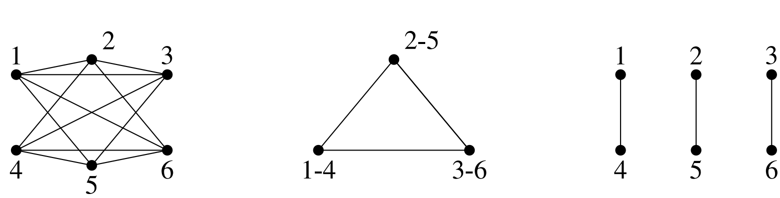

FIG. 3. *Separable graph of order* 3 (*left*); *its merged complete graph* (*middle*), *and its complement graph* (*right*).

3.2. *Preliminaries*. It follows from Theorem 1 that (5) is a necessary condition for stability of the system, regardless of the matching policy Φ. It is then natural to investigate what are the matching policies for which (5) is also a *sufficient* condition for stability of the matching queue.

DEFINITION 3. Let G be a connected graph. An admissible matching policy Φ on G is *maximal* if the stability region corresponding to G, Φ coincides with $\textsc{Ncond}_{\mathrm{C}}(G)$.

DEFINITION 4. A connected graph G is said to be:

- *matching-stable* if $\textsc{Ncond}_{\mathrm{C}}(G)$ is nonempty and all admissible matching policies on G are maximal;
- *matching-unstable* if the set $\textsc{Ncond}_{\mathrm{C}}(G)$ is empty.

In other words, if G is matching-stable the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ is stable for any admissible Φ and any $\lambda \in \textsc{Ncond}_{\mathrm{C}}(G)$. If G is matching-unstable, the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ is unstable for any admissible Φ and any arrival-rate vector λ. Clearly, a graph G might be neither matching-stable nor matching-unstable.

We have the following consequence to Theorem 2 in [23].

THEOREM 2. *For any connected graph G, the following hold*:

(i) *G is matching-unstable if and only if it is bipartite*;
(ii) *If G is non-bipartite, then the discipline* ML *is maximal*;
(iii) *If G is separable of order $q \geq 3$, then it is matching-stable*.

PROOF. The arguments in the proof of Theorem 1 are again employed to apply the results for the discrete-time model in [23] to the continuous-time model considered here. In particular, by Theorem 1 in [23], the set $\textsc{Ncond}_{\mathrm{C}}(G)$ is nonempty if and only if G is non-bipartite. Hence, the statements (i), (ii) and (iii) follow, respectively, from (16), (17) and (18) in [23]. □

We make the following observation: For a matching queue on a complete graph, all admissible matching policies are equivalent. Indeed, at most one class of items can be present in queue at any given time, so that an arriving item has no more than one choice for matching. Thus, the fact that all non-bipartite separable graphs are matching-stable is not surprising, given their relation to complete graphs, as described above.

Let G be a separable graph of order $q \geq 3$. Consider a matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ on G, where Φ is admissible, and the matching queue $(\tilde{G}, \tilde{\lambda}, \tilde{\Phi})_{\mathrm{C}}$, where $\tilde{G}$ is the "merged" complete graph of size q obtained from G, as described in Section 3.1, $\tilde{\Phi}$ is an arbitrary admissible policy and

$$\tilde{\lambda}_j = \bar{\lambda}_{\mathcal{I}_j}, \qquad j \in [\![1, q]\!].$$

(We add a tilde to all parameters associated with $\tilde{G}$.) Since $\tilde{G}$ is complete, the matching policy $\tilde{\Phi}$ is irrelevant, as long as it is admissible. In $(G, \lambda, \Phi)_{\mathrm{C}}$, only items of classes belonging to the same maximal independent set can be present in queue at any given time. Fix a time point t at which the system is nonempty, and let $\mathcal{I}_\ell$ be the (unique) maximal independent set having a nonempty queue at t. Suppose that an item enters the system at t. We have the following alternatives:

- If the new arriving item is of a class belonging to $\mathcal{I}_\ell$, then no matching occurs and the item joins the queue.
- If the new arrival is an item of a class $k \in \mathcal{I}_m$, $\ell \neq m$, then (no matter what the matching policy Φ is), the entering item will be matched with an item from a class in $\mathcal{I}_\ell$, and Φ only determines the class of its match in $\mathcal{I}_\ell$ (as several classes in $\mathcal{I}_\ell$ may have a nonempty queues at t).

Consequently, if the initial conditions satisfy $\sum_{k \in \mathcal{I}_j} Q_k(0) \stackrel{\mathrm{d}}{=} \tilde{Q}_j(0)$, $j \in [\![1, q]\!]$, then

$$\left\{\sum_{k \in \mathcal{I}_j} Q_k(t) : t \geq 0\right\} \stackrel{\mathrm{d}}{=} \{\tilde{Q}_j(t) : t \geq 0\}, \qquad j \in [\![1, q]\!]. \tag{7}$$

We conclude that, for *any* matching policy Φ, if one adopts a "macroscopic" view of the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$, by only keeping track of the maximal independent set present in queue at any time (there is at most one), and not of the particular classes of the items, then the matching queue on G amounts to a matching queue on $\tilde{G}$ having an arbitrary matching policy $\tilde{\Phi}$. As the latter model is stable at least for the policy ML since $\tilde{G}$ is non-bipartite [by assertion (ii) of Theorem 2 above], it is stable under any policy $\tilde{\Phi}$. From (7), the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ is stable regardless of the matching policy Φ, which is exactly assertion (iii) of Theorem 2.

3.3. *The main result.* Assertion (i) in Theorem 2 identifies the class of graphs rendering any matching queue unstable, regardless of the matching policy. By Assertion (ii) of the theorem, any matching queue $(G, \lambda, \mathrm{ML})_{\mathrm{C}}$ on a non-bipartite graph G is stable, provided λ satisfies NCOND. Assertion (iii) presents a class of graphs (the non-bipartite separable ones) that are matching-stable, namely, for any matching policy and arrival-rate vector satisfying NCOND the system is stable. Together, these results raise the question of whether the choice of the matching policy matters in terms of stability for graphs that are nonseparable and non-bipartite, that is, for graphs for which at least the discipline ML is maximal. The simplest such graph, namely, the pendant graph depicted in Figure 2, was considered in Section 5 of [23], where it is shown that this graph is not matching-stable. In particular, it was shown in [23] that, for a symmetric matching queue (with $\lambda_1 = \lambda_2$) there exists a matching policy for which the stability region is strictly included in $\textsc{Ncond}(G)$. Our main result, Theorem 3 below, provides a significant generalization of this result for a much larger class of graphs; en route, we also prove generalized versions of the results in [23], Section 5.

To present our main result, let $\mathscr{G}_7$ denote the set of all connected graphs inducing an odd cycle of size 7 or more, but no 5-cycle and no pendant graph, and let $\mathscr{G}_7^c$ denote its complement in the set of connected graphs.

THEOREM 3. *The only matching-stable graphs in* $\mathscr{G}_7^c$ *are separable of order* 3 *or more*.

In other words, except for the special case of graphs inducing an odd cycle of size 7 or more, but *no pendant graph and no 5-cycle*, the *only* matching-stable graphs are the non-bipartite separable graphs (i.e., separable graphs of order 3 or more). Therefore, separability of order at least 3 is not only *sufficient*, but also *necessary*, at least in $\mathscr{G}_7^c$, for the stability of any matching queue under $\textsc{Ncond}$. We conjecture that, among connected non-bipartite graphs, separability and matching-stability are equivalent or, in other words, that no graph in $\mathscr{G}_7$ is matching-stable. [The two statements are equivalent since all graphs in $\mathscr{G}_7$ are nonseparable; see Lemma 3(ii) below.] Even though we were not able to prove this result, we provide key steps in that direction; see Section 9 below.

Applying again the arguments of the proof of Theorem 1, we obtain the following immediate corollary to Theorem 3.

COROLLARY 1. *Theorem* 3 *also holds for the discrete-time matching model. In particular we have the following partial converse of assertion* (18) *in* [23]: *if* $G \in \mathscr{G}_7^c$ *is such that any discrete-time matching model* $(G, \mu, \Phi)_{\mathrm{C}}$ *is stable for* $\mu \in \textsc{Ncond}_{\mathrm{D}}(G)$, *then* G *is separable of order* $q \geq 3$.

3.4. *Strategy of the proof of Theorem* 3. To prove Theorem 3, we fix a non-bipartite and *nonseparable* graph G in $\mathscr{G}_7^c$, and show the existence of a nonmaximal *priority* matching policy Φ.

The proof hinges on the following fact, which will be proved in Section 6.1 [statement (i) in Lemma 3]: *any connected, non-bipartite and nonseparable graph induces a pendant graph or an odd cycle of length* 5 *or more.* Consequently, as G belongs to $\mathscr{G}_7^c$, it induces a graph $\breve{G}$ which is either a pendant graph or a 5-cycle. The remainder of the proof follows two main steps:

1. In Section 5, we construct a nonmaximal matching policy $\breve{\Phi}$ on the induced graph $\breve{G}$ (addressing successively the cases $\breve{G} =$ pendant graph and $\breve{G} =$ 5-cycle), by providing an arrival-rate vector $\breve{\lambda} \in \textsc{Ncond}_{\mathrm{C}}(\breve{G})$ such that $(\breve{G}, \breve{\lambda}, \breve{\Phi})_{\mathrm{C}}$ is unstable. In both cases, the instability of the system is shown using the fluid-limit arguments developed in Section 4.

2. We then prove that the instability of the matching queue on the induced graph $\breve{G}$ implies instability of the matching queue on the inducing graph G, by showing a "nonchaoticity" property in Section 6.2. In particular, we show that the influence of the arrivals to nodes of the complement of $\breve{G}$ in G can be bounded such that the unstable node in $(\breve{G}, \breve{\lambda}, \breve{\Phi})_{\mathrm{C}}$ remains unstable in $(G, \lambda, \Phi)_{\mathrm{C}}$, for a well-chosen arrival-rate vector $\lambda \in \mathrm{NCOND}_{\mathrm{C}}(G)$ and a well-chosen matching policy Φ.

4. Fluid stability. We now take a detour to develop the fluid limit which will be used in the proof of Theorem 3. Throughout this section, we fix a matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$, where Φ is of priority type, so that Q in (3) is a CTMC with state space

$$\mathbb{G} := \{z \in \mathbb{Z}_+^{|\mathcal{V}|} : z_i z_j = 0, \text{for any } i \in \mathcal{V} \text{ and } j \in \mathcal{E}(i)\}. \tag{8}$$

4.1. *Sample-path representation.* Before introducing the FWLLN for matching queues under priority policies, it is helpful to consider the sample-path representation of the CTMC Q. To that end, note that for each $i \in \mathcal{V}$, $Q_i(t)$ increases by 1 at time t if there is an arrival to node i at t and $Q_j(t) = 0$ for all $j \in \mathcal{E}(i)$; $Q_i(t)$ decreases by 1 at time t (when it is positive) if there is an arrival to one of the neighbors $j \in \mathcal{E}(i)$, and all the buffers in $\mathcal{E}(j)$ to which j gives a higher priority are empty. To express these dynamics, we introduce the following subsets of $\mathbb{G}$: for any $i \in \mathcal{V}$, we let

$$\begin{aligned} \mathcal{N}_i &:= \{z \in \mathbb{G}; z_i > 0\}; \\ \mathcal{O}_i &:= \{z \in \mathbb{G} : z_j = 0 \text{ for all } j \in \mathcal{E}(i)\}; \\ \mathcal{P}_j(i) &:= \{z \in \mathbb{G} : z_k = 0 \text{ for all } k \in \Phi_j(i)\}, \qquad j \in \mathcal{E}(i), \end{aligned} \tag{9}$$

where $\Phi_j(i)$ is the list of all the neighbors of node j to which node j gives a higher priority than to node i, namely,

$$\Phi_j(i) = \{k \in \mathcal{E}(j); j \text{ gives priority to } k \text{ over } i \text{ according to } \Phi\}. \tag{10}$$

Let $\mathscr{A}$ denote the infinitesimal generator of the queue process Q. Then, by the definition of the matching policy Φ, the only positive terms $\mathscr{A}(z, y)$, for all $y, z \in \mathbb{G}$, are given by

$$\begin{cases} \mathscr{A}(z, z + \mathbf{e}_i) = \lambda_i \mathbb{1}_{\mathcal{O}_i}(z), & i \in \mathcal{V}; \\ \mathscr{A}(z, z - \mathbf{e}_i) = \mathbb{1}_{\mathcal{N}_i}(z) \sum_{j \in \mathcal{E}(i)} (\lambda_j \mathbb{1}_{\mathcal{P}_j(i)}(z)), & i \in \mathcal{V}, \end{cases} \tag{11}$$

where $\mathbb{1}_A(\cdot)$ is the indicator function of the set A. Consequently (see, e.g., [25]), for all $i \in \mathcal{V}$, we can represent the sample path of Q_i using the independent Poisson arrival processes $N_i, i \in \mathcal{V}$, via

$$
\begin{aligned}
Q_i(t) &= Q_i(0) + \int_0^t \mathbb{1}_{\mathcal{O}_i}\big(Q(s-)\big)\, dN_i(s) \\
&\quad - \sum_{j \in \mathcal{E}(i)} \int_0^t \mathbb{1}_{\mathcal{N}_i}\big(Q(s-)\big)\mathbb{1}_{\mathcal{P}_j(i)}\big(Q(s-)\big)\, dN_j(s), \qquad t \geq 0,
\end{aligned} \tag{12}
$$

where $Q(t-)$ denotes the left limit of Q at the time point t.

EXAMPLE 1. For the pendant graph in which node 3 prioritizes nodes 1 and 2 over 4, as depicted in Figure 2, the subsets in (9) become

$$\mathcal{O}_4 = \{z \in \mathbb{G} : z_3 = 0\} \quad \text{and} \quad \mathcal{P}_3(4) = \{z \in \mathbb{G} : z_1 = z_2 = 0\}.$$

The sample paths of Q_4 in that case can be represented via

$$
\begin{aligned}
Q_4(t) &= Q_4(0) + \int_0^t \mathbb{1}_{\mathcal{O}_4}\big(Q(s-)\big)\, dN_4(s) \\
&\quad - \int_0^t \mathbb{1}_{\mathcal{N}_4 \cap \mathcal{P}_3(4)}\big(Q(s-)\big)\, dN_3(s), \qquad t \geq 0.
\end{aligned}
$$

4.2. *Marginal process corresponding to a particular node*. The fluid limit we are about to introduce is formulated for a particular class of models, exhibiting the following situation. When fixing the arrival-rate vector λ and the admissible matching policy Φ, and when isolating a single node for which we take the corresponding initial buffer content to be strictly positive, the content process of all the nodes different from that node and its neighbors, coincides in law with an ergodic CTMC.

Formally, fix a matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ and a node i_0 of G. Let

$$\mathcal{S} = \mathcal{V} \setminus \big(\{i_0\} \cup \mathcal{E}(i_0)\big),$$

and index the elements of $\mathcal{S}$ as follows:

$$\mathcal{S} = \{i_1, \ldots, i_{|\mathcal{S}|}\}. \tag{13}$$

Now, for any $z \in \mathbb{G}$ such that $z_{i_0} > 0$, the only positive terms $\mathscr{A}(z, y)$, $y \in \mathbb{G}$, are given by

$$
\begin{cases}
\mathscr{A}(z, z + \mathbf{e}_{i_0}) = \lambda_{i_0}; \\
\mathscr{A}(z, z - \mathbf{e}_{i_0}) = \sum_{j \in \mathcal{E}(i_0)} \big(\lambda_j \mathbb{1}_{\mathcal{P}_j(i_0)}(z)\big); \\
\mathscr{A}(z, z + \mathbf{e}_{i_\ell}) = \lambda_{i_\ell} \mathbb{1}_{\mathcal{O}_{i_\ell}}(z), & \ell \in [\![1, |\mathcal{S}|]\!]; \\
\mathscr{A}(z, z - \mathbf{e}_{i_\ell}) = \mathbb{1}_{\mathcal{N}_{i_\ell}}(z) \sum_{\substack{j \in \mathcal{E}(i_\ell): \\ i_0 \notin \Phi_j(i_\ell)}} \big(\lambda_j \mathbb{1}_{\mathcal{P}_j(i_\ell)}(z)\big), & \ell \in [\![1, |\mathcal{S}|]\!].
\end{cases} \tag{14}
$$

To see this, observe that, by the definition of Φ, $\mathbb{1}_{\mathcal{P}_j(i_\ell)}(z) = 0$ for all ℓ and $j \in \mathcal{E}(i_\ell)$ such that $i_0 \in \Phi_j(i_\ell)$, since $z_{i_0} > 0$.

