
Online matching on stochastic block model

Maria Cherifa

Université de Paris, MAP5, Paris, France
CREST, ENSAE, Palaiseau, France
Fairplay joint team
maria.cherifa@ensae.fr

Clément Calauzènes

Criteo AI Lab, Paris, France
Fairplay joint team
c.calauzenes@criteo.com

Vianney Perchet

CREST, ENSAE, Palaiseau, France
Criteo AI Lab, Paris, France
Fairplay joint team
vianney.perchet@normalesup.org

Abstract

While online bipartite matching has gained significant attention in recent years, existing analyses in stochastic settings fail to capture the performance of algorithms on heterogeneous graphs, such as those incorporating inter-group affinities or other social network structures. In this work, we address this gap by studying online bipartite matching within the stochastic block model (SBM). A fixed set of offline nodes is matched to a stream of online arrivals, with connections governed probabilistically by latent class memberships. We analyze two natural algorithms: a Myopic policy that greedily matches each arrival to the most compatible class, and the Ex-ante Balance algorithm, which accounts for both compatibility and remaining capacity. For the Myopic algorithm, we prove that the size of the matching converges, with high probability, to the solution of an ordinary differential equation (ODE), for which we provide a tractable approximation along with explicit error bounds. For the Ex-ante Balance algorithm, we demonstrate convergence of the matching size to a differential inclusion and derive an explicit limiting solution. Lastly, we explore the impact of estimating the connection probabilities between classes online, which introduces an exploration-exploitation trade-off.

Introduction

Finding matchings in bipartite graphs is a fundamental problem at the intersection of graph theory [17, 45], network science, and combinatorial optimization [31, 40]. In this context, a bipartite graph $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{E})$ is defined by two disjoint node sets, \mathcal{N} and \mathcal{T} , and a set of edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{T}$ connecting them. These graphs naturally model systems where entities from one group are to be matched with entities from another, such as tasks to agents, products to consumers, or ads to users. A *matching* in such a graph refers to a subset of edges with no shared endpoints — ensuring that each entity is involved in at most one pair. The core challenge lies in computing optimal matchings that respect resource constraints and maximize some utility or coverage. This problem has significant practical relevance, especially in operations research, where it is linked to the classical assignment problem [19].

Recent real-world applications, particularly in online advertising, ride-hailing platforms, and real-time job allocation, have attracted significant interest in the online version of the matching problem [38]. In this setting, nodes in \mathcal{N} (e.g., advertisers or servers) are fixed and known in advance, while nodes in \mathcal{T} (e.g., users or requests) arrive sequentially. When a new node $t \in \mathcal{T}$ arrives, the algorithm

must decide on the spot whether to match it to an available node $n \in \mathcal{N}$, such that $(n, t) \in \mathcal{E}$, with the constraint that each node can be matched at most once. These decisions are irrevocable, making the problem both practically and theoretically challenging. The central goal is to design algorithms that construct a maximum matching—that is, one that covers as many nodes as possible. Typically, such algorithms are analyzed in one of two ways: by approximating the size of the resulting matching, or by evaluating their competitive ratio, defined as the worst-case ratio between the size of the algorithm’s matching and that of an optimal matching computed with full knowledge of the graph in advance.

Online matching has been explored through several theoretical lenses, most notably the adversarial and stochastic frameworks [38]. In the adversarial setting, the graph and arrival sequence are designed to be worst-case, providing robust but conservative performance guarantees. In contrast, stochastic models assume randomness in either the graph structure or the arrival process [9], allowing for more realistic analyses and often stronger guarantees. Within this stochastic line of work, much attention has been given to Erdős–Rényi-type models [35], where edges are included independently with a fixed probability. While such models are analytically tractable and provide insights, they fail to capture complex patterns like community structure, heterogeneity, and group-based interactions observed in real-world systems [20].

To address these limitations, the *stochastic block model* (SBM) has emerged as a powerful alternative [20, 2]. SBM introduces latent classes (or communities) and allows edge probabilities to depend on class membership, thereby capturing structured heterogeneity—such as homophily (the tendency of similar nodes to connect more frequently)[36], core-periphery patterns (where a densely connected "core" group links to many others while "periphery" nodes have fewer connections) [6], or inter-group affinities (specific preferences or tendencies for nodes in one group to connect with nodes in another group)[3]. This makes SBM particularly well-suited for modeling interactions in social networks, recommendation engines, and online marketplaces, where the likelihood of a match depends not just on individual attributes but also on group-level dynamics [26]. In the context of online bipartite matching, this leads naturally to the *bipartite stochastic block model*, where one partition consists of fixed agents and the other of arriving users, both belonging to latent classes that govern connection probabilities [28].

We study online matching in the bipartite stochastic block model, focusing on the sparse regime, where the average degree of each node remains bounded as the system grows. This regime is particularly relevant in practice, as real-world platforms typically feature users or items that interact with only a small subset of the population. Sparsity not only reflects these empirical network structures but also introduces significant analytical challenges. Formally, we consider a bipartite graph $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{E})$, where \mathcal{N} is a fixed set of nodes and \mathcal{T} is a set of nodes that arrive sequentially. Each node $n \in \mathcal{N}$ is independently assigned a class $c(n) \in \mathcal{C}$ according to a distribution $\mu(n)$, and each arriving node $t \in \mathcal{T}$ is independently assigned a class $d(t) \in \mathcal{D}$ with distribution $\nu(t)$. Conditional on these class assignments, an edge between n and t is present independently with probability $p_{n,t}$, which depends on their respective classes. A defining feature of this model, motivated by real-world constraints, is that the set of edges incident to each arriving node is not known in advance. Instead, when a node $t \in \mathcal{T}$ arrives, the algorithm observes only then which edges exist between t and the nodes in \mathcal{N} . This assumption captures a key operational reality in many systems, where information about potential interactions is revealed only upon arrival or activation of a new entity. For example, in online job platforms, the compatibility between a newly posted job (a node in \mathcal{T}) and existing freelancers (nodes in \mathcal{N}) becomes clear only when the job description is published. The system cannot precompute all possible matches due to computational constraints and the dynamic, user-driven nature of postings. In recommendation systems, user preferences are inferred from behavior observed at login, and only then can the platform determine which content is relevant — effectively modeling the appearance of edges at the moment of interaction. Given this online nature of information revelation, an algorithm must operate under uncertainty: it observes each vertex in \mathcal{T} sequentially and must decide, upon arrival, whether and how to match it to an available node in \mathcal{N} , without knowledge of future arrivals or their connections. In this work, we introduce and analyze two natural algorithms designed for this setting:

- **Myopic:** Upon the arrival of a node $t \in \mathcal{T}$, the algorithm chooses a compatible class $c^* \in \mathcal{C}$ and attempts to match t with an available node from this class.

- **Ex-ante Balance:** This algorithm selects the class with the highest probability of a successful match, considering both compatibility and current availability.

We focus on the Myopic and Ex-ante Balance algorithms because they are simple, practical, and reflect decision-making heuristics commonly used in real-world systems. Notably, while the Myopic algorithm has been studied in stochastic models, the Ex-ante Balance algorithm—despite its widespread use—has only been analyzed in adversarial settings. Its theoretical performance in structured stochastic environments, such as the bipartite stochastic block model, remains unexplored. These algorithms are appealing not only for their practical relevance but also for their interpretability and ease of implementation, which is particularly valuable in applications like online marketplaces, content recommendation, or allocation systems. We first analyze them under the assumption that the compatibility probabilities $p_{n,t}$ between user and item classes are known. This setting provides a tractable analytical framework, allowing us to rigorously characterize the fluid-limit behavior of both algorithms. While this fully informed setting offers valuable insights, it often does not reflect the realities faced by many real-world systems. Motivated by these practical constraints, we then turn to the more realistic case where the probabilities $p_{n,t}$ are not known a priori. In many applications—such as recommendation engines or online platforms—the interaction propensities between user and item types must be inferred over time through observed outcomes. This naturally gives rise to a bandit setting, where the algorithm must estimate the unknown affinities $p_{n,t}$ from binary feedback (indicating whether a match succeeded or failed), while simultaneously making irrevocable matching decisions. This introduces an exploration-exploitation trade-off that is absent in the known-parameter regime. To address this challenge, we propose and analyze a bandit version of the Ex-ante Balance algorithm, which learns class affinities dynamically while aiming to preserve strong matching performance.