Let $S = \{S(t) : t \geq 0\}$ denote the restriction of the process Q to the nodes of $\mathcal{S}$, that is,

$$S = (S_1, S_2, \dots, S_{|\mathcal{S}|}) := (Q_{i_1}, \dots, Q_{i_{|\mathcal{S}|}}). \tag{15}$$

Then S achieves values in the following subset $\mathbb{G}^{\mathcal{S}}$ of $\mathbb{Z}_+^{|\mathcal{S}|}$:

$$\mathbb{G}^{\mathcal{S}} = \{x \in \mathbb{Z}_+^{|\mathcal{S}|} : x_k x_\ell = 0 \text{ for } k, \ell \in [\![1, |\mathcal{S}|]\!] \text{ such that } i_k \in \mathcal{E}(i_\ell)\}. \tag{16}$$

Analogously to (9), we define the following subsets of $\mathbb{G}^{\mathcal{S}}$: For any $\ell \in [\![0, |\mathcal{S}|]\!]$:

$$\mathcal{N}_{i_\ell}^{\mathcal{S}} := \{x \in \mathbb{G}^{\mathcal{S}} : x_\ell > 0\};$$

$$\mathcal{O}_{i_\ell}^{\mathcal{S}} := \{x \in \mathbb{G}^{\mathcal{S}} : x_j = 0 \text{ for all } j \in [\![1, |\mathcal{S}|]\!] \text{ such that } i_j \in \mathcal{E}(i_\ell)\};$$

and for any $\ell \in [\![0, |\mathcal{S}|]\!]$ and $j \in \mathcal{E}(i_\ell)$,

$$\mathcal{P}_j^{\mathcal{S}}(i_\ell) := \{x \in \mathbb{G}^{\mathcal{S}} : x_k = 0 \text{ for all } k \text{ such that } i_k \in \Phi_j(i_\ell)\}. \tag{17}$$

DEFINITION 5. The *marginal process* corresponding to node i_0 is the $\mathbb{G}^{\mathcal{S}}$-valued CTMC $\chi := \{\chi(t) : t \geq 0\}$, whose infinitesimal generator $\mathscr{A}^{\mathcal{S}}$ has the following positive terms:

$$\begin{cases} \mathscr{A}^{\mathcal{S}}(x, x + \mathbf{e}_\ell) = \lambda_{i_\ell} \mathbb{1}_{\mathcal{O}_{i_\ell}^{\mathcal{S}}}(x) & \ell \in [\![1, |\mathcal{S}|]\!]; \\ \mathscr{A}^{\mathcal{S}}(x, x - \mathbf{e}_\ell) = \mathbb{1}_{\mathcal{N}_{i_\ell}^{\mathcal{S}}}(x) \displaystyle\sum_{\substack{j \in \mathcal{E}(i_\ell): \\ i_0 \notin \Phi_j(i_\ell)}} (\lambda_j \mathbb{1}_{\mathcal{P}_j^{\mathcal{S}}(i_\ell)}(x)) & \ell \in [\![1, |\mathcal{S}|]\!]. \end{cases} \tag{18}$$

Observe that for any $z \in \mathbb{G}$, if $x \in \mathbb{G}^{\mathcal{S}}$ is defined by $x = (z_{i_1}, \dots, z_{i_{|\mathcal{S}|}})$, then for all $\ell \in [\![1, |\mathcal{S}|]\!]$ we have

$$\mathbb{1}_{\mathcal{O}_{i_\ell}}(z) = \mathbb{1}_{\mathcal{O}_{i_\ell}^{\mathcal{S}}}(x);$$

$$\mathbb{1}_{\mathcal{N}_{i_\ell}}(z) = \mathbb{1}_{\mathcal{N}_{i_\ell}^{\mathcal{S}}}(x),$$

and if $z_{i_0} > 0$, as we have $z_j = 0$ for all $j \in \mathcal{E}(i_0)$, by definition of Φ we obtain

$$\mathbb{1}_{\mathcal{P}_j(i_\ell)}(z) = \mathbb{1}_{\mathcal{P}_j^{\mathcal{S}}(i_\ell)}(x) \qquad \text{for all } j \in \mathcal{E}(i_\ell).$$

Therefore, in view of (14) and (18) we can provide a more intuitive definition of the marginal process associated to the node i_0: it is a Markov process on $\mathbb{G}^{\mathcal{S}}$ which coincides in distribution with the restriction S of the process Q to its coordinates in $\mathcal{S}$, conditionally on the i_0th coordinate of Q being positive.

EXAMPLE 2 (Example 1, continued). Set $i_0 = 4$. Then we have $\mathcal{S} = \{1, 2\}$. Set $i_1 = 1$ and $i_2 = 2$. We thus have

$$\mathbb{G}^{\mathcal{S}} = (\{0\} \times \mathbb{Z}_+) \cup (\mathbb{Z}_+ \times \{0\}) =: \mathbb{E}_2, \tag{19}$$

and the following subsets of $\mathbb{E}_2$:

$$\mathcal{O}_1^{\mathcal{S}} = \mathbb{Z}_+ \times \{0\}; \qquad \mathcal{P}_3^{\mathcal{S}}(1) = \mathcal{P}_2^{\mathcal{S}}(1) = \mathcal{N}_1^{\mathcal{S}} = \mathbb{Z}_{++} \times \{0\};$$
$$\mathcal{O}_2^{\mathcal{S}} = \{0\} \times \mathbb{Z}_+; \qquad \mathcal{P}_3^{\mathcal{S}}(2) = \mathcal{P}_1^{\mathcal{S}}(2) = \mathcal{N}_2^{\mathcal{S}} = \{0\} \times \mathbb{Z}_{++}.$$

Thus, the positive terms of the generator $\mathscr{A}^{\mathcal{S}}$ are

$$\begin{cases} \mathscr{A}^{\mathcal{S}}(x, x + \mathbf{e}_1) = \lambda_1 & \text{for } x \in \mathbb{Z}_+ \times \{0\}; \\ \mathscr{A}^{\mathcal{S}}(x, x - \mathbf{e}_1) = \lambda_3 + \lambda_2 & \text{for } x \in \mathbb{Z}_{++} \times \{0\}; \\ \mathscr{A}^{\mathcal{S}}(x, x + \mathbf{e}_2) = \lambda_2 & \text{for } x \in \{0\} \times \mathbb{Z}_+; \\ \mathscr{A}^{\mathcal{S}}(x, x - \mathbf{e}_2) = \lambda_3 + \lambda_1 & \text{for } x \in \{0\} \times \mathbb{Z}_{++}. \end{cases} \tag{20}$$

4.3. *The FWLLN.* Throughout this section, fix the matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ and the node $i_0 \in \mathcal{V}$. We consider the sequence of fluid-scaled processes $\{\bar{Q}^n : n \geq 1\}$, defined via

$$\bar{Q}^n(t) = \frac{Q^n(t)}{n} := \frac{Q(nt)}{n}, \qquad t \geq 0, n \geq 1. \tag{21}$$

Similarly, recalling (15) we define

$$\bar{S}^n(t) = \frac{S^n(t)}{n} := \frac{S(nt)}{n}, \qquad t \geq 0, n \geq 1. \tag{22}$$

For $n \in \mathbb{Z}_{++}$, we will use the notation $N_i^n(\cdot)$ for the time-scaled Poisson arrival process to node i, namely, $N_i^n(\cdot) = N_i(n\cdot)$, $i \in \mathcal{V}$. We also denote by χ^n the nth marginal process corresponding to i_0, defined by

$$\chi^n(t) = \chi(nt), \qquad t \geq 0, \tag{23}$$

and define

$$\bar{\chi}^n(t) = \frac{\chi^n(t)}{n}, \qquad t \geq 0, n \geq 1. \tag{24}$$

The insufficiency of NCOND to ensure the stability of a given matching queue will be shown via the following lemma.

LEMMA 1. *If there exists an initial condition $Q^n(0) \in \mathbb{G}$ such that $\bar{Q}^n \Rightarrow \bar{Q}$ in $\mathbb{D}^{|\mathcal{V}|}$ as $n \to \infty$ and $\mathcal{P} := \{i \in \mathcal{V} : \bar{Q}_i(0) > 0\} \neq \varnothing$, and if for some $i \in \mathcal{P}$ it holds that $\bar{Q}_i$ is nondecreasing, then Q is either transient or null recurrent. In particular, the corresponding matching queue $(G, \lambda, \Phi)_{\mathrm{C}}$ is unstable.*

PROOF. By Proposition 9.9 in [28], if Q is positive recurrent there exists a (possibly random) time $\mathcal{T}$, such that $\mathcal{T} < \infty$ w.p.1 and $\bar{Q}(t) = 0$ for all $t \geq \mathcal{T}$. $\square$

For the fluid analysis, we make two assumptions.

ASSUMPTION 1. $Q^n(0) \in \mathbb{G}$ for any $n \geq 1$, and $\bar{Q}^n(0) \Rightarrow \bar{Q}(0)$ as $n \to \infty$, where $\bar{Q}(0)$ is a deterministic element of $\mathbb{R}^{|\mathcal{V}|}$, with $\bar{Q}_{i_0}(0) > 0$ and $\bar{Q}_i(0) = 0$, $i \in \mathcal{V} \setminus \{i_0\}$.

ASSUMPTION 2. For all $n \geq 1$, the $\mathbb{G}^{\mathcal{S}}$-valued process χ^n is ergodic with stationary probability π^n.

For $n \geq 1$, let

$$\rho^n := \rho^n\big(Q^n(0)\big) := \inf\big\{t \geq 0 : Q^n_{i_0}(t) = 0\big\}, \tag{25}$$

with $\inf \varnothing := \infty$.

LEMMA 2. *Consider the sequence* $\{\bar{Q}^n : n \in \mathbb{Z}_{++}\}$ *corresponding to a system* $(G, \lambda, \Phi)_{\mathrm{C}}$. *Then there exist* $n_0 \in \mathbb{Z}_+$ *and* $\delta > 0$ *such that* $\rho^n(Q^n(0)) > \delta$ *w.p.*1 *for all* $n \geq n_0$. *In particular, there exists* $n_0 < \infty$, *such that*

$$\inf_{0 \leq t < \delta} \bar{Q}^n_{i_0}(t) > 0 \qquad \text{w.p.1 for all } n \geq n_0. \tag{26}$$

PROOF. We use a simple coupling argument. Consider the matching queue $(G, \lambda, \widetilde{\Phi})_{\mathrm{C}}$ (with the same graph G and arrival-rate vector λ as in the statement of the lemma), where $\widetilde{\Phi}$ is the priority policy under which each $i \in \mathcal{E}(i_0)$ gives the highest priority to node i_0. If the corresponding queue process $\widetilde{Q}^n$ is given the same Poisson processes $\{N^n_i : i \in \mathcal{V}\}$ of the original system, we clearly have

$$Q^n_{i_0}(t) \geq \widetilde{Q}^n_{i_0}(t) := Q^n_{i_0}(0) + N_{i_0}(nt) - \sum_{j \in \mathcal{E}(i_0)} N_j(nt), \qquad 0 \leq t \leq \delta^n,$$

where $\delta^n := \inf\{t > 0 : \widetilde{Q}^n_{i_0}(t) = 0\}$.

Dividing $\widetilde{Q}^n_{i_0}$ by n and taking $n \to \infty$, we obtain from the functional strong law of large numbers (FSLLN) for the Poisson process, that $\widetilde{Q}^n/n$ converges w.p.1 to $\widetilde{q}$, where

$$\widetilde{q}_{i_0}(t) := \bar{Q}_{i_0}(0) + \Bigg(\lambda_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \lambda_j\Bigg)t, \qquad 0 \leq t < \delta, \tag{27}$$

where

$$\delta := \frac{\bar{Q}_{i_0}(0)}{\sum_{j \in \mathcal{E}(i_0)} \lambda_j - \lambda_{i_0}} \qquad \text{if } \sum_{j \in \mathcal{E}(i_0)} \lambda_j - \lambda_{i_0} > 0 \quad \text{and} \quad \delta := \infty \text{ otherwise.}$$

The uniform convergence over compact subintervals of $[0, \delta)$ of the lower bounding processes $\widetilde{Q}^n_{i_0}/n$ to a strictly positive function gives (26). □

Before presenting the fluid limit, we explain the intuition behind the expression for $\bar{Q}_{i_0}$ that we obtain. Since $\bar{Q}_{i_0}$ is strictly positive over an interval $[0, \delta)$ by Lemma 2, if $S^n(0) \stackrel{\mathrm{d}}{=} \chi^n(0)$ for all $n \geq 1$, then from (14), (18) describes the infinitesimal rates of S^n over $[0, \delta)$, so that

$$\{S^n(t) : 0 \leq t < \delta\} \stackrel{\mathrm{d}}{=} \{\chi^n(t) : 0 \leq t < \delta\}. \tag{28}$$

Hence, S^n is locally (over $[0, \delta)$) distributed as a CTMC, which is ergodic by Assumption 2. Thus, it is not hard to show that $\bar{S}^n$ converges to 0 over that interval; see the proof of Theorem 4. Nevertheless, the dynamics of $\bar{S}^n$ determine those of $\bar{Q}^n_{i_0}$ for each n, as is clear from (12), and the affect of $\bar{S}^n$ on $\bar{Q}^n_{i_0}$ does not diminish as n increases to infinity. However, the "small" process S^n is also "fast" relative to $\bar{Q}^n_{i_0}$, since (23) implies that, regardless of the distribution of $\chi^n(0)$, $\chi^n(t) \stackrel{\mathrm{d}}{\approx} \chi(\infty)$, for any $t > 0$ and for all large-enough n, where $\chi(\infty)$ denotes a random variable having the stationary distribution of χ. [We write $\stackrel{\mathrm{d}}{\approx}$ if, in the limit as $n \to \infty$, the distribution of $\chi^n(t)$ at time t converges to the stationary distribution of χ, that is, $\chi^n(t) \Rightarrow \chi(\infty)$ in $\mathbb{G}^{\mathcal{S}}$.] Then (28) implies that $S^n(t)$ converges to $\chi(\infty)$ as well as $n \to \infty$, $0 < t < \delta$. Such a result is known in the queueing literature as a *pointwise stationarity*, for example, [33]. Of course, to obtain a FWLLN, the convergence must hold uniformly in t over the interval $[0, \delta)$, namely, the aforementioned stochastic AP must hold, but the intuition for the fast averaging phenomenon is similar.

Formally, let π denote the stationary distribution of the CTMC χ whose generator $\mathscr{A}^{\mathcal{S}}$ is given in (18), that is,

$$\pi(\mathcal{Z}) = P(\chi(\infty) \in \mathcal{Z}), \qquad \mathcal{Z} \subseteq \mathbb{G}^{\mathcal{S}}. \tag{29}$$

THEOREM 4 (FWLLN). *Let $(G, \lambda, \Phi)_{\mathrm{C}}$ be a matching queue such that Φ is of the priority type. If, for some node i_0*

$$\lambda_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \lambda_j \pi(\mathcal{P}^{\mathcal{S}}_j(i_0)) < 0, \tag{30}$$

for π in (29) and $\mathcal{P}^{\mathcal{S}}_j(i_0)$, $j \in \mathcal{E}(i_0)$ in (17), then $\rho^n \Rightarrow \rho$ in $\mathbb{R}$ as $n \to \infty$, for ρ^n in (25), where

$$\rho := \frac{\bar{Q}_{i_0}(0)}{\sum_{j \in \mathcal{E}(i_0)} \lambda_j \pi(\mathcal{P}^{\mathcal{S}}_j(i_0)) - \lambda_{i_0}}. \tag{31}$$

Otherwise, $\rho^n \Rightarrow \infty$. *In either case*, $\bar{Q}^n \Rightarrow \bar{Q}$ *in* $\mathbb{D}^{|\mathcal{V}|}[0,\rho)$ *as* $n \to \infty$, *where*

$$\begin{aligned} \bar{Q}_{i_0}(t) &= \bar{Q}_{i_0}(0) + \Bigl(\lambda_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \lambda_j \pi\bigl(\mathcal{P}_j^{\mathcal{S}}(i_0)\bigr)\Bigr)t, \\ \bar{Q}_i(t) &= 0, \qquad i \in \mathcal{V} \setminus \{i_0\}. \end{aligned} \tag{32}$$

The proof of Theorem 4 is given in Section 7. From Theorem 4 and Lemma 1, it immediately follows that we have the following.

COROLLARY 2. *If* $\rho^n \Rightarrow \infty$, *for* ρ^n *in* (25), *then* $(G, \lambda, \Phi)_{\mathrm{C}}$ *is unstable.*

It is significant that we can compute the stationary probabilities $\pi(\cdot)$ in (29) in some cases, using reversibility arguments.

5. The pendant graph and the 5-cycle.

In this section, we analyze matching queues defined on the pendant graph and the 5-cycle, using the fluid limit in Theorem 4. In both cases, the stationary probability π [on $\mathbb{E}_2$ in (19)] can be computed explicitly, so that the stability region of the corresponding matching queues can be fully characterized.

5.1. *The pendant graph.*

We start with the model depicted in Figure 2.

PROPOSITION 1. *Let* G *be the pendant graph and* Φ *the matching policy depicted in Figure* 2. *Consider an arrival-rate vector* $\lambda \in \mathrm{NCOND}_{\mathrm{C}}(G)$, *that is*,

$$\lambda_4 < \lambda_3 < \lambda_4 + \lambda_1 + \lambda_2, \qquad \lambda_4 + \lambda_1 < \lambda_3 + \lambda_2 \quad \text{and} \quad \lambda_4 + \lambda_2 < \lambda_3 + \lambda_1. \tag{33}$$

If $\bar{Q}^n(0) \Rightarrow x\mathbf{e}_4$ *in* $\mathbb{R}^4$ *for some* $x \in \mathbb{R}_{++}$, *then* $\bar{Q}^n \Rightarrow \bar{Q}$ *in* $\mathbb{D}^4[0,\rho)$ *as* $n \to \infty$, *where*

$$\bar{Q}(t) = \bigl(0, 0, 0, x + (\lambda_4 - \lambda_3\alpha)t\bigr), \qquad 0 \le t < \rho, \tag{34}$$

for $\rho := x/(\lambda_3\alpha - \lambda_4)$ *if* $\alpha > \lambda_4/\lambda_3$ *and* $\rho := \infty$ *otherwise*, *and for*

$$\alpha := \left[1 + \frac{\lambda_1}{\lambda_3 + \lambda_2 - \lambda_1} + \frac{\lambda_2}{\lambda_3 + \lambda_1 - \lambda_2}\right]^{-1} = \frac{(\lambda_3)^2 - (\lambda_1 - \lambda_2)^2}{\lambda_3(\lambda_3 + \lambda_1 + \lambda_2)}. \tag{35}$$

PROOF. The result follows from Theorem 4. In the present case, we set $i_0 = 4$, so that the marginal process χ is a $\mathbb{E}_2$-valued CTMC having the generator $\mathscr{A}^{\mathcal{S}}$ in (20) (see Example 2). For α in (35), let $\pi(0,0) = \alpha$ and

$$\pi(x) = \begin{cases} \alpha\left(\dfrac{\lambda_1}{\lambda_3 + \lambda_2}\right)^i & \text{for } x = (i,0), i \ge 1; \\ \alpha\left(\dfrac{\lambda_2}{\lambda_3 + \lambda_1}\right)^j & \text{for } x = (0,j), i \ge 1. \end{cases}$$

It is easy to check that under (33), π is a probability vector satisfying the detailed balance equations for χ, so that it is the unique stationary distribution of this reversible CTMC. In particular, Assumption 2 holds. Since items of class 3 give priority to 1 and 2 over 4, we have that $\mathcal{P}_3^{\mathcal{S}}(4) = (0, 0)$, so the stated convergence of Q^n to the fluid limit in (34) follows from (32). $\square$

PROPOSITION 2. *The matching queue* $(G, \lambda, \Phi)_{\mathrm{C}}$ *corresponding to the pendant graph* G *and the priority policy* Φ *represented in Figure 2 is stable if and only if* NCOND *holds together with*

$$\lambda_4 < \alpha\lambda_3, \tag{36}$$

for α *in* (35).

PROOF. The necessity of NCOND has been shown in Theorem 1. Also, it follows from Proposition 1 that, for any initial condition of the form $(0, 0, 0, x)$, $x > 0$, the fluid limit $\bar{Q}$ will hit the origin if and only if $\lambda_4 < \alpha\lambda_3$. The necessity of (36) then follows from Lemma 1. To show sufficiency, we apply Dai's result in [10]. To that end, we must consider *all possible initial conditions* for the fluid limit, and show that the origin is guaranteed to be hit in finite time.

First, assume that $\bar{Q}_3(0) > 0$, so that all other queues are empty initially. In that case, and as long as the class-3 queue is strictly positive, its drift down (toward 0) is $\lambda_4 + \lambda_1 + \lambda_2$, which is larger than the drift up λ_3 by (33). In particular, during the initial interval over which $\bar{Q}_3 > 0$, the class-3 queue is distributed as an ergodic birth and death (BD) process whose fluid limit is known to be (e.g., Proposition 5.16 in [28])

$$\bar{Q}_3(t) = \bar{Q}_3(0) + (\lambda_3 - \lambda_4 - \lambda_1 - \lambda_2)t, \qquad 0 \le t \le \frac{\bar{Q}_3(0)}{\lambda_4 + \lambda_1 + \lambda_2 - \lambda_3},$$

so that the fluid queue hits the origin in finite time.

Now assume that $\bar{Q}_4(0) > 0$. Then at most one of $\bar{Q}_1(0)$ or $\bar{Q}_2(0)$ can be strictly positive. Say $\bar{Q}_1(0) > 0$. In that case, the matching policy we consider implies that, as long as $\bar{Q}_1 > 0$, all the arriving items of classes 3 and 2 are matched with class-1 items. Hence, as long as the class-1 queue process is positive, it is distributed as a BD process having a constant birth rate λ_1 and a constant death rate $\lambda_3 + \lambda_2$. This BD process is ergodic due to (33), and its fluid limit is

$$\bar{Q}_1(t) = \bar{Q}_1(0) + (\lambda_1 - \lambda_3 - \lambda_2)t, \qquad 0 \le t \le \frac{\bar{Q}_1(0)}{\lambda_3 + \lambda_2 - \lambda_1}.$$

In particular, the fluid process $\bar{Q}_1$ will hit 0 in finite time, so that $\bar{Q}$ will hit the origin in finite time by Proposition 1. A similar argument applies when $\bar{Q}_2(0) > 0$.

Now, since the prelimit processes Q_i, $i = 1, 2, 3$ have drifts toward 0 whenever any of them is strictly positive, the fluid limit must remain in state 0 after hitting this state, and Proposition 1 shows that $\bar{Q}_4$ will also remain fixed at 0 after hitting that state. Thus, the ergodicity of the system follows from Theorem 4.2 in [10]. □

REMARK 1. A discrete version of Proposition 2 was proved in [23] for the symmetric model $(G, \mu, \Phi)_{\mathrm{D}}$ in which $\mu_\lambda(1) = \mu_\lambda(2)$ (so that a lower-dimensional process can be considered), via subtle Lyapunov-stability arguments. Plugging $\lambda_1 = \lambda_2$ in (35) and recalling (6), Proposition 2 gives that the discrete model $(G, \mu_\lambda, \Phi)_{\mathrm{D}}$ corresponding to Figure 2 is stable if and only if

$$\mu_\lambda(4) < \mu_\lambda(3) < \mu_\lambda(4) + 2\mu_\lambda(1) \quad \text{and} \quad \big(\mu_\lambda(3)\big)^2 > \mu_\lambda(4)\big(1 - \mu_\lambda(4)\big). \tag{37}$$

As $\mu_\lambda(2) = \mu_\lambda(1)$, the left-hand condition above is equivalent to

$$\begin{cases} \mu_\lambda(3) < \mu_\lambda(4) + \mu_\lambda(1) + \mu_\lambda(2) = \mu_\lambda\big(\mathcal{E}(3)\big); \\ \mu_\lambda(4) < \mu_\lambda(3) = \mu_\lambda\big(\mathcal{E}(4)\big); \\ \mu_\lambda(1) < \mu_\lambda(3) + \mu_\lambda(1) = \mu_\lambda(3) + \mu_\lambda(2) = \mu_\lambda\big(\mathcal{E}(1)\big); \\ \mu_\lambda(2) < \mu_\lambda(3) + \mu_\lambda(2) = \mu_\lambda(3) + \mu_\lambda(1) = \mu\big(\mathcal{E}(2)\big); \\ \mu_\lambda\big(\{1,4\}\big) = \mu_\lambda(1) + \mu_\lambda(4) < \mu_\lambda(2) + \mu_\lambda(3) = \mu_\lambda\big(\mathcal{E}(\{1,4\})\big); \\ \mu_\lambda\big(\{2,4\}\big) = \mu_\lambda(2) + \mu_\lambda(4) < \mu_\lambda(1) + \mu_\lambda(3) = \mu_\lambda\big(\mathcal{E}(\{2,4\})\big). \end{cases}$$

Thus, the measure μ_λ satisfies the condition $\textsc{Ncond}(G)$ in page 5 of [23]. It is easy to see that the second condition in (37) is equivalent to the right-hand condition defining the region denoted $\textsc{Stab}(A)$ in Proposition 3 of [23] (after re-indexing the nodes according to Figure 1 in [23]). In particular, a measure μ_λ with $\mu_\lambda(1) = \mu_\lambda(2)$ satisfies (37) if and only if it belongs to $\textsc{Stab}(A)$, so we retrieved the stability condition that was established in Proposition 3 of [23] for that particular case. [Note, however, that we do not require $\mu_\lambda(1) = \mu_\lambda(2)$.]

As the following shows, the stability region of the model in Figure 2, namely

$$\textsc{Ncond}_{\mathrm{C}}(G) \cap \big\{\lambda \text{ satisfying (36)}\big\},$$

is strictly contained in $\textsc{Ncond}_{\mathrm{C}}(G)$.

PROPOSITION 3. *We have the strict inclusion*

$$\big\{\lambda \textit{ satisfying } (36)\big\} \cap \textsc{Ncond}_{\mathrm{C}}(G) \subsetneq \textsc{Ncond}_{\mathrm{C}}(G).$$

PROOF. Fix $\epsilon \in (0, 2/5]$ and set

$$\begin{cases} \lambda_1 = \lambda_2 = \epsilon/2; \\ \lambda_3 = \dfrac{1}{2} - \epsilon/4; \\ \lambda_4 = \dfrac{1}{2} - 3\epsilon/4. \end{cases}$$

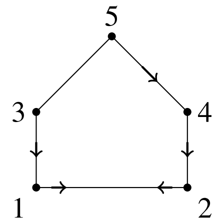

FIG. 4. *The* 5*-cycle*; *arrows indicate priorities*.

It is then easily checked that $\lambda \in \text{NCOND}_C(G)$. However, simple algebra shows that

$$\lambda_4 - \alpha\lambda_3 = \frac{1}{2} - \frac{3\epsilon}{4} - \left(\frac{1}{2} - \frac{\epsilon}{4}\right)\frac{1/2 - \epsilon/4}{1/2 + 3\epsilon/4} = \frac{\epsilon/4(1 - 5\epsilon/2)}{1/2 + 3\epsilon/4} \geq 0,$$

so that λ does not satisfy (36). □

5.2. *The* 5*-cycle*. We now consider the matching queue corresponding to the 5-cycle, under the priority policy depicted in Figure 4: nodes 1 and 2 prioritize each other, node 3 gives priority to node 1, and node 4 gives priority to node 2. For concreteness, we assume that node 5 gives priority to node 4.

As for the pendant graph analyzed above, the stability region of the 5-dimensional CTMC Q is challenging, even in symmetric cases; However, stability analysis is made considerably more simple *via* the fluid limits analysis. In particular, we now obtain a necessary and sufficient condition for stability of the matching queue corresponding to the 5-cycle and the matching policy Φ specified above, so that the stability region of the model is fully characterized.

Let

$$a := \frac{\lambda_3(\lambda_1 + \lambda_4)}{\lambda_1 + \lambda_4 - \lambda_2} + \frac{\lambda_4(\lambda_2 + \lambda_3)}{\lambda_2 + \lambda_3 - \lambda_1}. \tag{38}$$

PROPOSITION 4. *Let G be the* 5*-cycle*, Φ *be the priority matching policy depicted in Figure* 4, *and consider* $\lambda \in \text{NCOND}_C(G)$. *Assume that for some* $x \in \mathbb{R}_{++}$, $\bar{Q}^n(0) \Rightarrow x\mathbf{e}_5$ *in* $\mathbb{R}^5$. *Then* $\bar{Q}^n \Rightarrow \bar{Q}$ *in* $\mathbb{D}^5[0, \tilde{\rho})$ *as* $n \to \infty$, *where*, *for* a *in* (38),

$$\bar{Q}(t) = \big(0, 0, 0, 0, x + (\lambda_5 - a\tilde{\alpha})t\big), \qquad 0 \leq t < \tilde{\rho};$$

for

$$\tilde{\alpha} = \left[1 + \frac{\lambda_1}{\lambda_2 + \lambda_3 - \lambda_1} + \frac{\lambda_2}{\lambda_1 + \lambda_4 - \lambda_2}\right]^{-1} \tag{39}$$

and

$$\tilde{\rho} = \frac{x}{a\tilde{\alpha} - \lambda_5} \qquad \text{if } \lambda_5 < a\tilde{\alpha} \quad \text{and} \quad \tilde{\rho} := \infty \qquad \text{otherwise}.$$

It will be seen in the proof of Proposition 4 that $\tilde{\alpha}$ is the normalizing constant that makes the solution to the detailed balance equations a probability vector. Note that $\text{NCOND}_{\text{C}}(G)$ implies that a in (38) and $\tilde{\alpha}$ in (39) are well defined. To see this, consider the independent sets $\{1\}$ and $\{2\}$, whose neighboring sets are $\mathcal{E}(1) = \{2, 3\}$ and $\mathcal{E}(2) = \{1, 4\}$, respectively.

PROOF. Set $i_0 = 5$. In this case, from (18) the generator of the associated $\mathbb{E}_2$-valued marginal process χ has the following positive terms:

$$\begin{cases} \tilde{\mathscr{A}}^{\mathcal{S}}(x, x + \mathbf{e}_1) = \lambda_1 & \text{for } x \in \mathbb{Z}_+ \times \{0\}; \\ \tilde{\mathscr{A}}^{\mathcal{S}}(x, x - \mathbf{e}_1) = \lambda_3 + \lambda_2 & \text{for } x \in \mathbb{Z}_{++} \times \{0\}; \\ \tilde{\mathscr{A}}^{\mathcal{S}}(x, x + \mathbf{e}_2) = \lambda_2 & \text{for } x \in \{0\} \times \mathbb{Z}_+; \\ \tilde{\mathscr{A}}^{\mathcal{S}}(x, x - \mathbf{e}_2) = \lambda_1 + \lambda_4 & \text{for } x \in \{0\} \times \mathbb{Z}_{++}. \end{cases} \tag{40}$$

As in the proof of Proposition 1, one can easily check that under $\text{NCOND}_{\text{C}}(G)$,

$$\tilde{\pi}(x) = \begin{cases} \tilde{\alpha}\left(\dfrac{\lambda_1}{\lambda_3 + \lambda_2}\right)^i & \text{for } x = (i, 0), i \in \mathbb{Z}_+; \\ \tilde{\alpha}\left(\dfrac{\lambda_2}{\lambda_1 + \lambda_4}\right)^j & \text{for } x = (0, j), i \in \mathbb{Z}_+ \end{cases} \tag{41}$$

is the unique stationary distribution of χ. Since

$$\mathcal{P}_3^{\mathcal{S}}(5) = \{0\} \times \mathbb{Z}_+ \quad \text{and} \quad \mathcal{P}_4^{\mathcal{S}}(5) = \mathbb{Z}_+ \times \{0\},$$

the statement follows from Theorem 4. $\square$

COROLLARY 3. *A necessary condition for the matching queue in Figure 4 to be stable is*

$$\lambda_5 < a\tilde{\alpha} \qquad \textit{for } a \textit{ in } (38). \tag{42}$$

We next present a sufficient condition for the stability of the matching queue corresponding to Figure 4. Unlike the pendant graph, that sufficient condition is strictly stronger than the necessary condition of Corollary 3. First, note that nodes 1 and 2 are always stable, for any λ in $\text{NCOND}_{\text{C}}(G)$. This can be easily seen by observing that the drift down of the class 1 queue process, whenever the latter is positive, equals $\lambda_2 + \lambda_3$, which is strictly less than the drift up λ_1. Similarly, the downward drift to 0 of the class 2 queue process is $\lambda_1 + \lambda_4$, which is less than the upward drift λ_2. It remains to consider nodes 3 and 4.

Stability of node 3. If $\bar{Q}_3(0) > 0$, then $\bar{Q}_3$ is strictly positive over an interval $I \subset [0, \infty)$. Over this interval I, the class-2 and class-4 queue processes behave as a fast-time-scale CTMC. Just as Proposition 4, from Theorem 4 we obtain that, if

$\bar{Q}^n(0) \Rightarrow \bar{Q}_3(0)$ and $\bar{Q}_i^n(0) \Rightarrow 0$, $i \neq 3$ in $\mathbb{R}$ as $n \to \infty$, then $\bar{Q}^n \Rightarrow \bar{Q}$ in $\mathbb{D}^5[0, \rho_3)$ as $n \to \infty$, where

$$\begin{cases} \bar{Q}_3(t) = \bar{Q}_3(0) + [\lambda_3 - c\alpha_{(2,4)}]t & 0 \le t < \rho_3, \\ \bar{Q}_i(t) = 0 & 0 \le t < \rho_3, i \in [\![1,5]\!] \setminus \{3\}, \end{cases}$$

where

$$\alpha_{(2,4)} := \left(1 + \frac{\lambda_2}{\lambda_1 + \lambda_4 - \lambda_2} + \frac{\lambda_4}{\lambda_5 + \lambda_2 - \lambda_4}\right)^{-1},$$

$$c := \frac{\lambda_5(\lambda_1 + \lambda_4)}{\lambda_1 + \lambda_4 - \lambda_2} + \frac{\lambda_1(\lambda_5 + \lambda_2)}{\lambda_5 + \lambda_2 - \lambda_4},$$

$$\rho_3 := \frac{\bar{Q}_3(0)}{(c_0 + c_1)\alpha_{(2,4)} - \lambda_3} \qquad \text{if } \alpha_{(2,4)} > \frac{\lambda_3}{c_0 + c_1} \quad \text{and}$$

$$\rho_3 := \infty \qquad \text{otherwise.}$$

Hence, in addition to requiring that $\text{NCOND}_\text{C}(G)$ holds, we must have $\rho_3 < \infty$, so that the fluid limit $\bar{Q}_3$ reaches 0 in finite time. In particular, a necessary condition for the stability of the model is that

$$\lambda_3 < c\alpha_{(2,4)}, \tag{43}$$

which is not implied by the necessary condition in Corollary 3.