Our two main contributions, correspond to the two settings considered:

- **When $p_{n,t}$ are known:** We provide a fluid-limit analysis of both Myopic and Ex-ante Balance algorithms in the sparse bipartite stochastic block model. Specifically,
 - We prove that the matching size obtained by the Myopic algorithm is, with high probability, close to the solution of a specific ordinary differential equation. Due to the complexity of solving this equation in closed form, we derive a tractable approximation and show that the resulting error remains small.
 - For the Ex-ante Balance algorithm, we prove that the matching size converges with high probability to a solution of a differential inclusion—a generalization of ODEs that captures the algorithm’s discontinuous decision rules. To our knowledge, this is the first use of differential inclusions to analyze online matching problems.
 - We extend this analysis to a generalized version of Ex-ante Balance, where the decision rule incorporates both connection probabilities and real-time availability, and show that its performance similarly converges to a well-defined differential inclusion.
- **When $p_{n,t}$ are unknown:** We study ETC – balance algorithm a bandit extension of the Ex-ante Balance algorithm, where the compatibility probabilities $p_{n,t}$ are unknown and must be learned over time. In this setting, the algorithm receives binary feedback for each matching attempt and must estimate the latent affinities between classes while making sequential, irrevocable decisions. We analyze the regret of this learning-based algorithm and prove that it is of order $\mathcal{O}(T^{\frac{q+3}{4}})$ for $0 < q < 1$ using stochastic approximations and differential inclusions tools.

Related works

Online bipartite matching has been extensively studied, particularly in adversarial and stochastic models ([22, 37] for a survey). In the adversarial setting, the Greedy algorithm guarantees a $1/2$ competitive ratio, improving to $1 - 1/e$ under random arrivals [18]. The Ranking algorithm achieves the optimal $1 - 1/e$ bound in this setting and performs even better with random arrivals [25, 13, 32]. In contrast, stochastic models assume known distributions over vertex types, often in the i.i.d. setting. This allows improved performance, with algorithms reaching competitive ratios up to 0.711 [33, 23, 11, 21]. However, the i.i.d. model overlooks graph structure, and in many practical or average-case settings, simple heuristics can match or outperform these algorithms [8]. This has motivated the study of stochastic input models that better reflect real-world graphs. Consequently,

another line of work focuses on applying online algorithms to specific random graph families. A foundational example is online matching in Erdős–Rényi graphs, particularly in the sparse regime where each edge exists independently with probability c/N [35, 7, 14]. Even for simple strategies like Greedy, analysis in this setting is nontrivial and yields valuable insights. The configuration model further generalizes this approach by prescribing degree distributions for vertices [39, 1]. Additional generalizations of Erdős–Rényi have introduced dynamic elements—for instance, models where node degrees evolve over time to reflect changing environments or behaviors [12]. The stochastic block model (SBM), a structured extension of Erdős–Rényi that captures community structure, has also been studied in the online setting. In the dense regime, [42] analyze max-weight policies and show they can achieve asymptotically perfect matchings. In the general SBM, [10] extend lower bounds from the Erdős–Rényi case, demonstrating that the 0.837 bound from [35] remains tight when communities have equal expected degrees. They propose several efficient heuristics for online matching in SBMs, which perform well empirically, but none are proven to achieve asymptotic optimality. However, all these works focus primarily on dense regimes or heuristics without theoretical guarantees. In particular, the sparse regime of SBM, which is highly relevant for many real-world applications where connections are scarce and structured, has received limited attention. From another perspective, bandit algorithms offer a general framework for decision-making under uncertainty, where limited feedback guides sequential choices. These models focus on balancing between exploration and exploitation and have found broad application in online learning and resource allocation [30, 41]. While conceptually distinct, they share core challenges with online matching and offer complementary insights.

1 Model

We consider the online bipartite matching problem, where the nodes on one side arrive sequentially, with an additional graph structure given by a *stochastic block model*. The latter is defined by a bipartite graph $\mathcal{G} = (\mathcal{N}, \mathcal{T}, \mathcal{E})$, where $\mathcal{N} = [N] := \{1, \dots, N\}$ is the set of “offline” nodes, $\mathcal{T} = [T]$ is the set of “online” nodes, and $\mathcal{E} \subset \mathcal{N} \times \mathcal{T}$ is the set of edges. This underlying graph is random, in the sense that each edge $(n, t) \in \mathcal{N} \times \mathcal{T}$ belongs to \mathcal{E} independently with some probability. The block model assumes that each node belongs to a latent class: we denote by $\mathcal{C} := [C]$ the set of classes on the offline side, and by $\mathcal{D} := [D]$ the set of classes on the online side. Each offline node $n \in \mathcal{N}$ is assigned a class $c(n) \in \mathcal{C}$, and each online node $t \in \mathcal{T}$ is assigned a class $d(t) \in \mathcal{D}$. These assignments are drawn independently: nodes on the offline side are sampled from a distribution μ over \mathcal{C} , and nodes on the online side are sampled from a distribution ν over \mathcal{D} . Given the class labels, the edge (n, t) appears in \mathcal{E} with probability $p_{n,t} = p(c(n), d(t)) \in [0, 1]$, where $p = (p(c, d))_{c \in \mathcal{C} \times \mathcal{D}}$ is a class-to-class affinity matrix. In this work, we focus on the *sparse regime*, where the underlying graph remains sparse as the number of offline nodes N grows. More precisely, we assume the existence of a non-negative matrix $a = (a_{c,d}) \in \mathbb{R}_+^{C \times D}$ such that $p(c, d) = \frac{a_{c,d}}{N}$. Moreover, we assume that $a_{c,d} \leq a$ for all $c \in \mathcal{C}, d \in \mathcal{D}$, where $a \in (0, N)$ is a fixed constant. This choice ensures that each offline node has a bounded expected degree, even as $N \rightarrow \infty$, which reflects realistic constraints in large-scale platforms where individual users or items interact with only a limited number of others.

As mentioned in the introduction, in the online matching problem, an algorithm ALG observes sequentially the vertices in \mathcal{T} and constructs on the fly a matching (i.e., a subset of edges such that any vertex belongs to at most one of them) irrevocably: after seeing the vertex $t \in \mathcal{T}$, it can decide to add irrevocably an edge $(n, t) \in \mathcal{E}$ to the current matching if t does not belong to an edge of the matching yet.

We shall now introduce some notations. We denote by b_c the proportion of nodes of \mathcal{N} of class $c \in \mathcal{C}$. We also assume that there exists some scaling factor $\alpha > 0$ such that $T = \alpha N$ (as we will consider asymptotic results when N is large). We will also define the Boolean variable $m_n(t)$ equal to 1 if, and only if, the vertex u has been included in the matching by ALG before the vertex t arrives (otherwise $m_n(t) = 0$). Additionally, we denote by $\mathcal{N}_c := \{n \in \mathcal{N}, c(n) = c\}$ the set of nodes of class $c \in \mathcal{C}$, by $\mathcal{M}_c(t) = \{n \in \mathcal{N}_c, m_n(t) = 1\}$ the set of vertices of class c that are already matched before seeing vertex $t \in \mathcal{T}$ (we denote by $M_c(t)$ its cardinality) and by $M(t) := \sum_{c \in \mathcal{C}} M_c(t)$ the size of the matching constructed so far. We also denote by $\mathcal{F}_c(t) = \{u \in \mathcal{N}_c \setminus \mathcal{M}_c(t), (n, t) \in \mathcal{E}\}$ the set of vertices of class c that are not matched so far (thus free), but such that (n, t) belongs to \mathcal{E} (they are the “available neighbors” of t of class c), and by $\mathcal{C}(t) = \{c \in [C] | \mathcal{F}_c(t) \neq \emptyset\}$ the set of classes that have available nodes at time $t \in [T]$. Finally, we shall denote by e_i the i -th basis vector of \mathbb{R}^C .

2 Known compatibility probabilities

2.1 A warm-up: Myopic algorithm

In this section, we present a simple yet foundational algorithm, `Myopic`, which serves as a baseline for more sophisticated algorithms presented later. This algorithm is designed to make fast, greedy decisions for matching, without attempting to look ahead or anticipate future availability. Specifically, when a new vertex $t \in \mathcal{T}$ arrives (e.g., a request or a user), the policy *selects a class* $c_t \in \mathcal{C}$ according to a fixed probability distribution, and then *attempts to match* t to an available node within that class. The selection is made without verifying beforehand whether any nodes in the chosen class are actually available at time t . As a result, `Myopic` is computationally simple and immediate in its decisions, but it may sometimes fail to make a match due to resource unavailability. Despite its reactive nature, `Myopic` is carefully designed to respect class-specific budget constraints and to maximize the expected long-term success rate of matches. The key component of the algorithm is the computation of a probability matrix $Q^*(c, d)$, which solves the following optimization problem:

$$Q^* \in \arg \max_Q \sum_{c,d} Q(c, d)p(c, d),$$

$$\text{s.t. } \sum_d Q(c, d)\nu(d) = b_c, \forall c \in \mathcal{C} \quad \text{and} \quad \sum_c Q(c, d)\nu(d) = \nu(d), \forall d \in \mathcal{D}.$$

In particular, Q^* represents the optimal transport plan that maps the distribution ν to the budget vector b , minimizing the transport cost with respect to $-p(c, d)$. This matrix can be computed efficiently using the Hungarian algorithm, with a computational complexity of $\mathcal{O}(CD(C + D))$. Notably, the special case where there is only one class (i.e., $C = D = 1$) reduces to the setting studied in [34].

Algorithm 1: `Myopic` policy

Output: Updated matching $M(t)$

- 1 Compute the optimal transport plan Q^*
 - 2 **for** $t \in [T]$ **do**
 - 3 Choose $c_t \in \mathcal{C}$ at random with probability $Q^*(c_t, d_t)/\nu(d_t)$
 - 4 **if** $\mathcal{F}_{c_t}(t) = \emptyset$ **then**
 - 5 $M(t) = M(t - 1)$
 - 6 **else**
 - 7 $M(t) = M(t - 1) \cup \{(n_t, t)\}$ for $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$
-

In order to understand the behavior of the `Myopic` policy in large-scale matching markets, we consider a deterministic approximation via an ordinary differential equation (ODE). The guiding intuition is that, as the number of agents grows (nodes in the graph), stochastic variability in the system averages out, and the evolution of the system can be captured by a smooth deterministic trajectory. However, the ODE associated with the dynamics of the `Myopic` policy Equation (1) is nonlinear and does not generally admit a closed-form solution, especially due to the complex structure of the underlying graph encoded in the parameters $a_{c,d}$ and $Q^*(c, d)$. This makes direct analysis challenging. In certain structured settings, however, the ODE becomes more tractable. For instance, when $a_{c,d} = c$ for all (c, d) which corresponds to the Erdős–Rényi model, the ODE simplifies and admits a closed-form solution, as shown in [35]. This illustrates how the structure of the graph can significantly affect the tractability of the analysis and motivates the need for approximation techniques in the more general, heterogeneous case.

To address this, we introduce a simplified surrogate ODE, which approximates the behavior of the original system while being analytically tractable. This auxiliary equation has a closed-form solution $\tilde{y}_c(t)$, allowing us to extract structural insights such as convergence rates and equilibrium behavior. Crucially, we rigorously control the error between the true ODE solution $y_c(t)$ and the approximate solution $\tilde{y}_c(t)$ by bounding it with an explicit error term $e_c(t)$.

The next theorem formalizes this two-step approximation strategy. It shows that:

- The normalized matching size produced by the `Myopic` policy closely follows the ODE solution $y_c(t)$ with high probability.

- The ODE solution is well-approximated by the simpler function $\tilde{y}_c(t)$ with a small and explicitly controlled error.

Theorem 1. Let $y_c : [0, \alpha] \rightarrow \mathbb{R}$ be the solution of the following ODE

$$\begin{cases} \dot{y}_c(s) &= \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - y_c(s))}) Q^*(c, d) \\ y_c(0) &= 0 \end{cases} \quad (1)$$

Then, for each class $c \in \mathcal{C}$, the matching size $M_c(t)$ produced by **Myopic** satisfies, for all $t \in [T]$

$$\left| \frac{M_c(t)}{N} - y_c(t/N) \right| \leq \frac{3L_c e^{\alpha L_c}}{N^{1/3}}, \text{ where } L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d) \quad (2)$$

with probability at least $1 - 2Ce^{-N^{1/3}L_c^2/8\alpha}$. Moreover, for $c \in [C]$, $y_c(t) = \tilde{y}_c(t) - e_c(t)$, where $e_c(0) = 0$, and $\tilde{y}_c(t) = b_c - b_c \exp(-tL_c)$, and e_c satisfies,

$$e_c(t) \leq \frac{J_c}{L_c} (1 - e^{-L_c t})$$

Where $J_c = \frac{b_c^2}{2} \sum_{d=1}^D a_{c,d}^2 Q^*(c, d)$.

2.2 Ex-ante Balance

The **Ex-ante Balance** algorithm builds on the limitations of the **Myopic** policy by trying to take into account node availability. When a node of class $c(t)$ arrives, the **Myopic** policy selects a compatible class c purely based on potential compatibility between the classes—without considering whether any unmatched nodes from that class are actually available at that moment. In contrast, **Ex-ante Balance** takes a more informed approach: it chooses the class c that maximizes the probability that at least one unmatched node is available and connected to the arriving node. This probability is estimated by the expression $1 - (1 - \frac{a_{c,i}}{N})^{Nb_c - M_c(t)}$, which captures the likelihood of an edge existing under a stochastic block model, where edge probabilities between classes are governed by parameters $a_{c,j}$.

Algorithm 2: Ex-ante Balance

Output: Updated matching $M(t)$

```

1 for  $t \in [T]$  do
2   Choose  $c_t = \arg \max_{c \in [1, C]} \sum_{j=1}^D (1 - (1 - \frac{a_{c,j}}{N})^{Nb_c - M_c(t)}) \nu(j)$ 
3   if  $\mathcal{F}_{c_t}(t) = \emptyset$  then
4      $M(t) = M(t-1)$ 
5   else
6      $M(t) = M(t-1) \cup \{(n_t, t)\}$  for  $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$ 
```

A key distinction between the **Myopic** and **Ex-ante Balance** policies lies in how they respond to the evolving state of the system—and this difference has important implications for their large-scale behavior. The **Myopic** approach selects a class based solely on static compatibility between node types, leading to smooth dynamics. As the system scales, the randomness introduced by node arrivals averages out, and the evolution of the system can be accurately described by an ordinary differential equation (ODE). This continuous, deterministic approximation leverages the fact that the **Myopic** policy's decision rule is smooth and Lipschitz-continuous. In contrast, the **Ex-ante Balance** policy takes a more strategic decision by selecting the class that maximizes the actual match probability. This introduces discontinuities into the process—small variations in the system state can lead to abrupt changes in the selected class. As a result, the system no longer evolves smoothly, and the assumptions required for ODE convergence break down.

To address this, we turn to the framework of differential inclusions, a generalization of ODEs designed to handle such discontinuous dynamics. Rather than prescribing a single trajectory, a differential inclusion allows the system's evolution to follow a set of possible directions at each point, capturing the non-smooth transitions in behavior driven by abrupt changes in decision rules. The following

theorem formalizes this connection by proving that, with high probability, the normalized matching sizes produced by **Ex-ante Balance** converge to a solution of a differential inclusion. For an introduction to differential inclusions and their relevance in this context, see Appendix A.

Theorem 2. *Let m be the unique solution of the differential inclusion*

$$\dot{m} \in F(m) := \text{conv} \left\{ f_{c,b_c}(m_c) e_c ; c \in \arg \max_{k \in [C]} f_{k,b_k}(m_k) \right\},$$

which is the convex hull of the mappings

$$f_{c,b_c}(x) = \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - x)}) \nu(d).$$

*Then the matching built by **Ex-ante Balance** satisfies for all $t \in [T]$ and $c \in \mathcal{C}$, with probability at least $1 - \frac{b\alpha}{N\epsilon^2}$,*

$$\left| \frac{M_c(t)}{N} - m_c(t/N) \right| \leq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}}, \quad (3)$$

The different constants in are defined by $L = \max_{c \in [C]} \sum_{d=1}^D a_{c,d} \nu(d)$, $\delta_c = \frac{1}{N} \sum_{d=1}^D \frac{a_{c,d}}{e} \nu(d)$, $K_\alpha = (c\alpha + \epsilon) e^{c\alpha}/c$, ϵ as defined in Lemma 11 and c in Lemma 7, $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d} b_c}) \nu(d)$, $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta_c + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha$.