Stability of node 4. Since node 5 gives priority to class 4 over class 3, the instantaneous downward drift of the class 4 queue process, at any time t in which it is strictly positive, is $\lambda_5 + \lambda_2 \mathbb{1}_{\{Q_1(t)=0\}}$, while its upward drift is the constant λ_4. Then Theorem 4 implies again that, if $\bar{Q}_4^n(0) \Rightarrow \bar{Q}_4(0)$, for some $\bar{Q}_4(0) > 0$, and $\bar{Q}_i^n(0) \Rightarrow 0$ in $\mathbb{R}$ as $n \to \infty$, $i \neq 4$, then over some interval $I \subset [0, \infty)$ it holds that $\bar{Q}^n \Rightarrow \bar{Q}$ in $\mathbb{D}^5[0, \rho_4)$, where

$$\begin{cases} \bar{Q}_4(t) = \bar{Q}_4(0) + (\lambda_4 - \lambda_5 - \lambda_2\alpha_{(1,3)})t & 0 \le t < \rho_4, \\ \bar{Q}_i(t) = 0 & 0 \le t < \rho_4, i \in [\![1,5]\!] \setminus \{4\}, \end{cases}$$

for

$$\alpha_{(1,3)} := 1 - \frac{\lambda_1}{\lambda_2 + \lambda_3},$$

and

$$\rho_4 := \frac{\bar{Q}_4(0)}{\lambda_5 + \lambda_2\alpha_{(1,3)} - \lambda_4} \qquad \text{if } \alpha_{(1,3)} > (\lambda_4 - \lambda_5)/\lambda_2 \quad \text{and}$$

$$\rho_4 := \infty \qquad \text{otherwise.}$$

Therefore, in order for node 4 to be stable, we require that, in addition to having $\text{NCOND}_\text{C}(G)$ hold, $\rho_4 < \infty$ or, equivalently,

$$\lambda_4 < \lambda_2\alpha_{(1,3)} + \lambda_5. \tag{44}$$

The stability region of the model of Figure 4. Similar to the proof of Proposition 2, one can show that, if each of the necessary conditions for stability of each of the nodes holds, then the model is stable. In particular, we have the following.

PROPOSITION 5. *The model $(G, \lambda, \Phi)_{\mathrm{C}}$ corresponding to the 5-cycle and the matching policy in Figure 4 is stable if and only if $\lambda \in \mathrm{NCOND}_{\mathrm{C}}(G)$ and all three inequalities* (42), (43) *and* (44) *hold.*

Similar to the pendant graph (see Proposition 3), we can check that the stability region of the model is strictly included in $\mathrm{NCOND}_{\mathrm{C}}(G)$. Specifically, we have the following.

PROPOSITION 6. *We have the strict inclusion*:

$$\{\lambda \textit{ satisfying } (42)\} \cap \mathrm{NCOND}_{\mathrm{C}}(G) \subsetneq \mathrm{NCOND}_{\mathrm{C}}(G).$$

PROOF. Fix $\epsilon \in (0, 2/9]$ and set

$$\begin{cases} \lambda_1 = \lambda_2 = \epsilon/2; \\ \lambda_3 = \lambda_4 = 1/4 - \epsilon/8; \\ \lambda_5 = 1/2 - 3\epsilon/4. \end{cases}$$

Clearly, $\lambda \in \mathrm{NCOND}_{\mathrm{C}}(G)$, but (42) does not hold since

$$\lambda_5 - a\tilde{\alpha} = \frac{1}{2} - \frac{3\epsilon}{4} - \left(\frac{1}{2} - \frac{\epsilon}{4}\right)\frac{1/4 + 3\epsilon/8}{1/4 + 7\epsilon/8} = \frac{\epsilon/8(1 - 9\epsilon/2)}{1/4 + 7\epsilon/8} \geq 0. \qquad \square$$

6. Proof of the main result. In this section, we prove Theorem 3, after introducing several key auxiliary results.

6.1. *Graphs induced by separable graphs.* We start by proving the connection between separable graphs and the graphs investigated in Section 5.

LEMMA 3. *For any connected graph G*:

(i) *If G is non-bipartite and nonseparable, then it induces a pendant graph or an odd cycle of size* 5 *or more.*

(ii) *If G is separable, then it does not induce a pendant graph, nor any odd cycle of size* 5 *or more.*

PROOF. *Proof of* (i). Let G be a non-bipartite and nonseparable graph. It is a classical result of graph theory (see, e.g., Theorem 13.2.1 of [20]) that G contains an odd cycle $\breve{G}$ as a subgraph. We consider two cases separately: $\breve{G}$ is a triangle or $\breve{G}$ is of size 5 or more.

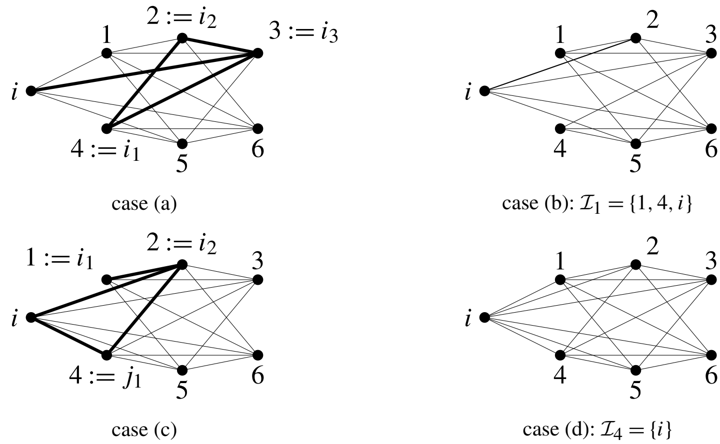

FIG. 5. *Completing a separable graph with one node always yields to a separable graph* [(b) *and* (d)], *or a nonseparable graph inducing a pendant graph* [(a) *and* (c)].

Case 1. Assume that $\check{G}$ is a triangle. It is not possible that $\check{G} = G$, because then G would be a complete graph, and thereby a separable graph. Thus, G has other nodes, in addition to the three ones of $\check{G}$.

We prove the claim by induction. Assume that G induces a graph $\bar{G}$ that consists of a separable graph $\check{G}$ of order $q \geq 3$ (the base case of the induction being $\check{G} = \check{G}$), connected to another node, which we denote as node i. We have the following four alternatives, which are depicted in an example in Figure 5.

(a) There may exist two independent sets $\mathcal{I}_1$ and $\mathcal{I}_2$ of $\check{G}$, and two nodes $i_1 \in \mathcal{I}_1$ and $i_2 \in \mathcal{I}_2$, such that $i \not- i_1$ and $i \not- i_2$. In that case, as $\check{G}$ is of order 3 or more, there exist an independent set $\mathcal{I}_3$, which is different from $\mathcal{I}_1$ and $\mathcal{I}_2$, and a node $i_3 \in \mathcal{I}_3$, such that $i - i_3$. Then $\check{G}$ induces the pendant graph $i - i_3 - i_1 - i_2$, where i is only adjacent to i_3 and (i_1, i_2, i_3) form a triangle. In particular, G induces a pendant graph.

(b) There may exist a maximal independent set $\mathcal{I}_1$ of $\check{G}$ such that $i - k$ for any $k \in \check{G} \setminus \mathcal{I}_1$, and $i \not- i_1$ for any $i_1 \in \mathcal{I}_1$. Then $\bar{G}$ is again a separable graph of order q, having the same maximal independent sets as $\check{G}$, except that $\mathcal{I}_1$ is replaced by $\mathcal{I}_1 \cup \{i\}$.

(c) There may exist a maximal independent set $\mathcal{I}_1$ of $\check{G}$ such that $i - k$ for any $k \in \check{G} \setminus \mathcal{I}_1$, and $i \not- i_1$ for some (not necessarily unique) $i_1 \in \mathcal{I}_1$ and $i - j_1$ for some (again, not necessarily unique) $j_1 \in \mathcal{I}_1$. In that case, $\bar{G}$ induces the pendant graph $i_1 - i_2 - j_1 - i$, where i_1 is only adjacent to i_2 and (i_2, j_1, i) form a triangle.

(d) We may have $i - j$ for any node $j \in \check{G}$. Then $\bar{G}$ is a separable graph of order $q + 1$ whose maximal independent sets are those of $\check{G}$, plus the independent set $\mathcal{I}_{q+1} := \{i\}$.

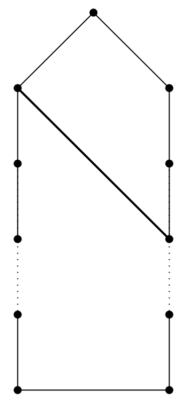

FIG. 6. *Drawing an edge between two nodes of an odd cycle creates an odd cycle.*

To summarize, in cases (a) and (c), $\bar{G}$ and, therefore, G, induce a pendant graph. In cases (b) and (d) $\bar{G}$ is separable, and we cannot have $G = \bar{G}$. Thus, there exists another node in G that is connected to $\bar{G}$, and we can re-iterate the same procedure for $\breve{G} \equiv \bar{G}$. Eventually, some $\bar{G}$ induced in G will exhibit either case (a) or (c), otherwise G would be separable. This concludes the proof in this first case.

Case 2. Now assume that $\breve{G}$ is of length 5 or more. Assume that G does not induce $\breve{G}$. Therefore, G admits an edge (i, j), where i and j are two nodes of $\breve{G}$. But drawing an edge inside an odd cycle always creates an odd cycle and an even cycle (see Figure 6).

By induction on the added edges, $\breve{G}$ finally induces an odd cycle of length $2p + 1 \geq 3$. If $2p + 1 \geq 5$, we are done. If $2p + 1 = 3$, we are back in case 1.

Proof of (ii). We prove the result by contradiction. Let G be a separable graph of order q and let $\mathcal{I}_1, \dots, \mathcal{I}_q$ be its disjoint maximal independent sets. Then for any two nodes i, j of G, the relation $i \not- j$ implies that i and j belong to the same independent set $\mathcal{I}_k$, for some $k \in [\![1, q]\!]$, but to no other independent set.

First, take the contradictory assumption that G induces a pendant graph $\breve{G}$, and label the nodes of $\breve{G}$ as in Figure 2. Suppose that $3 \in \mathcal{I}_i$. All the neighbors of 3 in $\breve{G}$ (and thus in G) cannot be in $\mathcal{I}_i$, so there exist $j, k, \ell \in [\![1, q]\!] \setminus \{i\}$, such that $4 \in \mathcal{I}_j$, $1 \in \mathcal{I}_k$ and $2 \in \mathcal{I}_\ell$. As $4 \not- 1$, we have $j = k$. Similarly, we have $j = \ell$, and thus $k = \ell$. But, since $1-2$, nodes 1 and 2 do not belong to the same independent set - a contradiction. Therefore, G cannot induce a pendant graph.

We next assume that G induces the $2p + 1$-cycle $\breve{G}$, with $p \geq 2$ (so that $q \leq 2p + 1$). For simplicity, label the nodes of $\breve{G}$ as $1, 2, \dots, 2p + 1$, in a way that

$$1-2-3-\cdots-2p-(2p+1)-1.$$

For all $j \in [\![1, 2p + 1]\!]$, let i_j be such that $j \in \mathcal{I}_{i_j}$. As $1-2, 2-3$ and $3 \not- 1$, we have $i_1 \neq i_2$ and $i_1 = i_3$. In general, we see that any odd node k is in $\mathcal{I}_{i_1}$, whereas any even node ℓ is in $\mathcal{I}_{i_2}$. It follows that nodes 1 and $2p + 1$ both belong to $\mathcal{I}_{i_1}$, but since $(2p+1)-1$ we again arrive at a contradiction, implying that G cannot induce an odd cycle of size 5 or more. □

6.2. *Nonchaoticity of matching queues*. As was already discussed, Theorem 3 is proved by employing Lemma 3, after showing that the pendant graph and the 5-cycle graph can be unstable, even if NCOND holds. However, if $\breve{G} = (\breve{\mathcal{V}}, \breve{\mathcal{E}})$ is induced in $G = (\mathcal{V}, \mathcal{E})$, and if we assume that i_0 is an unstable node of a matching queue on $\breve{G}$ when it is considered in isolation, then it might not be unstable in a matching queue on G. (In our case, $\breve{G}$ is either the pendant graph of the 5-cycle.) Therefore, to construct an unstable matching queue on G itself, we must show that the effect of the arrivals to $\mathcal{V} \setminus \breve{\mathcal{V}}$ can be controlled so that node i_0 remains unstable in the matching queue on G. If we think of the effect of the arrivals of items of $\mathcal{V} \setminus \breve{\mathcal{V}}$, as a perturbation of the number of items in the matching queue on $\breve{G}$ when considered in isolation, then we must show that perturbations of the matching queue on $\breve{G}$ do not get amplified within the matching queue on G. We refer to such a property as *nonchaoticity* of matching queues. See Lemma 5 below for the precise statement.

To prove this nonchaoticity property, we first prove an auxiliary result. Let $\|\cdot\|$ denote the 1-norm on $\mathbb{R}^p$, $p \in \mathbb{Z}_{++}$,

$$\|x\| = \sum_{i=1}^{p} |x_i|, \qquad x \in \mathbb{R}^p.$$

We say that two matching queues $(G, \lambda, \Phi)_C$ and $(G', \lambda', \Phi')_C$ such that $|\mathcal{V}| = |\mathcal{V}'|$ and $\lambda = \lambda'$ have *the same input*, if both are constructed using the same $|\mathcal{V}|$ Poisson processes (same sample paths of the arrival process).

For a given matching queue $(G, \lambda, \Phi)_C$ having a queue process defined on a state space $\mathbb{G}$, let $Q^z = \{Q^z(t) : t \geq 0\}$ denote the queue process when the initial condition is $Q^z(0) = z$, $z \in \mathbb{G}$.

LEMMA 4. *For any matching queue $(G, \lambda, \Phi)_C$ and any initial conditions x, y in $\mathbb{G}$, if the two systems have the same input we have that*

$$\|Q^x(t) - Q^y(t)\| \leq \|x - y\|, \qquad t \geq 0.$$

We remark that the Lipschitz-continuity property stated above follows (as will become clear in the proof) from a specific property of the DTMC embedded in arrival-time epoches, known as *nonexpensiveness* in the literature on stochastic recursions; see, for example, Section 2.11 in [3].

PROOF OF LEMMA 4. Let $T_1 < T_2 < \cdots$ be the arrival times of elements to the system. With some abuse of notation, denote for the time being, $Q^x(0) = x$, $Q^y(0) = y$, and for all $n \geq 1$, $Q^x(n) := Q^x(T_n)$ and $Q^y(n) := Q^y(T_n)$. Since both processes Q^x and Q^y are constant between arrival times, it suffices to show the result at any time T_n. We reason by induction. Let $n \in \mathbb{Z}_+$, and assume that

$$\|Q^x(n) - Q^y(n)\| \le \|x - y\|.$$

Let j be the class of the item drawn at time T_n. We have the following alternatives:

1. Assume that the new item is matched in both systems, with an item of the same class $k \in \mathcal{E}(j)$. Then in both cases the kth coordinate decreases by one and we have $\|Q^x(n+1) - Q^y(n+1)\| = \|Q^x(n) - Q^y(n)\|$.

2. Assume that the new item is matched with k^x (resp., k^y) in the system initiated by x (resp., by y), where $k^x \neq k^y$. Then, by the definition of a priority policy, we must have that

$$\big(\big(Q^x_{k^x}(n) > 0\big) \text{ and } \big(Q^y_{k^x}(n) = 0\big)\big) \quad \text{or} \quad \big(\big(Q^x_{k^y}(n) = 0\big) \text{ and } \big(Q^y_{k^y}(n) > 0\big)\big)$$

(or both), since otherwise, the new arrival of class j would be matched with the same item in both systems.

Assume that we are in the first case, the other one is symmetric. We have

$$\begin{aligned}
&\|Q^x(n+1) - Q^y(n+1)\| \\
&\quad = \|Q^x(n) - Q^y(n)\| - \big|Q^x_{k^x}(n) - Q^y_{k^x}(n)\big| + \big|\big(Q^x_{k^x}(n) - 1\big) - Q^y_{k^x}(n)\big| \\
&\qquad - \big|Q^x_{k^y}(n) - Q^y_{k^y}(n)\big| + \big|\big(Q^y_{k^y}(n) - 1\big) - Q^x_{k^y}(n)\big| \\
&\quad \le \|Q^x(n) - Q^y(n)\| - Q^x_{k^x}(n) + \big(Q^x_{k^x}(n) - 1\big) - \big|Q^x_{k^y}(n) - Q^y_{k^y}(n)\big| \\
&\qquad + \big|Q^y_{k^y}(n) - Q^x_{k^y}(n)\big| + 1 \\
&\quad = \|Q^x(n) - Q^y(n)\|.
\end{aligned}$$

3. Assume that the new arrival to class j is matched in the system initiated by x, say, with $k^x \in \mathcal{E}(j)$, but not in the one initiated by y (the other way around is symmetric). Then we must have $Q^y_{k^x}(n) = 0$ and in turn

$$\begin{aligned}
&\|Q^x(n+1) - Q^y(n+1)\| \\
&\quad = \|Q^x(n) - Q^y(n)\| - \big|Q^x_{k^x}(n) - Q^y_{k^x}(n)\big| + \big|\big(Q^x_{k^x}(n) - 1\big) - Q^y_{k^x}(n)\big| \\
&\qquad - \big|Q^x_j(n) - Q^y_j(n)\big| + \big|\big(Q^y_j(n) + 1\big) - Q^x_j(n)\big| \\
&\quad \le \|Q^x(n) - Q^y(n)\| - Q^x_{k^x}(n) + \big(Q^x_{k^x}(n) - 1\big) - \big|Q^y_j(n) - Q^x_j(n)\big| \\
&\qquad + \big|Q^y_j(n) - Q^x_j(n)\big| + 1 \\
&\quad = \|Q^x(n) - Q^y(n)\|.
\end{aligned}$$

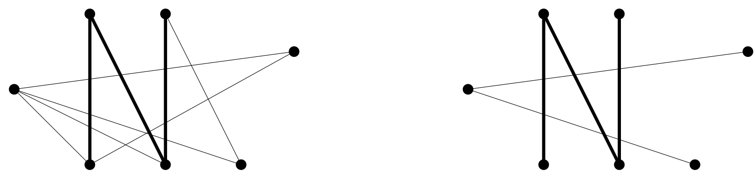

FIG. 7. *Right, the disconnected graph $\tilde{G}$, if $\breve{G}$ is the "N" graph.*

4. Finally, assume that $Q^x_k(n) = Q^y_k(n) = 0$ for any $k \in \mathcal{E}(j)$, that is, the incoming item is not matched upon arrival in both systems. Then the jth coordinate increases by one in both cases, so $\|Q^x(n+1) - Q^y(n+1)\| = \|Q^x(n) - Q^y(n)\|$. □

For a connected graph $G = (\mathcal{V}, \mathcal{E})$ and $\breve{\mathcal{V}} \cup \hat{\mathcal{V}}$ a partition of $\mathcal{V}$, we denote by $\breve{G}$ and $\hat{G}$ the graphs induced, respectively, by $\breve{\mathcal{V}}$ and $\hat{\mathcal{V}}$ in G. Then the *disconnected* graph $\tilde{G}$ corresponding to the partition $\breve{\mathcal{V}} \cup \hat{\mathcal{V}}$ is the graph $\tilde{G} = (\mathcal{V}, \tilde{\mathcal{E}})$ such that $\tilde{\mathcal{E}} = \mathcal{E} \setminus ((\breve{\mathcal{V}} \times \hat{\mathcal{V}}) \cup (\hat{\mathcal{V}} \times \breve{\mathcal{V}}))$. In other words, the graph $\tilde{G}$ is obtained from G by erasing the edges between elements of $\breve{\mathcal{V}}$ and $\hat{\mathcal{V}}$; an example is depicted in Figure 7.

Consider a connected graph G, a partition $\breve{\mathcal{V}} \cup \hat{\mathcal{V}}$ of $\mathcal{V}$, and the disconnected graph $\tilde{G}$ as defined above. For two priority matching policies Φ on G and $\tilde{\Phi}$ on $\tilde{G}$, we say that the restrictions to $\breve{G}$ of Φ and $\tilde{\Phi}$ coincide if [recall (10)]

$$\Phi_j(i) \cap \breve{\mathcal{V}} = \tilde{\Phi}_j(i) \cap \breve{\mathcal{V}}, \qquad i, j \in \breve{\mathcal{V}}.$$

That is, for any elements i, j, k of $\breve{\mathcal{V}}$, j prioritizes k over i according to Φ if and only if it does so according to $\tilde{\Phi}$. We then have the following result.

LEMMA 5. *Let $G = (\mathcal{V}, \mathcal{E})$ be a connected graph, $\breve{\mathcal{V}} \cup \hat{\mathcal{V}}$ be a partition of $\mathcal{V}$ and $\tilde{G}$ be the corresponding disconnected graph. Consider the two matching queues $\Sigma := (G, \lambda, \Phi)_{\mathrm{C}}$ and $\tilde{\Sigma} := (\tilde{G}, \lambda, \tilde{\Phi})_{\mathrm{C}}$, where the restrictions to $\breve{G}$ of the priority matching policies Φ and $\tilde{\Phi}$ coincide. Let Q and $\tilde{Q}$ be the respective queue processes of Σ and $\tilde{\Sigma}$. Then, if $Q(0) = \tilde{Q}(0)$, $Q_j(0) = \tilde{Q}_j(0) = 0$ for all $j \in \hat{\mathcal{V}}$, and the two systems have the same input, we have that*

$$\sum_{i \in \breve{\mathcal{V}}} \left| Q_i(t) - \tilde{Q}_i(t) \right| \leq \hat{N}(t), \qquad t \geq 0,$$

where $\hat{N}(t)$ is the number of arrivals of items to $\hat{\mathcal{V}}$ up to t.

PROOF. Let $\hat{T}_n$, $n \geq 1$, be the points of $\hat{N}$ (i.e., the arrival times of elements of $\hat{\mathcal{V}}$). Let also $\hat{U}_0 = 0$ and for all $n \geq 0$,

$$\begin{aligned}\hat{U}_{n+1} = \inf\{ & t \geq \hat{U}_n; \text{a matching occurs at } t \text{ in } \Sigma \\ & \text{between an element of } \breve{\mathcal{V}} \text{ and one of } \hat{\mathcal{V}}\}.\end{aligned}$$

Notice that the times $\hat{U}_n, n \geq 1$ can either coincide with points of $\hat{N}$, or with arrival times of elements of $\breve{\mathcal{V}}$ that are matched with an element of $\hat{\mathcal{V}}$ in Σ.

Let $n \geq 0$. In the time interval $(\hat{U}_n, \hat{U}_{n+1})$, the restrictions to $\breve{G}$ of Σ and of $\tilde{\Sigma}$ both behave exactly as the matching queue $(\breve{G}, \lambda_{\breve{\mathcal{V}}}, \breve{\Phi})_{\mathrm{C}}$, where $\breve{\Phi}$ is the restriction of Φ and $\tilde{\Phi}$ to $\breve{G}$. Therefore, we can apply Lemma 4 to the latter model and to the initial conditions $Q_{\breve{\mathcal{V}}}(\hat{U}_n)$ and $\tilde{Q}_{\breve{\mathcal{V}}}(\hat{U}_n)$, where we recall that $X(t^-)$ denotes the left-limit of X at t. We conclude that for any $n \geq 0$

$$\sum_{i \in \breve{\mathcal{V}}} \left| Q_i(\hat{U}_{n+1}^-) - \tilde{Q}_i(\hat{U}_{n+1}^-) \right| \leq \sum_{i \in \breve{\mathcal{V}}} \left| Q_i(\hat{U}_n) - \tilde{Q}_i(\hat{U}_n) \right|. \tag{45}$$

Furthermore, we have the following alternatives:

1. If $\hat{U}_{n+1}$ is an arrival time of an item of class in $\hat{\mathcal{V}}$ (it coincides with some $\hat{T}_k$), this item is matched immediately with an items of a class in $\breve{\mathcal{V}}$ (say, of class j), which leaves the buffer of Σ. Hence, in the restriction of Q to $\breve{\mathcal{V}}$, all coordinates remain unchanged except for the jth coordinate, which decreases by 1. On the other hand, the restriction of $\tilde{Q}$ to $\breve{\mathcal{V}}$ does not change.
2. If $\hat{U}_{n+1}$ is an arrival time of an item of class $j \in \breve{\mathcal{V}}$ that is matched immediately with an element of $\hat{\mathcal{V}}$ in Σ, then the buffer content of Σ restricted to $\breve{\mathcal{V}}$ does not change, but that of $\tilde{\Sigma}$ does:

2a. if the arriving item is matched in $\tilde{\Sigma}$ with an item of class $k \in \mathcal{E}(j) \cap \breve{\mathcal{V}}$, the kth coordinate of $\tilde{Q}$ decreases by 1, while all other coordinates remain unchanged;
2b. if the arriving item does not find a match in $\breve{\mathcal{V}}$ in $\tilde{\Sigma}$, it is stored in the buffer and the jth coordinate of $\tilde{Q}$ increases by 1, while all other coordinates remain unchanged.

In all cases, we obtain that

$$\sum_{i \in \breve{\mathcal{V}}} \left| Q_i(\hat{U}_{n+1}) - \tilde{Q}_i(\hat{U}_{n+1}) \right| \leq \sum_{i \in \breve{\mathcal{V}}} \left| Q_i(\hat{U}_{n+1}^-) - \tilde{Q}_i(\hat{U}_{n+1}^-) \right| + 1. \tag{46}$$

Finally, gathering (45) and (46), as $Q(0)$ and $\tilde{Q}(0)$ coincide, we obtain that for any $t \geq 0$,

$$\sum_{i \in \breve{\mathcal{V}}} \left| Q_i(t) - \tilde{Q}_i(t) \right| \leq \sum_{n \geq 1} \mathbb{1}_{\{\hat{U}_n \leq t\}}. \tag{47}$$

Finally, observe that the instants $\hat{U}_n$ are departure times of items of class in $\hat{\mathcal{V}}$. As there are no such items in storage initially, the number of such instants up to t cannot exceed the number of arrivals of items of class in $\hat{\mathcal{V}}$ up to t. Thus, we have

$$\sum_{n \geq 1} \mathbb{1}_{\{\hat{U}_n \leq t\}} \leq \sum_{n \geq 1} \mathbb{1}_{\{\hat{T}_n \leq t\}},$$

which, together with (47), completes the proof. □

6.3. *Proof of Theorem* 3. We are now in position to prove Theorem 3. The strategy is the following: We fix a non-bipartite and nonseparable graph $G \in \mathscr{G}_7^c$. Lemma 3 entails that such a graph induces particular types of graphs, for which the corresponding matching queues $(G, \lambda, \Phi)_{\mathrm{C}}$ are proved to be possibly unstable for some $\lambda \in \textsc{Ncond}_{\mathrm{C}}(G)$ and some matching policy Φ (Sections 5.1 and 5.2). The instability of the matching queue under consideration can then be deduced from the nonchaoticity property in Lemma 5.

PROOF OF THEOREM 3. Let $G = (\mathcal{V}, \mathcal{E})$ be a non-bipartite and non-separable graph in $\mathscr{G}_7^c$. By Lemma 3, G induces a pendant graph or an odd cycle of size 5. Let $\breve{G} = (\breve{\mathcal{V}}, \breve{\mathcal{E}})$ be that induced subgraph. Then there exists an arrival-rate vector $\breve{\lambda} \in (\mathbb{R}_{++})^{|\breve{\mathcal{V}}|}$ and a matching policy $\breve{\Phi}$, such that the matching queue $(\breve{G}, \breve{\lambda}, \breve{\Phi})_{\mathrm{C}}$ is unstable, whereas $\breve{\lambda} \in \textsc{Ncond}_{\mathrm{C}}(\breve{G})$. (This latter claim follows from Proposition 3 if $\breve{G}$ is a pendant graph, and from Proposition 6 if $\breve{G}$ is a 5-cycle.) We fix the latter $\breve{\lambda}$ until the end of the proof and set [recall (2)]

$$\tau := \min\{\bar{\breve{\lambda}}_{\breve{\mathcal{E}}(\breve{\mathcal{I}})} - \bar{\breve{\lambda}}_{\breve{\mathcal{I}}} : \breve{\mathcal{I}} \in \mathbb{I}(\breve{G})\}. \tag{48}$$

Let $\hat{\mathcal{V}} = \mathcal{V} \setminus \breve{\mathcal{V}}$ and denote $\hat{G} = (\hat{\mathcal{V}}, \hat{\mathcal{E}})$ the induced subgraph in G. In view of Proposition 1 and Proposition 4, there exists a node $i_0 \in \breve{V}$ ($i_0 = 4$ if $\breve{G}$ is a pendant graph and $i_0 = 5$ if $\breve{G}$ is a 5-cycle) and a measure $\breve{\pi}$ on $\mathbb{E}_2$ such that the drift of the i_0-coordinate of the fluid limit reads

$$\beta := \breve{\lambda}_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \breve{\lambda}_j \breve{\pi}(\mathcal{P}_j^{\mathcal{S}}(i_0)) > 0. \tag{49}$$

Set $\gamma = \frac{1}{2}(\tau \wedge \beta)$, and let $\lambda \in \mathbb{R}_{++}^{|\mathcal{V}|}$ satisfy

$$\begin{cases} \lambda_{\breve{\mathcal{V}}} = \breve{\lambda}; \\ \bar{\lambda}_{\hat{\mathcal{V}}} \le \gamma. \end{cases} \tag{50}$$

We first prove that $\lambda \in \textsc{Ncond}_{\mathrm{C}}(G)$. For $\mathcal{I} \in \mathbb{I}(G)$, observe that

$$\breve{\mathcal{E}}(\mathcal{I} \cap \breve{\mathcal{V}}) \cup \hat{\mathcal{E}}(\mathcal{I} \cap \hat{\mathcal{V}}) \subset \mathcal{E}(\mathcal{I}).$$

Therefore,

$$\bar{\lambda}_{\mathcal{E}(\mathcal{I})} - \bar{\lambda}_{\mathcal{I}} \ge \bar{\lambda}_{\breve{\mathcal{E}}(\mathcal{I} \cap \breve{\mathcal{V}})} + \bar{\lambda}_{\hat{\mathcal{E}}(\mathcal{I} \cap \hat{\mathcal{V}})} - \bar{\lambda}_{\mathcal{I} \cap \breve{\mathcal{V}}} - \bar{\lambda}_{\mathcal{I} \cap \hat{\mathcal{V}}}. \tag{51}$$

First, as $\breve{G}$ is induced in G, $\mathcal{I} \cap \breve{\mathcal{V}} \in \mathbb{I}(\breve{G})$. Thus, (48) implies that

$$\bar{\lambda}_{\breve{\mathcal{E}}(\mathcal{I} \cap \breve{\mathcal{V}})} - \bar{\lambda}_{\mathcal{I} \cap \breve{\mathcal{V}}} > \gamma.$$

Moreover, from (50) we clearly have that

$$\bar{\lambda}_{\hat{\mathcal{E}}(\mathcal{I} \cap \hat{\mathcal{V}})} - \bar{\lambda}_{\mathcal{I} \cap \hat{\mathcal{V}}} \ge -\gamma.$$

These two observations, together with (51), yield $\bar{\lambda}_{\mathcal{E}(\mathcal{I})} - \bar{\lambda}_{\mathcal{I}} > 0$. Therefore, $\lambda \in \text{NCOND}_{\text{C}}(G)$.

It remains to construct a matching policy Φ on G rendering the matching queue $(G, \lambda, \Phi)_{\text{C}}$ unstable. To this end, it suffices to consider any Φ whose restriction to $\breve{G}$ is $\breve{\Phi}$. Consider the process $\tilde{Q}$ constructed in Lemma 5, associated with the disconnected graph corresponding to the partition $\breve{\mathcal{V}} \cup \hat{\mathcal{V}}$. If $Q(0)$ and $\tilde{Q}(0)$ are equal, and satisfy $Q_j(0) = \tilde{Q}_j(0) = 0$ for all $j \in \hat{\mathcal{V}}$, then it follows from Lemma 5 that

$$Q_{i_0} \geq_{\text{st}} \tilde{Q}_{i_0} - \hat{N}, \tag{52}$$

where $\hat{N}$ denotes again the arrival process of items of class in $\hat{\mathcal{V}}$. For $n \geq 1$, let $\hat{N}^n$ be a Poisson process with intensity $n\bar{\lambda}_{\hat{\mathcal{V}}}$. Then, for any initial condition $\breve{Q}^n(0)$ such that $\breve{Q}^n_{i_0}(0) = nx$, the following convergence holds in $\mathbb{D}^{|\breve{\mathcal{V}}|}$:

$$\left(\frac{1}{n}\breve{Q}^n(t) - \frac{1}{n}\hat{N}^n(t)\mathbf{e}_{i_0}, t \geq 0\right)$$
$$\Rightarrow \quad \left(\left(x + \breve{\lambda}_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \breve{\lambda}_j \breve{\pi}\big(\mathcal{P}_j^{\mathcal{S}}(i_0)\big) - \bar{\lambda}_{\hat{\mathcal{V}}}\right) t\mathbf{e}_{i_0}, t \geq 0\right).$$

Together with (49) and (50), the above convergence implies that the Markov process $\breve{Q} - \hat{N}\mathbf{e}_{i_0}$ is transient and that $\breve{Q}_{i_0} - \hat{N} \Rightarrow \infty$ as $t \to \infty$.

Finally, observe that by definition, the restriction to $\breve{\mathcal{V}}$ of $\tilde{Q}$ has the same distribution as $\breve{Q}$ if $\breve{Q}(0)$ is set equal to the restriction to $\breve{\mathcal{V}}$ of $\tilde{Q}(0)$. Thus, the $\mathbb{Z}_+^{|\mathcal{V}|}$-valued Markov process $\tilde{Q} - \hat{N}\mathbf{e}_{i_0}$ is transient and its i_0-coordinate converges in distribution to ∞ as $t \to \infty$. By (52), this is also the case for the i_0-coordinate of Q, so that Q is transient. □

REMARK 2. Observe that, for any non-bipartite and nonseparable graph G, the the proof of Theorem 3 not only shows the existence of a nonmaximal priority policy on G, but also provides a simple way of constructing that policy. Specifically, we have proven that for *any* priority matching policy Φ, if the restriction of Φ to the induced sub-graph $\breve{G}$ is nonmaximal for $\breve{G}$, then Φ is also nonmaximal for G. Consequently, (i) if G induces a pendant graph (whose nodes are labeled as in Figure 2), then any priority policy Φ on G such that node 3 prioritizes nodes 1 and 2 over node 4 is nonmaximal; (ii) if G induces a 5-cycle (whose nodes are labeled as in Figure 4), then any priority policy Φ on G such that node 3 prioritizes node 1 over 5, node 4 prioritizes node 2 over 5, node 1 prioritizes 2 over 3 and node 2 prioritizes 1 over 4, is nonmaximal.

7. Proof of the FWLLN. Our proof of the FWLLN will follow the precompactness approach [4, 34]. In particular, we will show that $\{\bar{Q}^n : n \geq 1\}$ is tight in $\mathbb{D}^{|\mathcal{V}|}$ and uniquely characterize the limit.