As **Ex-ante Balance** algorithm always picks the class with the highest probability of connection, which decreases in case of match, it tends to progressively equalize these probabilities across the classes in \mathcal{C} . Once two classes have their probabilities (almost) equalized, these stay equal over time by decreasing at the same rate. This generates phases, in which the k "first" classes (with highest initial probability) have their probabilities equalized, decreasing at the same rate, up until reaching the probability of the $(k+1)^{\text{th}}$ class. This phasing allows to obtain an explicit formula for m . The remainder of the section introduces the exact formula and some intuitions of how it is built, while the verification proof is deferred to Appendix C.2.

In the large- N limit, the marginal probability that the next online node is connected to at least one node in class $c \in \mathcal{C}$, given a vector $\beta \in \mathbb{R}_+^C$ representing the proportions of available nodes in each class, is given by: $f_{c,\beta_c}(z) = \sum_{d \in \mathcal{D}} (1 - e^{-a_{c,d}(\beta_c - z)}) \nu(d)$. W.l.o.g., we assume for the analysis that the elements of \mathcal{C} are ordered by decreasing marginal probability of receiving at least one edge at the initial time step – i.e. $f_{1,b_1}(0) \geq f_{2,b_2}(0) \geq \dots \geq f_{C,b_C}(0)$. Additionally, as $a_{c,d} > 0$, f_{c,β_c} is strictly decreasing and thus invertible, with f_{c,β_c}^{-1} also strictly decreasing.

During phase k , the $C - k$ last classes are not selected at all, while the k first classes have their probabilities equalized, decreasing at the same rate. Given the budgets $\beta \in \mathbb{R}_+^C$ at the beginning of the phase, the number of nodes matched during the phase, in each of the k classes, at time t , ends up evolving following $\mu_{k,\beta}(t)$ defined as the solution of the following separable ODE

$$\begin{cases} \frac{d\mu_{k,\beta}}{dt} = \text{dt} \\ \mu_{k,\beta}(0) = 0 \end{cases} \quad \text{where} \quad F_{k,\beta} = \left(\sum_{c=1}^k f_{c,\beta_c}^{-1} \right)^{-1}.$$

Note that, as $\frac{d\mu_{k,\beta}}{dt} > 0$, $\mu_{k,\beta}$ is strictly increasing, positive and invertible.

To assemble the phases, and provide the full expression of m , two sequences need to be defined. First, the sequence of time-steps $(t_k)_{k \in [C]}$ defines the phases. Then $(\beta^{(k)})_{k \in [C]}$ describes the proportion of available nodes per class at the start of each phase. They are defined as follows:

$$\forall c, k \in [C], \beta_c^{(k)} = \begin{cases} b_c & \text{if } k \leq c \\ \beta_c^{(k-1)} - f_{c,\beta^{(k-1)}}^{-1}(f_{k,\beta^{(1)}}(0)) & \text{otherwise} \end{cases} \quad (4)$$

$$\forall k \in [C+1], t_k = \begin{cases} 0 & \text{if } k = 1 \\ \min \left(T, t_{k-1} + \mu_{k-1,\beta^{(k-1)}}^{-1} \left(F_{k-1,\beta^{(k-1)}}^{-1}(f_{k,\beta^{(1)}}(0)) \right) \right) & \text{otherwise} \end{cases} \quad (5)$$

Theorem 3. For any $c \in \mathcal{C}$, the function

$$m_c^* : t \mapsto (b_c - \beta_c^{(k_t)}) + \left(f_{c, \beta^{(k_t)}}^{-1} (F_{k_t, \beta^{(k_t)}} (\mu_{k_t, \beta^{(k_t)}} (t - t_{k_t}))) \right)_+ \quad (6)$$

where $k_t = \max\{k \in [C] : t > t_k\}$, is the solution of the differential inclusion defined in Theorem 2.

Figure 1 illustrates the accuracy of $m_c^*(t/N)$ to estimate $\frac{M_c(t)}{N}$ in the sparse regime.

2.3 Ex-post Balance

While the Ex-ante Balance policy improves upon the purely compatibility-driven Myopic approach by considering the probability of successful matches, it still relies on an expected availability model—it selects the class that likely has unmatched nodes, without verifying their presence. This can lead to wasted opportunities when the chosen class ends up being empty. To address this limitation, we introduce the Ex-post Balance policy, which takes an additional, more grounded step. Instead of only maximizing the expected match probability, Ex-post Balance ensures that the selected class actually contains at least one available node at the time of decision. This additional check prevents the algorithm from targeting unavailable options, making it more efficient. We prove that the matching size under Ex-post Balance converges with high probability to a differential inclusion. This differential inclusion differs from the one in Theorem 2, as it explicitly accounts for real-time availability constraints. Full details and results are provided in Appendix C.3.

3 Unknown compatibility probabilities

In many real-world applications, the underlying parameters of the graph—such as the connection probabilities between node classes or the distribution of arriving node types—are not known *a priori*. These parameters may be shaped by latent variables, evolve over time, or be inferred only through noisy and partial observations. Consequently, algorithms must operate under uncertainty and learn these parameters dynamically in order to make effective matching decisions. In this section, we study the setting in which the connection probabilities $a_{c,d}$ are unknown and must be estimated online. This transforms the problem into a bandit setting, where each class $c \in \mathcal{C}$ can be viewed as an arm, and upon choosing class c_t at time t , a Bernoulli reward is observed—indicating whether a successful match occurred between the arriving node and an available node in class c_t .

3.1 ETC – balance

Algorithm 3 presents the ETC – balance policy, which combines a fixed-duration Explore-Then-Commit (ETC) strategy with the Ex-ante Balance rule for class selection. The algorithm proceeds in two phases. During the exploration phase, which lasts for a fixed number of rounds T_{explore} , arriving nodes are matched by selecting classes uniformly at random. This allows the algorithm to collect data on the outcomes of triplets (c, d, m) , where $c \in \mathcal{C}$, $d \in \mathcal{D}$, and m is the current matching size of the class c . We let $T_{c,d,m}$ denote the number of times the triplet (c, d, m) has been observed. After this phase, the algorithm enters the commitment phase and uses the collected data to estimate the match failure probabilities. Specifically, it estimates $D_{c,d}(m) = \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - m}$ using an estimator $\hat{D}_{c,d}(m)$, whose form and concentration properties are given in ???. These estimates are then incorporated into the Ex-ante Balance rule to select the class c that maximizes the expected success probability, weighted by the type distribution $\nu(d)$.

3.2 Regret

For each class $i \in \mathcal{C}$, let $M_i(t)$ and $\hat{M}_i(t)$ denote the number of matches made by the Ex-ante Balance and ETC – balance algorithms, respectively, up to time t . We define the regret of ETC – balance as the total difference in matching performance across all classes when compared to Ex-ante Balance, which has full knowledge of the match probabilities. In other words, the regret quantifies the cumulative loss incurred by ETC – balance due to not knowing the match probabilities in advance. The following result shows that this regret grows at most on the order of $\mathcal{O}(N^{(q+3)/4})$ for some $0 < q < 1$.

Algorithm 3: ETC – balance

Input: $T, \nu(d), N, T_{\text{explore}},$ confidence level δ **Output:** Matching $M(t)$

```
1 Init:  $\mathcal{M}(0) = \emptyset, M_c(0) = 0, T_{c,d,m}(0) = 0, \hat{D}_{c,d}(m) = 1$ 
2 for  $t = 1$  to  $T$  do
3   Node  $t$  of type  $d(t)$  arrives
4   if  $t \leq T_{\text{explore}}$  then
5     Select  $c_t$  uniformly at random
6   else
7     foreach  $c \in [C]$  do
8       Let  $m = M_c(t-1), s(m) = \frac{b_c}{b_c - m/N}$ 
9       Compute  $\hat{D}_{c,d}(m)$ 
10    Choose  $c_t = \arg \max_c \sum_d 1 - \hat{D}_{c,d}(M_c(t-1))\nu(d)$ 
11  if  $\mathcal{F}_{c_t}(t) \neq \emptyset$  then
12    Match  $(n_t, t)$  for  $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$ , update  $\mathcal{M}(t), M_{c_t}(t)$ , and  $Y_{c_t,d(t)}(t) = 1$ 
13  else
14    No match,  $Y_{c_t,d(t)}(t) = 0$ 
15  Update  $T_{c_t,d(t),M_{c_t}(t)} \leftarrow T_{c_t,d(t),M_{c_t}(t)} + 1$ 
```

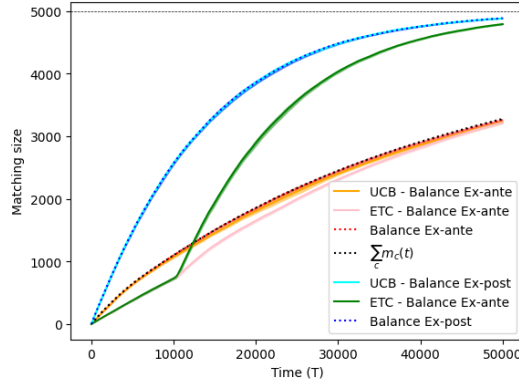


Figure 1: Illustration of the matching size for the different methods. $T = 50000$, $N = 5000$, $C = 5$, $D = 6$ and simulations are averaged over 20 trajectories. **1.** As expected the *ex-post* versions of the algorithms perform better than their *ex-ante* counterparts. **2.** The fluid limit m^* (dark) is an accurate estimator of the actual empirical trajectory of Ex-ante Balance (red). **3.** A UCB version of balance can be built using the confidence set from ?? and performs empirically better than ETC. Yet, its analysis remains an open question.