Recall that $\mathbb{G}^{\mathcal{S}}$ defined by (16) denotes the state space of S^n for all $n \geq 1$. For $\delta > 0$, let $\mathcal{M} := \mathcal{M}(\mathcal{L}_\delta)$ denote the space of (finite) measures on the space $\mathcal{L}_\delta := [0, \delta) \times \mathbb{G}^{\mathcal{S}}$ such that $\mu([0,t] \times \mathbb{G}^{\mathcal{S}}) = t$ for all $\mu \in \mathcal{M}$ and $t \in [0, \delta)$, endowed with the Prohorov metric [4]. Next, define a sequence of random elements $\{\nu^n : n \geq 1\} \subset \mathcal{M}$ via

$$\nu^n([0,t] \times \mathcal{Z}) := \int_0^t \mathbb{1}_{\mathcal{Z}}(S^n(u-))\,du, \qquad 0 \leq t < \delta, \mathcal{Z} \subseteq \mathbb{G}^{\mathcal{S}}. \tag{53}$$

LEMMA 6. *If $\bar{Q}^n(0)$ is tight in $\mathbb{R}^{|\mathcal{V}|}$, then $\{(\bar{Q}^n, \nu^n) : n \geq 1\}$ is tight in $\mathbb{D}^{|\mathcal{V}|}[0,\delta) \times \mathcal{M}$, for δ in Lemma 2. Moreover, for any limit point $(\bar{Q}, \nu)$,*

$$\nu([0,t] \times \mathcal{Z}) = \int_0^t p_u(\mathcal{Z})\,du; \qquad 0 \leq t < \delta, \mathcal{Z} \subseteq \mathbb{G}^{\mathcal{S}} \tag{54}$$

for some family of probability measures $\{p_u : 0 \leq u \leq t < \delta\}$ on $\mathbb{G}^{\mathcal{S}}$.

PROOF. Tightness of the vector $\{(\bar{Q}^n, \nu^n) : n \geq 1\}$ follows from the tightness of each component separately; for example, Theorem 11.6.7 in [34]. We start by showing that $\{\nu^n : n \geq 1\}$ is tight in $\mathcal{M}$.

It follows from (28) that, for any compact set $\mathcal{Z} \subset \mathbb{G}^{\mathcal{S}}$ and $0 \leq t < \delta$,

$$\begin{aligned} E[\nu^n([0,t] \times \mathcal{Z})] &= E\left[\int_0^t \mathbb{1}_{\mathcal{Z}}(S^n(u-))\,du\right] \\ &= E\left[\int_0^t \mathbb{1}_{\mathcal{Z}}(\chi^n(u-))\,du\right] \\ &= E\left[\frac{1}{n}\int_0^{nt} \mathbb{1}_{\mathcal{Z}}(\chi(u-))\,du\right] \\ &\to P(\chi(\infty) \in \mathcal{Z})t \\ &= \pi(\mathcal{Z})t. \end{aligned} \tag{55}$$

The limit in (55) holds as $n \to \infty$ and follows from the ergodicity of χ and the bounded convergence theorem.

Take $\epsilon > 0$ and $t \in [0, \delta)$. Observe that, for any finite N and all $n \leq N$, the right-hand side of the second equality in (55) implies that for large-enough $c \in \mathbb{Z}_+$ and for $K_c := [0, c]^{|\mathcal{S}|}$, it holds that $E[\nu^n([0,t] \times K_c)] \geq (1-\epsilon)t$. The limit in (55) shows that this latter inequality holds for all n, that is, we have

$$\inf_n E[\nu^n([0,t] \times K_c)] \geq (1-\epsilon)t,$$

for a sufficiently large c. Hence, tightness of $\{\nu^n : n \geq 1\}$ follows from Lemma 1.3 in [19]. The fact that any limit point of this tight sequence has the form in (54) follows from Lemma 1.4 in [19].

We next show that $\{\bar{Q}^n : n \geq 1\}$ is $\mathbb{C}$-tight in $\mathbb{D}^{|\mathcal{V}|}$, that is, it is tight and any of its limit points is continuous. We first note that, since all the jumps in $\bar{Q}^n$ are of size $1/n$, and are thus converging to 0 as $n \to \infty$, we can work with the modulus of continuity defined for continuous functions (see, e.g., [4], page 123)

$$w(y, \eta, T) := \sup\{|y(t_2) - y(t_1)| : 0 \le t_1 \le t_2 \le T, |t_2 - t_1| \le \eta\},$$
$$\eta > 0, T > 0.$$

Hence, we can establish the result by applying Theorem 11.6.3 in page 389 in [34]. Conditions (6.3) in that theorem, namely, tightness of $\{\bar{Q}^n(0) : n \ge 1\}$ in $\mathbb{R}^{|\mathcal{V}|}$ is assumed. To show that condition (6.4) in [34], Theorem 11.6.3, also holds, observe that, for all $i \in \mathcal{V}$ and $n \ge 1$,

$$|\bar{Q}_i^n(t_2) - \bar{Q}_i^n(t_1)| \le \frac{N_i^n(t_2) - N_i^n(t_1)}{n} + \frac{Z_i^n(t_2) - Z_i^n(t_1)}{n},$$

where N_i^n denotes the time-scaled Poisson arrival process to node i (see Section 4.3), and $Z_i^n(t)$, $n \ge 1$, is the number of type-i items that were matched (and left the system) by time t. Then we have that

$$w(\bar{Q}_i^n, \eta, T) \le w(N_i^n/n, \eta, T) + w(Z_i^n/n, \eta, T), \qquad \eta, T > 0.$$

It follows from the representation of Q_i^n in (12) that the oscillations of $\bar{Q}_i^n$ are bounded by those of the time-scaled Poisson processes N_i^n and N_j^n, $j \in \mathcal{E}(i)$. Hence,

$$w(\bar{Q}_i^n, \eta, T) \le w(N_i^n/n, \eta, T) + \sum_{j \in \mathcal{E}(i)} w(N_j^n/n, \eta, T), \qquad j \in \mathcal{E}(i). \tag{56}$$

By the FWLLN for Poisson processes, all the moduli of continuity for the scaled Poisson processes in (56) are controlled, so that, for every $\epsilon > 0$ and $\zeta > 0$, there exists $\eta > 0$ and $n_0 \in \mathbb{Z}_+$, such that

$$P(w(\bar{Q}_i^n, \eta, T) \ge \epsilon) < \zeta \qquad \text{for all } n \ge n_0.$$

Thus, $\{\bar{Q}_i^n : n \ge 1\}$ is $\mathbb{C}$-tight for each $i \in \mathcal{V}$, implying that $\{\bar{Q}^n : n \ge 1\}$ is $\mathbb{C}$-tight in $\mathbb{D}^{|\mathcal{V}|}$, and in particular, it is $\mathbb{C}$-tight in $\mathbb{D}^{|\mathcal{V}|}[0, \delta)$. □

In the proof of Theorem 4, we employ the standard result that, for all $i \in \mathcal{V}$ and $n \ge 1$, the following process is a square integrable martingale (with respect to the filtration generated by the Poisson processes); see, for example, [25]:

$$\begin{aligned} M_i^n(t) := &\sum_{j \in \mathcal{E}(i)} \int_0^t \mathbb{1}_{\mathcal{N}_i}(Q^n(s-)) \mathbb{1}_{\mathcal{P}_j(i)}(Q^n(s-))\, dN_j^n(s) \\ &- \int_0^t \mathbb{1}_{\mathcal{N}_i}(Q^n(s-)) \Big(\sum_{j \in \mathcal{E}(i)} n\lambda_j \mathbb{1}_{\mathcal{P}_j(i)}(Q^n(s-)) \Big)\, ds, \qquad t \ge 0. \end{aligned} \tag{57}$$

PROOF OF THEOREM 4. The assumed convergence of the initial condition $\bar{Q}^n(0)$ implies that it is also tight in $\mathbb{R}^{|\mathcal{V}|}$. To characterize the limit, let us first consider $\bar{Q}^n_{i_0}$. Fix $t < \delta$ and $n \geq n_0$, where n_0 is defined by Lemma 2. Then (12) can be written as follows for node i_0:

$$\begin{aligned}\bar{Q}^n_{i_0}(t) &= \bar{Q}^n_{i_0}(0) + \int_0^t \mathbb{1}_{\mathcal{O}_{i_0}}\big(Q(s-)\big)\,dN^n_{i_0}(s)\\ &\quad - \sum_{j\in\mathcal{E}(i_0)} \lambda_j \int_0^t \mathbb{1}_{\mathcal{P}_j(i_0)}\big(Q^n(s)\big)\,ds - M^n_{i_0}(t)/n,\end{aligned} \tag{58}$$

for $M^n_{i_0}$ in (57). Now, observe that for all $j \in \mathcal{E}(i_0)$ and $s \leq t$, we have that $Q^n_k(s) = 0$ for all $k \in \mathcal{E}(i_0) \cap \mathcal{E}(j)$. Therefore, for all such n, j and s we have that

$$\mathbb{1}_{\mathcal{P}_j(i_0)}\big(Q^n(s)\big) = \prod_{\substack{\ell\in[\![1,|\mathcal{S}|]\!]:\\ i_\ell\in\Phi_j(i_0)}} \mathbb{1}_{\{0\}}\big(Q_{i_\ell}(s)\big) = \mathbb{1}_{\mathcal{P}^{\mathcal{S}}_j(i_0)}\big(S^n(s)\big).$$

Thus, we obtain from (58) that, for all $n \geq n_0$ and $t < \delta$,

$$\begin{aligned}\bar{Q}^n_{i_0}(t) &= \bar{Q}^n_{i_0}(0) + \int_0^t \mathbb{1}_{\mathcal{O}_{i_0}}\big(Q(s-)\big)\,dN^n_{i_0}(s)\\ &\quad - \sum_{j\in\mathcal{E}(i_0)} \lambda_j \nu^n\big([0,t]\times\mathcal{P}^{\mathcal{S}}_j(i_0)\big) - M^n_{i_0}(t)/n,\end{aligned} \tag{59}$$

for ν^n in (53). Now, the equality in distribution (28) and the ergodicity of the CTMC χ imply [similar to (55)] that

$$\nu^n\big([0,T]\times\mathcal{Z}\big) \Rightarrow \pi(\mathcal{Z})T \qquad \text{as } n\to\infty, \qquad \mathcal{Z}\subseteq\mathbb{G}^{\mathcal{S}}, \tag{60}$$

for π in (29). By Lemma 2 and the FWLLN for the Poisson process, the second argument to the right of the equality in (59) converges weakly to $\lambda_{i_0}e$ in $\mathbb{D}[0,\delta)$, where e denotes the identity function $e(t) = t$. Since $\mathbb{1}_{\mathcal{N}_i}(Q(s-))$ is identically equal to 1 for all n large enough, again by Lemma 2, and $M^n_{i_0}/n \Rightarrow 0e$ in $\mathbb{D}[0,\delta)$ as $n \to \infty$, by virtue of Doob's martingale inequality, the limit (32) follows from (60) and Lemma 7.3 in [27] (a simple extension to the continuous mapping theorem).

Next, recall (22) and (24), and consider $\{\bar{S}^n : n \geq 1\}$. By Assumption 2, each element $\chi^n, n \geq 1$, is an ergodic CTMC, and thus $\{\bar{\chi}^n : n \geq 1\}$ is a $\mathbb{C}$-tight sequence of ergodic CTMC's. It follows from Proposition 9.9 in [28] that there exists an a.s.-finite time $\mathcal{T}$, such that $\bar{\chi}^n \Rightarrow 0e$ in $\mathbb{D}^{|\mathcal{S}|}(\mathcal{T},\infty)$ as $n \to \infty$. In particular, with d_P denoting the Prohorov metric [4, 13] (here, denoted in terms of the distance between the random elements instead of the distance between their corresponding probability measures in the underlying probability space), it holds that, for any $t > \mathcal{T}$ and for any $\epsilon > 0$,

$$d_P\big(\bar{\chi}^n(t), 0\big) < \epsilon/2 \qquad \text{for all } n \text{ large enough.} \tag{61}$$

Since any limit point of the tight sequence $\{\bar{\chi}^n : n \geq 1\}$ is continuous, if it is ever larger than 0, then it must be strictly positive over an interval.

Fix $\epsilon > 0$ and let $\|\cdot\|_{\text{tv}}$ denote the total-variation norm; see, for example, [13]. It follows from (23) that for any $t > 0$, there exists n_1, such that $\|\chi^n(t) - \chi(\infty)\|_{\text{tv}} < \epsilon/2$ for all $n > n_1$. Hence, by the triangle inequality, for any $s < \mathcal{T}$ and $t > \mathcal{T}$ there exists n_2, such that for all $n > n_2$,

$$\|\chi^n(s) - \chi^n(t)\|_{\text{tv}} < \epsilon/2 \quad \text{or, equivalently,} \quad \|\bar{\chi}^n(s) - \bar{\chi}^n(t)\|_{\text{tv}} < \epsilon/2. \tag{62}$$

Now, since the Prohorov metric and the total-variation metric (induced by the total variation norm) are equivalent in discrete state spaces, (62) implies that $d_P(\bar{\chi}^n(s), \bar{\chi}^n(t)) < \epsilon/2$ for all $n > n_2$. Together with (61) and the triangle inequality, we obtain

$$d_P(\bar{\chi}^n(s), 0) < \epsilon, \qquad 0 < s < \mathcal{T}.$$

The pointwise convergence of $\bar{\chi}^n$ to 0 implies that no limit point of the $\mathbb{C}$-tight sequence $\{\bar{\chi}^n : n \geq 1\}$ can be strictly positive over an interval, so that $\bar{\chi}^n \Rightarrow 0e$, and in turn, by (28), $\bar{S}^n \Rightarrow 0e$ as $n \to \infty$ in $\mathbb{D}^{|\mathcal{S}|}[0, \delta)$.

Increasing the interval of convergence. The representation of $\bar{Q}^n_{i_0}$ in (58) holds as long as $\bar{Q}^n_{i_0} > 0$, and in particular, over $[0, \rho^n)$. Since $P(\bar{Q}^n_{i_0}(\delta-) > 0) \to 1$ as $n \to \infty$ we conclude from the $\mathbb{C}$-tightness of $\bar{Q}^n$ over $[0, \delta]$ that the convergence of $\bar{Q}^n$ in fact holds over $[0, \delta]$. We can then treat $\bar{Q}(\delta)$ as an initial condition, and apply Lemma 2 for this new initial condition to conclude that there exists a $\delta_2 > \delta$ such that $\bar{Q}^n_{i_0} > 0$ w.p.1 over $[0, \delta_2)$ for all n large enough. Hence, the FWLLN holds over $[0, \delta_2)$ as well. Repeating the same arguments inductively, we can continue increasing the interval of convergence as long as $\bar{Q}_{i_0}$ is guaranteed to be strictly positive, where in the induction step k we take $\bar{Q}^n_{i_0}(\delta_k)$ as an initial condition and apply Lemma 2 to find a $\delta_{k+1} > \delta_k$ such that $\bar{Q}^n \Rightarrow \bar{Q}$ in $\mathbb{D}^{|\mathcal{V}|}[0, \delta_{k+1})$. If (30) does not hold, then it follows from (32) that $\bar{Q}_{i_0}$ is nondecreasing. Necessarily, $\rho^n \Rightarrow \infty$ as $n \to \infty$, and the convergence of $\bar{Q}^n$ to $\bar{Q}$ can be extended indefinitely. On the other hand, if (30) does hold, then the fluid limit $\bar{Q}_{i_0}$ is strictly decreasing. The first passage time ρ^n in (25) is a continuous mapping by, for example, Theorem 13.6.4 in [34], so that $\rho^n \Rightarrow \rho$ in $\mathbb{R}$ as $n \to \infty$, for ρ in (31), and the convergence of $\bar{Q}^n$ can be extended from $[0, \delta)$ to $[0, \rho)$. □

8. Uniform matching policy.

Our main result shows that, for G in $\mathscr{G}_7^c$, there always exists a "bad" choice of priority matching policy, leading to a stability region that is strictly smaller than $\text{NCOND}_{\text{C}}(G)$. In this section, we show that the methods developed to prove this result can be applied to other policies. In particular, we now consider a matching policy in which the choice of which class to match with an arriving class-i is drawn uniformly at random from the *available* classes in $\mathcal{E}(i)$ at the arrival epoch. We refer to this policy as *uniform*, and denote it by $\Phi = \text{U}$. To formally describe U, let t be an arrival epoch of a class-i item, and consider the set

$$\mathcal{U}_i(t) = \{j \in \mathcal{E}(i) : Q_j(t) > 0\}. \tag{63}$$

(i) If $\mathcal{U}_i(t) = \varnothing$, then no matching occurs, and the arriving item is placed in the buffer.

(ii) If $\mathcal{U}_i(t) \neq \varnothing$, then the matching class is chosen uniformly at random, namely, the arriving class-i item will be matched with a class-j item with probability $1/|\mathcal{U}_i(t)|$, for each $j \in \mathcal{U}_i(t)$.

The main question Theorem 3 answered was whether the choice of the (admissible) matching policy affects the stability region of matching queues having non-bipartite and nonseparable graphs. The next result demonstrates that nonmaximality of such graphs is not restricted to strict priority policies.