Theorem 4. Let $R(T) = \sum_{i \in C} M_i(T) - \hat{M}_i(T)$ denote the regret of ETC – balance. Suppose the exploration phase lasts for $T_{\text{explore}} = T^{\frac{q+3}{4}}$, for some $0 < q < 1$. Then the regret satisfies

$$R(T) = \mathcal{O}(T^{\frac{q+3}{4}}).$$

4 Conclusion

We studied online bipartite matching within the Stochastic Block Model (SBM), capturing structured heterogeneity in real-world networks through class-dependent connection probabilities. We analyzed two main algorithms under known probabilities: the Myopic policy, which is simple but limited by ignoring availability, and the Ex-ante Balance algorithm, which accounts for compatibility and capacity, and is shown to converge to a differential inclusion. When the probabilities are unknown, we introduced ETC – balance, a bandit-based extension that learns affinities over time and

achieves sublinear regret. Simulations confirm that UCB-based methods outperform ETC — balance, while Ex-ante Balance under known probabilities closely matches actual outcomes, validating its effectiveness. A promising future research direction is to analyze UCB in this setting using differential inclusion tools, to better understand its asymptotic behavior and theoretical guarantees.

References

- [1] A. Aamand, J. Y. Chen, and P. Indyk. (optimal) online bipartite matching with degree information. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022.
- [2] E. Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.
- [3] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed membership stochastic blockmodels. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2008.
- [4] J.-P. Aubin and A. Cellina. *Differential Inclusions: Set-Valued Maps and Viability Theory*, volume 264 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1984.
- [5] B. Bollobás. *Random Graphs*. Cambridge University Press, 2001.
- [6] S. P. Borgatti and M. G. Everett. Models of core/periphery structures. *Social Networks*, 21(4): 375–395, 2000.
- [7] A. Borodin, C. Karavasilis, and D. Pankratov. Greedy bipartite matching in random type poisson arrival model, 2018.
- [8] A. Borodin, C. Karavasilis, and D. Pankratov. An experimental study of algorithms for online bipartite matching. *ACM J. Exp. Algorithmics*, 25, mar 2020. ISSN 1084-6654. doi: 10.1145/3379552.
- [9] A. Borodin, C. MacRury, and A. Rakheja. Bipartite stochastic matching: Online, random order, and i.i.d. models. 04 2020. doi: 10.48550/arXiv.2004.14304.
- [10] A. Brandenberger, B. Chin, N. S. Sheffield, and D. Shyamal. Matching Algorithms in the Sparse Stochastic Block Model. In C. Mailler and S. Wild, editors, *35th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2024)*, volume 302 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 16:1–16:21, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-329-4. doi: 10.4230/LIPIcs.AofA.2024.16.
- [11] B. Brubach, K. A. Sankararaman, A. Srinivasan, and P. Xu. Online stochastic matching: New algorithms and bounds. *Algorithmica*, 82:2737 – 2783, 2016.
- [12] M. Cherifa, C. Calauzenes, and V. Perchet. Dynamic online matching with budget refills. 05 2024. doi: 10.48550/arXiv.2405.09920.
- [13] N. R. Devanur, K. Jain, and R. D. Kleinberg. Randomized primal-dual analysis of ranking for online bipartite matching. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA ’13*, page 101–107, USA, 2013. Society for Industrial and Applied Mathematics. ISBN 9781611972511.
- [14] M. Dyer, A. Frieze, and B. Pittel. The average performance of the greedy matching algorithm. *The Annals of Applied Probability*, 3(2):526–552, 1993.
- [15] A. F. Filippov. *Differential Equations with Discontinuous Right-Hand Sides*, volume 18 of *Mathematics and Its Applications (Soviet Series)*. Kluwer Academic Publishers, Dordrecht, 1988. Translated from the Russian.
- [16] N. Gast and B. Gaujal. Markov chains with discontinuous drifts have differential inclusions limits. Application to stochastic stability and mean field approximation. Research Report RR-7315, Apr. 2011.

- [17] C. D. Godsil. Matchings and walks in graphs. *J. Graph Theory*, 5:285–297, 1981.
- [18] G. Goel and A. Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '08, page 982–991, USA, 2008. Society for Industrial and Applied Mathematics.
- [19] E. F. Grove, M.-Y. Kao, P. Krishnan, and J. S. Vitter. Online perfect matching and mobile computing. In *Algorithms and Data Structures: 4th International Workshop, WADS'95 Kingston, Canada, August 16–18, 1995 Proceedings 4*, pages 194–205. Springer, 1995.
- [20] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983.
- [21] Z. Huang, X. Shu, and S. Yan. The power of multiple choices in online stochastic matching. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 91–103, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392648. doi: 10.1145/3519935.3520046.
- [22] Z. Huang, Z. G. Tang, and D. Wajc. Online matching: A brief survey. *SIGecom Exch.*, 22(1): 135–158, Oct. 2024. doi: 10.1145/3699824.3699837.
- [23] P. Jaillet and X. Lu. Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, 39(3):624–646, 2014. ISSN 0364765X, 15265471.
- [24] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011.
- [25] R. M. Karp, U. V. Vazirani, and V. V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing*, STOC '90, page 352–358, New York, NY, USA, 1990. Association for Computing Machinery. ISBN 0897913612.
- [26] B. Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.
- [27] H. K. Khalil. *Nonlinear systems; 3rd ed.* Prentice-Hall, Upper Saddle River, NJ, 2002. The book can be consulted by contacting: PH-AID: Wallet, Lionel.
- [28] M. Kim and S. Oh. Stochastic blockmodel with cluster-dependent connection probabilities for modeling bipartite networks. *Journal of Machine Learning Research*, 21(234):1–43, 2020.
- [29] M. Kunze. *Non-Smooth Dynamical Systems*, volume 1744 of *Lecture Notes in Mathematics*. Springer, Berlin, 2000. ISBN 978-3-540-67993-6. doi: 10.1007/BFb0103843.
- [30] T. Lattimore and C. Szepesvari. *Bandit Algorithms*. 2019.
- [31] L. Lovász and M. D. Plummer. *Matching Theory*, volume 367. American Mathematical Society, 2009.
- [32] M. Mahdian and Q. Yan. Online bipartite matching with random arrivals: An approach based on strongly factor-revealing lps. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, page 597–606, New York, NY, USA, 2011. Association for Computing Machinery.
- [33] V. H. Manshadi, S. O. Gharan, and A. Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012. ISSN 0364765X, 15265471.
- [34] A. Mastin and P. Jaillet. Greedy online bipartite matching on random graphs. 07 2013.
- [35] A. Mastin and P. Jaillet. Greedy online bipartite matching on random graphs. *ArXiv*, abs/1307.2536, 2013.
- [36] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.

- [37] A. Mehta. 2013. doi: 10.1561/04000000057.
- [38] A. Mehta. Online matching and ad allocation. 8 (4):265–368, 2013.
- [39] N. Noiry, V. Perchet, and F. Sentenac. Online matching in sparse random graphs: Non-asymptotic performances of greedy algorithm. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021.
- [40] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency*, volume 24. Springer Science & Business Media, 2003.
- [41] A. Slivkins. Introduction to multi-armed bandits. *Found. Trends Mach. Learn.*, 12(1–2):1–286, Nov. 2019. ISSN 1935-8237. doi: 10.1561/22000000068.
- [42] N. Soprano-Loto, M. Jonckheere, and P. Moyal. Online matching for the multiclass stochastic block model. *arXiv preprint arXiv:2303.15374*, 2023.
- [43] L. Warnke. On wormald’s differential equation method. *ArXiv*, abs/1905.08928, 2019.
- [44] N. C. Wormald. Differential equations for random processes and random graphs. *Annals of Applied Probability*, 5:1217–1235, 1995.
- [45] L. Zdeborová and M. Mézard. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05003, may 2006. doi: 10.1088/1742-5468/2006/05/P05003.

A Differential inclusions

This section aims to introduce the fundamental concepts of differential inclusions. Unlike ordinary differential equations (ODE), where the derivative of the unknown function is determined by a single-valued map, differential inclusions generalize this by allowing the derivative to lie within a set-valued map. This broader framework is well-suited for modeling dynamical systems that involve uncertainty, discontinuities, or control constraints.

A.1 Set-Valued Maps

Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denote a set-valued map, i.e a mapping which assigns to each point $x \in \mathbb{R}^n$ a subset $F(x) \subseteq \mathbb{R}^n$. We consider in general that $F(x)$ is nonempty for all $x \in \mathbb{R}^n$. The notation $\langle x, y \rangle$ denotes the standard inner product on \mathbb{R}^d , and the norm is given by $\|x\| = \sqrt{\langle x, x \rangle}$. For a set $A \subset \mathbb{R}^d$, we define its norm as $\|A\| = \sup_{x \in A} \|x\|$.

A set-valued map F is said to be:

- **Upper semicontinuous (u.s.c.):** A set-valued map $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is upper semicontinuous at a point $y \in \mathbb{R}^n$ if for every sequence $y^{(n)} \rightarrow y$ and any sequence $x_n \in F(y^{(n)})$ such that $x_n \rightarrow x$, it holds that $x \in F(y)$.
- **Locally bounded** if for every compact set $K \subset \mathbb{R}^n$, there exists a constant M such that,

$$\sup_{x \in K} \sup_{v \in F(x)} \|v\| \leq M.$$

- **Measurable** if its graph is measurable in the product sigma-algebra on $\mathbb{R}^n \times \mathbb{R}^n$.
- **One sided Lipschitz** with constant L if for all $z, \tilde{z} \in \mathbb{R}^n$ and for $z \in F(y), \tilde{z} \in F(\tilde{y})$,

$$\langle y - \tilde{y}, z - \tilde{z} \rangle \leq L \|y - \tilde{y}\|^2$$

A.2 Definition of differential inclusions

Definition 1. Let $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a set valued and $T > 0$. A differential inclusion is defined as,

$$\dot{x}(t) \in F(x(t)) \quad \text{for } t \in [0, T] \quad (7)$$

together with an initial condition,

$$x(0) = x_0$$

Let $I \subseteq \mathbb{R}$, a function $x : I \rightarrow \mathbb{R}^n$ is a solution to the differential inclusion defined in Equation (68) with initial condition $x(0) = x_0$ if there exists a function $\phi : I \rightarrow \mathbb{R}^n$ such that:

- For all $t \in I$: $x(t) = x(0) + \int_0^t \phi(s) ds$.
- For almost every $t \in I$ $\phi(t) \in F(x(t))$.

A.3 Existence and uniqueness of the solution

Proposition. ([15, 4, 29])

- If F is upper semicontinuous and if there exists c such that $\|F(y)\| \leq c(1 + \|y\|)$ then for any initial condition x_0 , $\dot{x} \in F(x)$ has at least one solution on $[0, +\infty)$ with $x(0) = x_0$.
- If F is one-sided Lipschitz, then for all $T > 0$ there exists at most one solution of $\dot{x} \in F(x)$ on $[0, T]$

If F is upper semicontinuous and one-sided Lipschitz then for $T > 0$, $\dot{x} \in F(x)$ has a unique solution on $[0, T]$.

B Myopic algorithm

This section is organized into two parts. In the first part, we analyze the `Myopic` algorithm. In the second part, we show that, under certain assumptions, the model reduces to the Erdős–Rényi case studied in [35]. We begin by considering the following `Myopic` algorithm, as introduced in the main paper:

Algorithm 4: Myopic policy

Output: Updated matching $M(t)$

```

1 Compute the optimal transport plan  $Q^*$ 
2 for  $t \in [T]$  do
3   Choose  $c_t \in \mathcal{C}$  at random with probability  $Q^*(c_t, d_t)/\nu(d_t)$ 
4   if  $\mathcal{F}_{c_t}(t) = \emptyset$  then
5      $M(t) = M(t-1)$ 
6   else
7      $M(t) = M(t-1) \cup \{(n_t, t)\}$  for  $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$ 

```

B.1 Proof of Theorem 1

Theorem 1. Let $y_c : [0, \alpha] \rightarrow \mathbb{R}$ be the solution of the following ODE

$$\begin{cases} \dot{y}_c(s) &= \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - y_c(s))}) Q^*(c, d) \\ y_c(0) &= 0 \end{cases} \quad (1)$$

Then, for each class $c \in \mathcal{C}$, the matching size $M_c(t)$ produced by `Myopic` satisfies, for all $t \in [T]$

$$\left| \frac{M_c(t)}{N} - y_c(t/N) \right| \leq \frac{3L_c e^{\alpha L_c}}{N^{1/3}}, \text{ where } L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d) \quad (2)$$

with probability at least $1 - 2Ce^{-N^{1/3}L_c^2/8\alpha}$. Moreover, for $c \in [C]$, $y_c(t) = \tilde{y}_c(t) - e_c(t)$, where $e_c(0) = 0$, and $\tilde{y}_c(t) = b_c - b_c \exp(-tL_c)$, and e_c satisfies,

$$e_c(t) \leq \frac{J_c}{L_c} (1 - e^{-L_c t})$$

Where $J_c = \frac{b_c^2}{2} \sum_{d=1}^D a_{c,d}^2 Q^*(c, d)$.

The proof of Theorem 1 is based on the Wormald theorem [43, 44] and is structured as follows:

- We first define the evolution of $M_c(t)$, the size of the matching constructed by `Myopic` in class $c \in [C]$ at time $t \in [T]$.
- We then verify that $M_c(t)$ satisfies the conditions required to apply the Wormald theorem [43, 44].
- We apply the Wormald theorem [43, 44] to analyze the behavior of $M_c(t)$.
- Next, we construct an approximate solution to the differential equation that serves as a continuous approximation of $M_c(t)$.
- Finally, we derive an explicit bound on the error between the true solution and its approximation.

Let $\mathbf{M}(t) = (M_1(t), \dots, M_C(t))$ denote the vector of matching sizes in each class $c \in [C]$ constructed by the `Myopic` algorithm. For each class $c \in [C]$, the matching size evolves according to the following dynamics:

$$M_c(t+1) = M_c(t) + \mathbb{1}_{\{\exists n \in \mathcal{N}_c(t) \text{ s.t. } c_t^* = c \text{ and } m_n(t+1)=1\}} \quad (8)$$

Here, c_t^* represents the class selected by `Myopic` at time t .

The first step is to compute the expected one-step change in $M_c(t)$, the size of the matching constructed by `Myopic`. This is formalized in the following lemma.

Lemma 1. For $t \in [T]$, $c \in [C]$ and for any $B = (b_1, \dots, b_C)$, the expectation of the one-step change of $M_c(t)$, when matching is constructed using **Myopic** is given by,

$$\mathbb{E}[M_c(t+1) - M_c(t) | \mathbf{M}(t), B] = F_c(t, M_1(t), \dots, M_C(t)) \quad (9)$$

where $F_c(t, M_1(t), \dots, M_C(t)) = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) Q^*(c, d)$

Proof. Moving to conditional expectation gives,

$$\mathbb{E}[M_c(t+1) - M_c(t) | \mathbf{M}(t), B] \quad (10)$$

$$= \mathbb{E}[\mathbb{1}_{\{\exists n \in \mathcal{N}_c(t), c_t^* = c, m_n(t+1) = 1\}} | \mathbf{M}(t), B] \quad (11)$$

$$= \mathbb{P}(\exists n \in \mathcal{N}_c(t), c_t^* = c, m_n(t+1) = 1 | \mathbf{M}(t), B) \quad (12)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t), c_t^* = c, m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d) \nu(d) \quad (13)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t), m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d, c_t^* = c) \nu(d) \quad (14)$$

$$= \sum_{d=1}^D \left(1 - (1 - p(c, d))^{Nb_c - M_c(t)}\right) \nu(d) Q^*(c, d) / \nu(d) \quad (15)$$

$$= \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) Q^*(c, d) \quad (16)$$

□

The following lemma establishes the Lipschitz continuity of the function $f_c(x) = \sum_{d=1}^D (1 - e^{-a_{c,d}(Nb_c - x)}) Q^*(c, d)$.