PROPOSITION 7. *The only graphs in $\mathscr{G}_7^c$ for which* $\text{NCOND}_\text{C}(G)$ *is non-empty and the policy* U *is maximal are separable of order* 3 *or more.*

The proof of Proposition 7 follows the same steps of the proof of Theorem 3. We therefore specify only the arguments in the proof that need to be modified. The main step that needs to be modified is the proof of Lemma 4, which needs to be adapted to the policy U. To this end, as in the proof of Lemma 4, we must couple two systems with initial buffer contents x and $y \in \mathbb{G}$, such that both systems are fed by the same Poisson processes. (We henceforth refer to those systems as "system x" and "system y.")

Let $\{T_n : n \geq 0\}$ and $\{C_n : n \geq 0\}$ denote, respectively, the sequences of arrival times and of the classes of the entering items, in arrival order. Consequently, for any $n \geq 0$, the nth arriving item makes a uniform choice from the set $\mathcal{U}^x_{C_n}(T_n)$ in system x and the set $\mathcal{U}^y_{C_n}(T_n)$ in system y, where $\mathcal{U}^x_{C_n}(T_n)$ and $\mathcal{U}^y_{C_n}(T_n)$ denote the sets defined in (63) for systems x and y, respectively. In the present case, the difficulty stems from the fact that the sets $\mathcal{U}^x_{C_n}(T_n)$ and $\mathcal{U}^y_{C_n}(T_n)$ *a priori* differ, even though both systems are constructed with the same input $\{(T_n, C_n) : n \geq 0\}$. Nevertheless, we can couple these two systems as follows.

Let $\{K^x_n, n \geq 0\}$ and $\{K^y_n, n \geq 0\}$ denote two independent sequences of independent random variables, where for all $n \geq 0$, K^x_n and K^y_n follow the discrete uniform distribution on $\mathcal{U}^x_{C_n}(T_n)$ and $\mathcal{U}^y_{C_n}(T_n)$, respectively. Set $K^x_n = 0$ (resp., $K^y_n = 0$) if $\mathcal{U}^x_{C_n}(T_n) = \varnothing$ [resp., $\mathcal{U}^y_{C_n}(T_n) = \varnothing$], and for all $n \geq 0$, denote the event

$$\mathscr{U}_n = \{(K^x_n, K^y_n) \in (\mathcal{U}^x_{C_n}(T_n) \cap \mathcal{U}^y_{C_n}(T_n))^2\}.$$

Finally, set

$$\tilde{K}^y_n = K^x_n \mathbb{1}_{\mathscr{U}_n} + K^y_n \mathbb{1}_{\mathscr{U}^c_n}. \tag{64}$$

Then the random variables $\tilde{K}^y_n, n \geq 0$, are independent, and for all $n \geq 0$, $\tilde{K}^y_n$ is uniformly distributed on $\mathcal{U}^y_{C_n}(T_n)$. To see this, observe that, if $\mathcal{U}^y_{C_n}(T_n) \neq \varnothing$, then:

- for all $k \in \mathcal{U}^y_{C_n}(T_n) \setminus \mathcal{U}^x_{C_n}(T_n)$,

$$P(\tilde{K}^y_n = k) = P(\{K^y_n = k\} \cap \mathscr{U}^c_n) = P(K^y_n = k) = \frac{1}{|\mathcal{U}^y_{C_n}(T_n)|};$$

- for all $k \in \mathcal{U}^y_{C_n}(T_n) \cap \mathcal{U}^x_{C_n}(T_n)$,

$$\begin{aligned} P(\tilde{K}^y_n = k) &= P(\{\tilde{K}^y_n = k\} \cap \mathscr{U}_n) + P(\{\tilde{K}^y_n = k\} \cap \mathscr{U}^c_n) \\ &= P(\{K^x_n = k\} \cap \{K^y_n \in \mathcal{U}^x_{C_n}(T_n) \cap \mathcal{U}^y_{C_n}(T_n)\}) \\ &\quad + P(\{K^y_n = k\} \cap \{K^x_n \in \mathcal{U}^x_{C_n}(T_n) \setminus \mathcal{U}^y_{C_n}(T_n)\}) \\ &= \frac{1}{|\mathcal{U}^x_{C_n}(T_n)|} \frac{|\mathcal{U}^x_{C_n}(T_n) \cap \mathcal{U}^y_{C_n}(T_n)|}{|\mathcal{U}^y_{C_n}(T_n)|} \\ &\quad + \frac{1}{|\mathcal{U}^y_{C_n}(T_n)|}\left(1 - \frac{|\mathcal{U}^x_{C_n}(T_n) \cap \mathcal{U}^y_{C_n}(T_n)|}{|\mathcal{U}^x_{C_n}(T_n)|}\right) \\ &= \frac{1}{|\mathcal{U}^y_{C_n}(T_n)|}. \end{aligned}$$

We have the following analogue to Lemma 4.

LEMMA 7. *Fix G and the matching policy $\Phi = \textsc{u}$. Let x and y be two elements in the state space $\mathbb{G}$ of Q, and denote by Q^x and Q^y the buffer content processes of the two models having initial values x and y, respectively, and respectively fed by the inputs $\{(T_n, C_n, K^x_n) : n \geq 0\}$ and $\{(T_n, C_n, \tilde{K}^y_n) : n \geq 0\}$. Then, for all $t \geq 0$,*

$$\|Q^x(t) - Q^y(t)\| \leq \|x - y\|.$$

PROOF. We reason by induction, as in the proof of Lemma 4, keeping the notation therein. Suppose that we have at time T_n, $\|Q^x(n) - Q^y(n)\| \leq \|x - y\|$. Then we are in the following alternative:

1. On $\mathscr{U}_n$, we have by construction $K^x_n = \tilde{K}^y_n$, so the newly arrived item of class C_n is matched with an item of the same class K^x_n in both systems. We are in the case 1 of the proof of Lemma 4.
2. On $\mathscr{U}^c_n$, we have three possible cases:

- If both $\mathcal{U}^x_{C_n}(T_n)$ and $\mathcal{U}^y_{C_n}(T_n)$ are nonempty, then $K^x_n = k^x$ for some $k^x \in \mathcal{U}^x_{C_n}(T_n)$ and $K^y_n = k^y$ for some $k^y \in \mathcal{U}^y_{C_n}(T_n)$. However, it must hold that $k^x \notin \mathcal{U}^y_{C_n}(T_n)$ or that $k^y \notin \mathcal{U}^x_{C_n}(T_n)$ (or both), otherwise we would be in $\mathscr{U}_n$. In the first case (the other one is symmetric), we have that $Q^x_{k^x}(n) > 0$ and $Q^y_{k^x}(n) = 0$, and we are in case 2 of the proof of Lemma 4.

- If exactly one of the two sets is empty, say $\mathcal{U}^y_{C_n}(T_n)$ is empty and $\mathcal{U}^x_{C_n}(T_n)$ is not (the other way around is symmetric), then the incoming item at T_n is matched in the system initiated by x and not in the system initiated by y, so we are in case 3 of the proof of Lemma 4.
- If the two sets are empty, then the incoming item at T_n is matched in none of the two systems, so we are in case 4 of Lemma 4. □

PROOF OF PROPOSITION 7. Fix a connected graph $G = (\mathcal{V}, \mathcal{E}) \in \mathscr{G}_7^c$, and $\Phi = \textsc{u}$. Fix a node $i_0 \in \mathcal{V}$ and denote again $\mathcal{S} := \mathcal{V} \setminus (\{i_0\} \cup \mathcal{E}(i_0)) = \{i_1, \dots, i_{|\mathcal{S}|}\}$. For any $j \in \mathcal{V}$, denote

$$\mathcal{S}(j) = \{\ell \in [\![1, |\mathcal{S}|]\!] : i_\ell \in \mathcal{E}(j)\},$$

and for any $r \in [\![0, |\mathcal{S}(j)|]\!]$, let

$$\mathcal{V}^{\mathcal{S}}_{j,r} = \{x \in \mathbb{G}^{\mathcal{S}} : \mathrm{Card}\{\ell \in \mathcal{S}(j) : x_\ell > 0\} = r\}, \tag{65}$$

where for notational convenience, we let Card A denote the cardinality of the set A.

Suppose that Assumption 2 holds for the sequence of processes $\{\chi^n : n \geq 1\}$ corresponding to the marginal process χ of infinitesimal generator

$$\begin{cases} \mathscr{A}^{\mathcal{S},\textsc{u}}(x, x + \mathbf{e}_\ell) = \lambda_{i_\ell} \mathbb{1}_{\mathcal{O}_{i_\ell}}(x) & \ell \in [\![1, |\mathcal{S}|]\!]; \\ \mathscr{A}^{\mathcal{S},\textsc{u}}(x, x - \mathbf{e}_\ell) = \mathbb{1}_{\mathcal{N}^{\mathcal{S}}_\ell}(x) \displaystyle\sum_{j \in \mathcal{E}(i_\ell)} \sum_{r=0}^{\mathcal{S}(j)} \frac{\lambda_j}{r} \mathbb{1}_{\mathcal{V}^{\mathcal{S}}_{j,r}}(x) & \ell \in [\![1, |\mathcal{S}|]\!]. \end{cases} \tag{66}$$

As is easily seen from the definition of the policy $\textsc{u}$, and similar to (28), the process χ^n coincides in distribution with the restriction S^n of the process Q^n to its coordinates in $\mathcal{S}$, as long as $Q^n_{i_0}$ remains strictly positive. Provided that at time t, $Q^n_{i_0}(t) > 0$, $S^n(t) \in \mathcal{V}^{\mathcal{S}}_{j,r}$, and an item of a class $j \in \mathcal{E}(i_0)$ enters the system, the match of the incoming item is drawn uniformly among all r classes of $\mathcal{S}(j)$ having items in line, and the class i_0. Consequently, under assumption 1, an analogous result to Theorem 4 holds, with the following drift for the fluid limit of the i_0-coordinate:

$$\lambda_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \sum_{r=0}^{\mathcal{S}(j)} \frac{\lambda_j}{r+1} \pi(\mathcal{V}^{\mathcal{S}}_{j,r}). \tag{67}$$

Now, as a consequence of Lemma 7, Lemma 5 still holds true for the matching queues $(G, \lambda, \textsc{u})_{\textsc{c}}$ and $(\tilde{G}, \lambda, \textsc{u})_{\textsc{c}}$, where the disconnected graph $\tilde{G}$ is constructed from G and any induced subgraph $\breve{G}$, as in Figure 7. From Lemma 3, G induces a pendant graph or an odd cycle of size 5, and we let $\breve{G} = (\breve{\mathcal{V}}, \breve{\mathcal{E}})$ be that induced subgraph. In view of (67), provided that we exhibit an arrival-rate vector $\breve{\lambda} \in (\mathbb{R}_{++})^{|\breve{\mathcal{V}}|}$ such that

$$\beta := \breve{\lambda}_{i_0} - \sum_{j \in \mathcal{E}(i_0)} \sum_{r=0}^{\mathcal{S}(j)} \frac{\breve{\lambda}_j}{r+1} \pi(\mathcal{V}^{\mathcal{S}}_{j,r}) > 0,$$

the matching queue $(\breve{G}, \breve{\lambda}, \text{U})_{\text{C}}$ is unstable, and the proof follows the same arguments as the proof of Theorem 3. Thus, it remains to prove the existence of an unstable matching queue $(\breve{G}, \breve{\lambda}, \text{U})_{\text{C}}$ for $\breve{G}$ the pendant graph or the odd cycle. This is done as follows.

Pendant graph. Set $i_0 = 4$. For $\Phi = \text{U}$, from (66) the generator of the marginal process χ is the same as (20), replacing the arrival rate $\breve{\lambda}_3 := \lambda_3$ to node 3 by $\breve{\lambda}_3/2$. (We add the "breve" to the notation of the arrival rates since we are now considering the pendant graph as the induced graph $\breve{G}$ in G.) Then, similar to (35), we obtain that

$$\alpha := \pi\big((0,0)\big) = \frac{(\breve{\lambda}_3/2)^2 - (\breve{\lambda}_1 - \breve{\lambda}_2)^2}{(\breve{\lambda}_3/2)(\breve{\lambda}_3 + \breve{\lambda}_1 + \breve{\lambda}_2)}.$$

The drift in (67) reads

$$\begin{aligned} &\breve{\lambda}_4 - \breve{\lambda}_3\pi\big((0,0)\big) - \frac{\breve{\lambda}_3}{2}\pi\big(\{0\}\times\mathbb{Z}_{++}\big) - \frac{\breve{\lambda}_3}{2}\pi\big(\mathbb{Z}_{++}\times\{0\}\big) \\ &\quad = \breve{\lambda}_4 - \breve{\lambda}_3\alpha - \frac{\breve{\lambda}_3}{2}(1-\alpha) = \breve{\lambda}_4 - \frac{\breve{\lambda}_3}{2}(1+\alpha). \end{aligned} \tag{68}$$

Fix $\epsilon \in (0, 7/15]$ and set (see Figure 8)

$$\begin{cases} \breve{\lambda}_1 = \breve{\lambda}_2 = \epsilon; \\ \breve{\lambda}_3 = \dfrac{1}{2} - \epsilon/2; \\ \breve{\lambda}_4 = \dfrac{1}{2} - 3\epsilon/4. \end{cases}$$

It can be easily checked that $\breve{\lambda} \in \text{NCOND}(\breve{G})$. However, the drift in (68) becomes

$$\frac{1}{2} - \frac{3\varepsilon}{4} - \frac{1}{2}\left(\frac{1}{2} - \frac{\varepsilon}{2}\right)\left(1 + \frac{1/2 - \epsilon/2}{1/2 - \epsilon/2 + 4\epsilon}\right) = \frac{\epsilon}{4(1+7\epsilon)}(7 - 15\epsilon) > 0.$$

5-cycle. Set $i_0 = 4$. For $\Phi = \text{U}$, from (66) the generator of the marginal process is the same as (40), replacing the arrival rates to nodes 3 and 4, $\breve{\lambda}_i = \lambda_i$, by $\breve{\lambda}_i/2$,

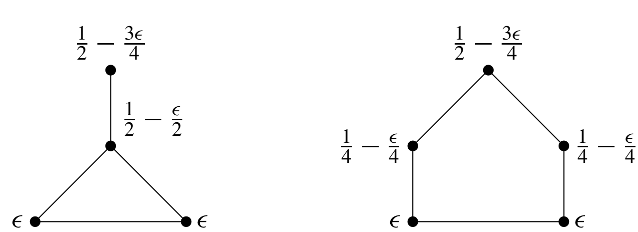

FIG. 8. *Unstable uniform matching queues: left, on the pendant graph and right, on the 5-cycle.*

$i = 3, 4$. The drift in (67) reads

$$\breve{\lambda}_5 - \breve{\lambda}_3 \tilde{\pi}(\{0\} \times \mathbb{Z}_+) - \breve{\lambda}_4 \tilde{\pi}(\mathbb{Z}_+ \times \{0\}) \tag{69}$$
$$- \frac{\breve{\lambda}_3}{2} \tilde{\pi}(\mathbb{Z}_{++} \times \{0\}) - \frac{\breve{\lambda}_4}{2} \tilde{\pi}(\{0\} \times \mathbb{Z}_{++}),$$

where $\tilde{\pi}$ is the stationary distribution of the fast process, obtained similarly to $\tilde{\pi}$ in (4), replacing the intensities at nodes 3 and 4, $\breve{\lambda}_i = \lambda_i$, by $\breve{\lambda}_i/2$, $i = 3, 4$.

Now, if we fix $\epsilon \in (0, 7/23]$ and set (see Figure 8)

$$\begin{cases} \breve{\lambda}_1 = \breve{\lambda}_2 = \epsilon; \\ \breve{\lambda}_3 = \breve{\lambda}_4 = \dfrac{1}{4} - \epsilon/4; \\ \breve{\lambda}_5 = \dfrac{1}{2} - 3\epsilon/4. \end{cases}$$

Again, we can easily check that $\breve{\lambda} \in \text{NCOND}(\breve{G})$, and the drift in (69) equals

$$\frac{1}{2} - \frac{3\varepsilon}{4} - \left(\frac{1}{2} - \frac{\varepsilon}{2}\right)\frac{1+7\varepsilon}{1+15\varepsilon} - \left(\frac{1}{4} - \frac{\varepsilon}{4}\right)\frac{8\varepsilon}{1+15\varepsilon}$$
$$= \frac{\epsilon}{4(1+15\epsilon)}(7 - 23\epsilon) > 0. \qquad \square$$

9. Summary and future research. In this paper, we proved that matching queues on graphs in $\mathscr{G}_7^c$ satisfying NCOND need not be stable. Our proof employed a fluid-limit whose characterization builds on estimating the stationary distribution of a related marginal process.

There are many directions for future research. We specify four, which we are currently investigating.

Generalizing the result. It follows from Lemma 3 that any non-bipartite and nonseparable graph induces an odd cycle of size 5 or more, or the pendant graph. For both the pendant and the 5-cycle graphs, we have shown that NCOND is not a sufficient condition for stability (Propositions 3 and 6). If a similar instability result could be shown for any odd-cycle, then the following conjecture would be proved via an application of the arguments in the proof of Theorem 3.

CONJECTURE 1. *The* only *connected and non-bipartite graphs G for which the matching queue* $(G, \lambda, \Phi)_C$ *is stable for any admissible matching policy* Φ *and any* $\lambda \in \text{NCOND}_C(G)$, *are the separable graphs of order* 3 *or more.*

A direct demonstration of instability of a matching queue on a p-cycle (where p is odd) requires computing the stationary distribution of the associated $(p-3)$-dimensional marginal process χ. Unfortunately, if $p \geq 7$, the marginal process is not reversible, so that obtaining closed-form expressions for the stationary distributions of all possible $(p-3)$-dimensional CTMCs seems prohibitively hard. Nevertheless, one might be able to appropriately bound these stationary distributions and prove a result analogous to Propositions 3 and 6.

Identifying bottlenecks. The fluid limit may be used to construct a procedure determining the "bottlenecks" (namely, unstable) nodes of general unstable matching queues. Moreover, when the stationary distribution of the fast-time-scale CTMC can be computed explicitly, the fluid limit provides the exact rate of increase of the queues corresponding to the unstable nodes.

Matching models on hypergraphs. In our model, items depart the system by pairs. However, in many applications (e.g., manufacturing and assemble-to-order systems) matchings can occur in groups that are larger than 2 (as, e.g., in [17]). Thus, it remains to establish an analogue to NCOND when the compatibility between items cannot be represented by a graph, but more generally, by an hypergraph.

APPENDIX: MATCHING ALGORITHMS ON RANDOM GRAPHS

In graph theory, a *matching on a graph* $\mathbf{G}$ is a subgraph $\check{\mathbf{G}}$ of $\mathbf{G}$ in which each node has exactly one neighbor. The matching is said to be *perfect* if $\mathbf{G}$ and $\check{\mathbf{G}}$ have the same set of nodes. It is a consequence of Tutte's theorem (a generalization of Hall's marriage theorem to arbitrary graphs), that a necessary condition for the existence of a perfect matching on $\mathbf{G}$ is given by

$$|\mathbf{I}| \leq |\mathcal{E}_{\mathbf{G}}(\mathbf{I})| \qquad \text{for any independent set } \mathbf{I} \in \mathbb{I}(\mathbf{G}), \tag{70}$$

where $\mathcal{E}_{\mathbf{G}}(\mathbf{I})$ denotes the set of neighbors of the elements of $\mathbf{I}$ in $\mathbf{G}$. A *matching algorithm* is a procedure for constructing a matching.

It is a well-known fact that, even when a perfect matching exists on $\mathbf{G}$, an online matching algorithm under which, at each step a node is chosen uniformly at random among all unmatched nodes and its match is chosen uniformly at random among all its unmatched neighbors, fails in general to lead to a perfect matching.

A matching on a graph is not to be confused with the stochastic matching of items discussed thus far. However, clear connections can be drawn between the two problems, as we briefly illustrate by a simple example.

Consider a *random* graph in which the nodes are of p different types, such that the types of the various nodes are random and i.i.d., having a common distribution μ on the set of types $[\![1, p]\!]$. Nodes of the same type are not neighbors of each other, and have the same neighbors. Specifically, we fix a given auxiliary (simple) graph $G = (\mathcal{V}, \mathcal{E})$ of size p, which we call *template graph*, and the set of types of the nodes of $\mathbf{G}$ is identified with $\mathcal{V}$. The edges of $\mathbf{G}$ are fully determined by the types of the nodes according to the following rule: two nodes u and v of **G**, of respective types i and j, are neighbors in **G** if and only if $i - j$ in G.

We aim to construct a matching on the resulting graph **G**. Our approach is to construct the random graph **G** *together* with the matching on **G** sequentially. To this end, we define p independent Poisson processes $N_1, \dots, N_p$ with respective intensities $\mu(i), i \in [\![1, p]\!]$, and let $T_1, T_2, \dots$ the points of the superposition N of the p processes. We also fix a matching queue $(G, \mu, \Phi)_{\text{C}}$, where the arrival-rate vector is denoted $\mu = (\mu(1), \dots, \mu(p))$. We proceed by induction, at each point of N. For all $n \geq 1$:

(i) Let i be the element of $[\![1, p]\!]$ such that T_n is a point of N_i. At T_n, create a node u of **G**, and assign to u the type i. Then create an edge in **G** between u and all the previously created nodes of all types j such that $i - j$ in the template graph G.

(ii) If the set of unmatched neighbors of u in **G** is nonempty, apply Φ to select a unique node v in the latter set, exactly as we choose a match for an item of class i in the matching queue $(G, \mu, \Phi)_{\text{C}}$. We call v the *match* of u, and say that both u and v are *matched* nodes. If no neighbor of u is unmatched, we set the status of u to be *unmatched*.

At any time T_n, the graph **G** has exactly n nodes, some of which are matched and the other are unmatched. We let $\mathscr{M}(\mathbf{G})$ and by $\bar{\mathscr{M}}(\mathbf{G})$ denote the sets of matched and unmatched nodes of **G**, respectively. The *matching* on **G** is the set of nodes $\mathscr{M}(\mathbf{G})$, together with the edges between matched couples. An example of such a construction when **G** is the pendant graph is given in Figure 9.

Almost surely for a large enough n, the resulting graph **G** at T_n is p-partite (all types of nodes are represented in **G**, and there is no edge between any two nodes of the same type). Moreover, as a straightforward application of the Strong Law of Large Numbers (SLLN), the proportions of nodes of the various types tend to the measure μ as n increases to infinity.

The construction just described is related to the procedure of uniform random pairing; see [35]. This latter procedure leads to the so-called configuration model introduced by Bollobas [5] (see also [32]), generating a realization of a random graph where the *degree*, that is, the number of neighbors of each node, is fixed beforehand. We also refer to [14] and the references therein, for general results concerning matching on random graphs.

It is then easy to couple the construction described above with the matching queue (G, μ, Φ): if both are constructed with the same Poisson processes, then:

(i) the creation of a node of type i in the random graph **G** corresponds to the arrival of a class-i item in the matching queue;

(ii) a matching between two nodes of respective types i and j in **G** occurs if and only if in the matching queue, at the same instant, two items of respective classes i and j are matched, and depart the system.

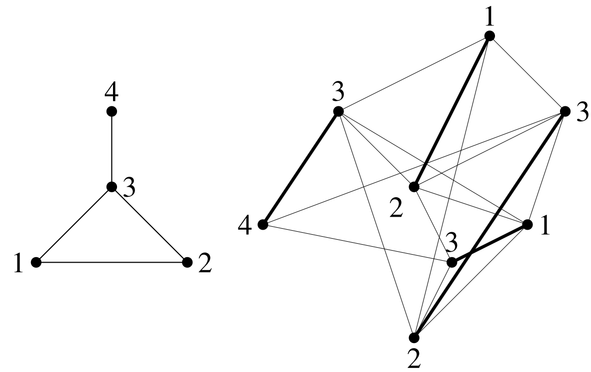

FIG. 9. *Construction of a matching on a* 4*-partite graph having the pendant graph as template.*

Consequently, at any time the list of classes of queued items in the matching queue coincides with the list of types of the unmatched nodes in $\mathbf{G}$.

For all $j \in [\![1, p]\!]$ and $t \geq 0$, let $\mathbf{Q}_j(t)$ denote the number of unmatched nodes of type j in $\mathbf{G}$ at time t, and let $\mathbf{Q} = (\mathbf{Q}_1, \ldots, \mathbf{Q}_p)$. As before, let Q denote the queue process of the matching queue in (3). It follows that, if $Q(0) = \mathbf{Q}(0)$, then

$$\mathbf{Q}_j(t) = Q_j(t), \qquad j \in [\![1, p]\!], t \geq 0.$$

Therefore, Theorems 2 and 7 imply the following.

COROLLARY 4. *Assume that the template graph G is connected and non-bipartite*:

(i) *If $\Phi = \text{ML}$, then for all $\mu \in \text{NCOND}_{\text{C}}(G)$, the Markov process $\mathbf{Q}$ is positive recurrent*.

(ii) *If $\Phi = \text{U}$ and $G \in \mathscr{G}_7^c$, then the Markov process $\mathbf{Q}$ is positive recurrent for all $\mu \in \text{NCOND}_{\text{C}}(G)$ if and only if G is separable*.

Let us now observe that just after time T_n, the size of the matching on $\mathbf{G}$ is given by

$$|\mathscr{M}(\mathbf{G})| = n - |\bar{\mathscr{M}}(\mathbf{G})| = n - \sum_{i=1}^{p} \mathbf{Q}_i(T_n). \tag{71}$$

In view of (71), Corollary 4 suggests that the separable graphs are the only template graphs in $\mathscr{G}_7^c$ guaranteeing that under NCOND and the uniform policy, $|\bar{\mathscr{M}}(\mathbf{G})|$ becomes negligible with respect to n as n increases. Therefore, the size of the matching and the size of the graph tend to coincide.

Now, for any time T_n and for any set of nodes $A \subset \mathcal{V}$, let $\mathbf{X}_A(T_n)$ denote the set of nodes of $\mathbf{G}$ at T_n having types in A. We have

$$\mathbf{X}_A(T_n) = \sum_{j \in A} N_j\big((0, T_n]\big), \tag{72}$$

where $N_j((0, T_n])$ is the number of points of N_j up to time T_n. By construction, for any two nodes u and v in $\mathbf{G}$ of respective types i and j, $u \not\sim v$ in $\mathbf{G}$ entails $i \not\sim j$ in G. Thus, for any independent set $\mathbf{I}$ of $\mathbf{G}$ at T_n, there exists a unique independent set $\mathcal{I}$ of G, such that $\mathbf{I} \subset \mathbf{X}_{\mathcal{I}}(T_n)$, and all types in $\mathcal{I}$ are represented in $\mathbf{I}$. Moreover, the set of all the neighbors in $\mathbf{G}$ of the elements of $\mathbf{I}$ is exactly the set of all nodes of $\mathbf{G}$ having types that are neighbors in G of the types belonging to $\mathcal{I}$, that is,

$$\mathcal{E}_{\mathbf{G}}(\mathbf{I}) = \mathbf{X}_{\mathcal{E}(\mathcal{I})}(T_n). \tag{73}$$

It follows from (70), (72) and (73) that

$$\sum_{i \in \mathcal{I}} N_i(T_n) \le \sum_{j \in \mathcal{E}(\mathcal{I})} N_j(T_n) \qquad \text{for all independents sets } \mathcal{I} \in \mathbb{I}(G) \tag{74}$$

is a necessary condition for the existence of a perfect matching on $\mathbf{G}$ at time T_n. Dividing both sides of the equality in (74) by n and taking n to infinity, the SLLN implies that $\mu(\mathcal{I}) \le \mu(\mathcal{E}(\mathcal{I}))$, for all independents sets $\mathcal{I} \in \mathbb{I}(G)$.

We conclude that a necessary condition for the existence of a perfect matching on $\mathbf{G}$ in the large graph limit is closely related to NCOND(G). (Specifically, the strict inequality in NCOND is replaced by a weak inequality.) Thus, Corollary 4 is reminiscent of the aforementioned result concerning the construction of matchings using uniform on-line algorithms: aside from the case of separable graphs (for which all matching policies are equivalent in terms of types, see Section 3.2), under a condition on the connectivity of $\mathbf{G}$ that is closely related to (70), a matching policy that is uniform in terms of types of nodes also may fail in general to construct a perfect matching on $\mathbf{G}$.

Acknowledgments. The first author thanks the Industrial Engineering and Management Sciences (IEMS) Department and the McCormick School of Engineering at Northwestern University for their hospitality and support.

REFERENCES

[1] ADAN, I. and WEISS, G. (2012). Exact FCFS matching rates for two infinite multitype sequences. *Oper. Res.* **60** 475–489. MR2935072

[2] ADAN, I. and WEISS, G. (2014). A skill based parallel service system under FCFS-ALIS—Steady state, overloads, and abandonments. *Stoch. Syst.* **4** 250–299. MR3353219

[3] BACCELLI, F. and BRÉMAUD, P. (2003). *Elements of Queueing Theory: Palm Martingale Calculus and Stochastic Recurrences*, 2nd ed. *Applications of Mathematics* (*New York*) **26**. Springer, Berlin. MR1957884

[4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[5] BOLLOBÁS, B. (2001). *Random Graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge Univ. Press, Cambridge. MR1864966

[6] BRAMSON, M. (1994). Instability of FIFO queueing networks. *Ann. Appl. Probab.* **4** 414–431. MR1272733

[7] BRAMSON, M. (2008). *Stability of Queueing Networks*. *Lecture Notes in Math.* **1950**. Springer, Berlin. MR2445100

[8] BUŠIĆ, A., GUPTA, V. and MAIRESSE, J. (2013). Stability of the bipartite matching model. *Adv. in Appl. Probab.* **45** 351–378. MR3102455

[9] CALDENTEY, R., KAPLAN, E. H. and WEISS, G. (2009). FCFS infinite bipartite matching of servers and customers. *Adv. in Appl. Probab.* **41** 695–730. MR2571314

[10] DAI, J. G. (1995). On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Probab.* **5** 49–77. MR1325041

[11] DAI, J. G. (1996). A fluid limit model criterion for instability of multiclass queueing networks. *Ann. Appl. Probab.* **6** 751–757. MR1410113

[12] DAVIS, A. E., MEHROTRA, S., KILAMBI, V., PERRY, O., FRIEDEWALD, J. J. and LADNER, D. P. (2015). Addressing US national geographic disparity in kidney transplantation by creating sharing partnerships. Working paper.

[13] DURRETT, R. (1991). *Probability: Theory and Examples*. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA. MR1068527

[14] GAMARNIK, D. and GOLDBERG, D. A. (2010). Randomized greedy algorithms for independent sets and matchings in regular graphs: Exact results and finite girth corrections. *Combin. Probab. Comput.* **19** 61–85. MR2575098

[15] GAMARNIK, D. and HASENBEIN, J. J. (2005). Instability in stochastic and fluid queueing networks. *Ann. Appl. Probab.* **15** 1652–1690. MR2152240

[16] GURVICH, I., LARIVIERE, M. and MORENO, A. (2016). Operations in the on-demand economy: Staffing services with self-scheduling capacity. Available at SSRN 2336514.

[17] GURVICH, I. and WARD, A. (2014). On the dynamic control of matching queues. *Stoch. Syst.* **4** 479–523. MR3353225

[18] IBRAHIM, R. (2016). Staffing a service system where capacity is random. Working paper.

[19] KURTZ, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis* (*New Brunswick, NJ*, 1991). *Lect. Notes Control Inf. Sci.* **177** 186–209. Springer, Berlin. MR1169928

[20] LOVÁSZ, L., PELIKÁN, J. and VESZTERGOMBI, K. (2003). *Discrete Mathematics: Elementary and Beyond*. Springer, New York. MR1952453

[21] LU, S. H. and KUMAR, P. R. (1991). Distributed scheduling based on due dates and buffer priorities. *IEEE Trans. Automat. Control* **36** 1406–1416.

[22] LUO, J. and ZHANG, J. (2013). Staffing and control of instant messaging contact centers. *Oper. Res.* **61** 328–343. MR3046113

[23] MAIRESSE, J. and MOYAL, P. (2016). Stability of the stochastic matching model. *J. Appl. Probab.* **53** 1064–1077. MR3581242

[24] MEYN, S. P. (1995). Transience of multiclass queueing networks via fluid limit models. *Ann. Appl. Probab.* **5** 946–957. MR1384361

[25] PANG, G., TALREJA, R. and WHITT, W. (2007). Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probab. Surv.* **4** 193–267. MR2368951

[26] PERRY, O. and WHITT, W. (2011). An ODE for an overloaded X model involving a stochastic averaging principle. *Stoch. Syst.* **1** 59–108. MR2948918

[27] PERRY, O. and WHITT, W. (2013). A fluid limit for an overloaded X model via a stochastic averaging principle. *Math. Oper. Res.* **38** 294–349. MR3062009

[28] ROBERT, P. (2003). *Stochastic Networks and Queues*, French ed. *Applications of Mathematics* (*New York*) **52**. Springer, Berlin. MR1996883

[29] RYBKO, A. N. and STOLYAR, A. L. (1992). On the ergodicity of random processes that describe the functioning of open queueing networks. *Problemy Peredachi Informatsii* **28** 3–26. MR1189331

[30] SIDDHARTHA, B., RIQUELME, C. and JOHARI, R. (2015). Pricing in ride-share platforms: A queueing-theoretic approach. Available at SSRN 2568258.

[31] TALREJA, R. and WHITT, W. (2008). Fluid models for overloaded multiclass many-server queueing systems with first-come, first-served routing. *Manage. Sci.* **54** 1513–1527.

[32] VAN DER HOFSTAD, R. (2013). *Random Graphs and Complex Networks*. Lecture notes.

[33] WHITT, W. (1991). The pointwise stationary approximation for $M_t/M_t/s$ queues is asymptotically correct as the rates increase. *Manage*. *Sci*. **37** 207–314.

[34] WHITT, W. (2002). *Stochastic-Process Limits*: *An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer, New York. MR1876437

[35] WORMALD, N. C. (1995). Differential equations for random processes and random graphs. *Ann*. *Appl*. *Probab*. **5** 1217–1235. MR1384372

[36] WU, S., ZHANG, J. and ZHANG, R. Q. Management of a shared spectrum network in wireless communications. Working paper, HKUST.

LABORATOIRE DE MATHÉMATIQUES APPLIQUÉES—EA2222
UNIVERSITÉ DE TECHNOLOGIE DE COMPIÈGNE
DÉPARTEMENT GÉNIE INFORMATIQUE
CENTRE DE RECHERCHES DE ROYALLIEU
60200 COMPIÈGNE
FRANCE
E-MAIL: pascal.moyal@utc.fr

INDUSTRIAL ENGINEERING
AND MANAGEMENT SCIENCES
NORTHWESTERN UNIVERSITY
2145 SHERIDAN RD
EVANSTON, ILLINOIS 60208
USA
E-MAIL: ohad.perry@northwestern.edu