Lemma 2. For $x \leq Nb_c$ and $c \in [C]$, the function f_c is L_c -Lipschitz with $L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d)$.

Proof. For $x \leq Nb_c$, the function f_c is a sum of continuous and differentiable functions on \mathbb{R} . Its derivative is given by

$$|f'_c(x)| = \sum_{d=1}^D e^{-a_{c,d}(Nb_c - x)} a_{c,d} Q^*(c, d) \leq \sum_{d=1}^D a_{c,d} Q^*(c, d). \quad (17)$$

Since f_c is differentiable, the Mean Value Theorem implies that for any $x, y \in \mathbb{R}$ with $x, y \leq Nb_c$, there exists $\xi \in (x, y)$ such that

$$|f_c(x) - f_c(y)| = |f'_c(\xi)| |x - y| \leq L_c |x - y|, \quad (18)$$

where the Lipschitz constant is defined by $L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d)$. □

The following lemma is a technical lemma,

Lemma 3. For $n > 0$, $a \leq n/2$ and $0 \leq w \leq 1$,

$$0 \leq e^{-aw} - \left(1 - \frac{a}{n}\right)^{nw} \leq \frac{a}{ne}$$

Proof. Using the following inequalities: $1 - x \geq e^{-x-x^2}$ for $x \leq \frac{1}{2}$ and $1 - x \leq e^{-x}$ for $x \geq 0$, we obtain $e^{-aw} \left(1 - \frac{a^2 w}{n}\right) \leq \left(1 - \frac{a}{n}\right)^{nw} \leq e^{-aw}$. The result follows by rearranging terms and using that $awe^{-aw} \leq 1/e$. □

The next lemma bounds the distance between F_c defined in Lemma 1 and f_c ,

Lemma 4. For $c \in [C]$,

$$|F_c(t, M_1(t), \dots, M_C(t)) - f_c(M_c(t))| \leq \sum_{d=1}^D \frac{a_{c,d}}{Ne} Q^*(c, d) \quad (19)$$

Proof. For $c \in [C]$,

$$|F_c(t, M_1(t), \dots, M_C(t)) - f_c(M_c(t))| \quad (20)$$

$$= \left| \sum_{d=1}^D e^{-a_{c,d}(Nb_c - M_c(t))} Q^*(c, d) - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)} Q^*(c, d) \right| \quad (21)$$

$$\leq \sum_{d=1}^D \frac{a_{c,d}}{Ne} Q^*(c, d) \quad (\text{Lemma 3}) \quad (22)$$

□

Now we are ready to prove Theorem 1.

Proof. From Lemma 1, we have

$$\mathbb{E}[M_c(t+1) - M_c(t) | \mathbf{M}(t), B] = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) Q^*(c, d) \quad (23)$$

Let $Y_c(s) = \frac{M_c(sN)}{N}$ denote the normalized matching size in class $c \in [C]$, and define the vector $\mathbf{Y}(s) = (Y_1(s), \dots, Y_C(s))$ for $0 \leq s \leq T/N$. Then, we obtain:

$$\frac{\mathbb{E}[Y_c(s+1/N) - Y_c(s) | \mathbf{Y}(s), B]}{1/N} = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - NY_c(s)}\right) Q^*(c, d) \quad (24)$$

As $N \rightarrow \infty$, we find:

$$\text{for } s \in \left[\frac{T}{N}\right] \quad \dot{y}_c(s) = \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - y_c(s))}) Q^*(c, d), \text{ and } y_c(0) = 0$$

Applying Wormald's theorem [43], and considering the domain $0 \leq y_s \leq 1$ with $\beta = 1$ (by the nature of the matching process), we define the Lipschitz constant as $L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d)$ (see Lemma 2). Additionally, we set $\delta = \sum_{d=1}^D \frac{a_{c,d}}{Ne} Q^*(c, d)$ as in Lemma 4, which gives $\lambda = N^{-1/3} \sum_{d=1}^D a_{c,d} Q^*(c, d)$. Therefore, with probability at least $1 - 2Ce^{-N^{1/3} L_c^2 / 8\alpha}$ the approximation holds.

$$|M_c(t) - Ny_c(t/n)| \leq 3e^{L_c \alpha} N^{2/3} L_c \quad (25)$$

Where y_c satisfies for $s \in [T/n]$,

$$\begin{aligned} \dot{y}_c(s) &= \sum_{d=1}^D \left(1 - e^{-a_{c,d}(b_c - y_c(s))}\right) Q^*(c, d) \\ y_c(0) &= 0 \end{aligned} \quad (26)$$

Now we want to approximate the solution of Equation (26). To simplify the analysis, we introduce the change of variable $z_c(t) = y_c(t) - b_c$ where $t \in [T/N]$. Under this transformation, Equation (26) becomes:

$$\begin{cases} \dot{z}_c(t) &= \sum_{d=1}^D (1 - e^{a_{c,d} z_c(t)}) Q^*(c, d) \\ z_c(0) &= -b_c \end{cases} \quad (27)$$

The right-hand side of Equation (27) involves terms of the form $1 - e^{a_{c,d}z_c}$. Since $-b_c \leq z_c \leq 0$, we can use the Taylor expansion:

$$1 - e^{a_{c,d}z_c} = -a_{c,d}z_c + \mathcal{O}(z_c^2).$$

Consider now \tilde{z}_c be the solution of the following linearized differential equation with initial condition $\tilde{z}_c(0) = -b_c$ and $t \in [T/N]$:

$$\tilde{z}'_c(t) = -\tilde{z}_c(t) \sum_{d=1}^D a_{c,d} Q^*(c, d). \quad (28)$$

The solution to Equation (28) is explicitly given by:

$$\tilde{z}_c(t) = -b_c \exp \left(-t \sum_{d=1}^D a_{c,d} Q^*(c, d) \right).$$

With \tilde{z}_c in hand, we are ready to bound the error between the exact solution of Equation (27) and its approximation \tilde{z}_c . Let $e_c = \tilde{z}_c - z_c$ be the approximation error, its derivative for $t \in [T/N]$ is given by,

$$e'_c(t) = \tilde{z}'_c(t) - z'_c(t) \quad (29)$$

$$= \sum_{d=1}^D (-a_{c,d} \tilde{z}_c(t) - 1 + e^{a_{c,d}z_c(t)}) Q^*(c, d) \quad (30)$$

Since $z_c \in [-b_c, 0]$, we have the inequality

$$1 - e^{a_{c,d}z_c(t)} \leq -a_{c,d}z_c(t) \quad \text{for all } d = 1, \dots, D.$$

This implies,

$$\dot{z}_c(t) = \sum_{d=1}^D \left(1 - e^{a_{c,d}z_c(t)} \right) Q^*(c, d) \leq -z_c(t) \sum_{d=1}^D a_{c,d} Q^*(c, d).$$

This upper bound matches the right-hand side of the linearized system Equation (28) evaluated at $z_c(t)$. Since both solutions share the same initial condition, $z_c(0) = \tilde{z}_c(0) = -b_c$, we can apply the comparison principle [27, Chapter 3, Lemma 3.4].

It follows that

$$z_c(t) \leq \tilde{z}_c(t) \quad \text{for all } t \in [T/N],$$

which implies that the approximation error

$$e_c(t) := \tilde{z}_c(t) - z_c(t) \geq 0.$$

For all $t \in [T/N]$, using Taylor expansion of order 2,

$$e'_c(t) \leq \sum_{d=1}^D \left(-a_{c,d} \tilde{z}_c(t) + a_{c,d} z_c(t) + \frac{a_{c,d}^2 z_c^2(t)}{2} \right) Q^*(c, d) \quad (31)$$

$$e'_c(t) \leq \sum_{d=1}^D \left(-a_{c,d} e_c(t) + \frac{a_{c,d}^2 b_c^2}{2} \right) Q^*(c, d) \quad (\text{using } z_c \in [-b_c, 0]) \quad (32)$$

$$e'_c(t) \leq -e_c(t) L_c + J_c \quad (33)$$

We multiply Equation (33) by $e^{L_c t}$,

$$e'_c(t) e^{L_c t} + L_c e^{L_c t} e_c(t) \leq J_c e^{L_c t} \quad (34)$$

Integrating both sides,

$$e^{t L_c} e_c(t) - e_c(0) \leq J_c \frac{e^{L_c t} - 1}{L_c} \quad (35)$$

Thus,

$$e_c(t) \leq \frac{J_c}{L_c}(1 - e^{-L_c t}) + e(0)e^{-L_c t} \quad (36)$$

$$e_c(t) \leq \frac{J_c}{L_c}(1 - e^{-L_c t}) \quad (37)$$

Thus, for $t \in [0, \frac{T}{N}]$,

$$z_c(t) = \tilde{z}_c(t) - e_c(t)$$

where $e_c(t)$ satisfies Equation (37), with $L_c = \sum_{d=1}^D a_{c,d} Q^*(c, d)$ and $J_c = \sum_{d=1}^D \frac{a_{c,d} b_c^2}{2} Q^*(c, d)$.

Thus replacing $z_c(t)$ by $y_c(t) - b_c$ we get the final result. \square

B.2 Recovering the Erdős–Rényi case

Consider the special case where the connection probability depends only on the class c , that is, $p(c, d) = \frac{a_c}{n}$. In this setting, the graph structure introduced in Section 1 simplifies to an Erdős–Rényi random graph [5, 24]. Under this assumption, Equation (26) reduces to:

$$\dot{z}_c(t) = \left(1 - e^{-a_c z_c(t)}\right) \sum_{d=1}^D Q^*(c, d), \quad (38)$$

$$\frac{-a_c \dot{z}_c(t) e^{-a_c z_c(t)}}{e^{-a_c z_c(t)} - 1} = -a_c \sum_{d=1}^D Q^*(c, d). \quad (39)$$

Integrating both sides with respect to time yields:

$$\ln |e^{-a_c z_c(t)} - 1| - \ln |e^{-a_c b_c} - 1| = -a_c t \sum_{d=1}^D Q^*(c, d). \quad (40)$$

Solving for $z_c(t)$, we obtain the closed-form expression:

$$z_c(t) = -\frac{1}{a_c} \ln \left(1 + (e^{-a_c b_c} - 1) e^{-a_c t \sum_{d=1}^D Q^*(c, d)}\right). \quad (41)$$

This shows that when the model is reduced to the Erdős–Rényi setting, we obtain an exact solution to the corresponding differential equation. Consequently, with high probability, the size of the matching in each class $c \in [C]$ concentrates around z_c as given in Equation (41). These results are in close agreement with those found in [34], which also examined the dynamics of matching in Erdős–Rényi graphs and observed similar asymptotic behavior.

C Balance algorithm

This section is organized into two parts. First, we analyze the case where matching is performed using the Ex-ante Balance algorithm, as defined in the main paper. Next, we extend the analysis to the Ex-post Balance algorithm.

C.1 Ex-ante Balance

We consider the following Ex-ante Balance algorithm.

Algorithm 5: Ex-ante Balance

Output: Updated matching $M(t)$

```

1 for  $t \in [T]$  do
2   Choose  $c_t = \arg \max_{c \in [1, C]} \sum_{d=1}^D (1 - (1 - \frac{a_{c,d}}{N})^{N b_c - M_c(t)}) \nu(d)$ 
3   if  $\mathcal{F}_{c_t}(t) = \emptyset$  then
4      $M(t) = M(t-1)$ 
5   else
6      $M(t) = M(t-1) \cup \{(n_t, t)\}$  for  $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$ 
```

C.1.1 Proof of Theorem 2

Theorem 2. Let m be the unique solution of the differential inclusion

$$\dot{m} \in F(m) := \text{conv} \left\{ f_{c,b_c}(m_c) e_c ; c \in \arg \max_{k \in [C]} f_{k,b_k}(m_k) \right\},$$

which is the convex hull of the mappings

$$f_{c,b_c}(x) = \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - x)}) \nu(d).$$

Then the matching built by **Ex-ante Balance** satisfies for all $t \in [T]$ and $c \in \mathcal{C}$, with probability at least $1 - \frac{b\alpha}{N\epsilon^2}$,

$$\left| \frac{M_c(t)}{N} - m_c(t/N) \right| \leq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}}, \quad (3)$$

The different constants in are defined by $L = \max_{c \in [C]} \sum_{d=1}^D a_{c,d} \nu(d)$, $\delta_c = \frac{1}{N} \sum_{d=1}^D \frac{a_{c,d}}{e} \nu(d)$, $K_\alpha = (c\alpha + \epsilon)e^{c\alpha}/c$, ϵ as defined in Lemma 11 and c in Lemma 7, $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d}b_c}) \nu(d)$, $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta_c + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha$.

The proof of Theorem 2 is structured as follows:

- We first characterize the drift of the process M_c .
- Next, we verify that M_c satisfies the assumptions required by Theorem 1 in [16].
- Finally, we define the associated differential inclusion and apply Theorem 4 from [16] to derive an explicit rate of convergence.

The following lemma computes the drift of the process M_c for $c \in [C]$, defined as the conditional expectation $\mathbb{E}[M_c(t+1) - M_c(t) | B, \mathbf{M}(t)]$, where $B = (b_1, \dots, b_C)$ and $\mathbf{M}(t) = (M_1(t), \dots, M_C(t))$.

Lemma 5. For $c \in [C]$,

$$\mathbb{E}[M_c(t+1) - M_c(t) | B, \mathbf{M}(t)] = H_{c,b_c,N}(M_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(M_k(t)) = H_{c,b_c,N}(M_c(t))\}}$$

where $H_{c,b_c,N}(x) = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - x}\right) \nu(d)$.

Proof. For $c \in [C]$, as defined previously, $M_c(t)$ is the number of matched nodes in the class c , for $t \in [T]$, $M_c(t)$ follows the dynamics,

$$M_c(t+1) = M_c(t) + \mathbb{1}_{\{\exists n \in \mathcal{N}_c(t) \text{ s.t. } c_t^* = c \text{ and } m_n(t+1)=1\}} \quad (42)$$

Let c_t^* denote the class selected by the **Ex-ante Balance** algorithm. We define the expected one-step change of the process $M_c(t)$, for each $c \in [C]$ and $t \in [T]$, as follows:

$$\mathbb{E}[M_c(t+1) - M_c(t) | N, \mathbf{M}(t)] \quad (43)$$

$$= \mathbb{P}(\exists n \in \mathcal{N}_c(t) \text{ s.t. } c_t^* = c \text{ and } m_n(t+1) = 1 | \mathbf{M}(t), B) \quad (44)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t) \text{ s.t. } c_t^* = c \text{ and } m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d) \nu(d) \quad (45)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t) \text{ s.t. } m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d, c_t^* = c) \nu(d) \quad (46)$$

$$= H_{c,b_c,N}(M_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(M_k(t)) = H_{c,b_c,N}(M_c(t))\}} \quad (47)$$

where $H_{c,b_c,N}(x) = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - x}\right) \nu(d)$. \square

The next lemma defines a martingale difference sequence based on the process M_c , and proves that its second moment is bounded.

Lemma 6. *Let $c \in [C]$ and $t \in [T]$. Define the process $Q_c(t+1) = M_c(t+1) - M_c(t) - \mathbb{E}[M_c(t+1) - M_c(t) | \mathbf{M}(t), B]$. Then $(Q_c(t))_{t \in [T]}$ is a martingale difference sequence with respect to the filtration generated by $\mathbf{M}(t)$. Moreover, there exists a constant $b > 0$ such that:*

$$\mathbb{E}[Q_c(t+1) | \mathbf{M}(t), B] = 0$$

and

$$\mathbb{E}[|Q_c(t+1)|^2 | \mathbf{M}(t), B] \leq b$$

Proof. By direct computation, we obtain

$$\mathbb{E}[Q_c(t+1) | \mathbf{M}(t), B] = 0.$$

Furthermore, from the definition of the matching process, for all $c \in [C]$ we have

$$|M_c(t+1) - M_c(t)| \leq 1, \quad \forall t \in [T].$$

This implies that the second moment of $Q_c(t+1)$ is bounded. \square

The next result proves the first assumption of Theorem 1 in [16].

Lemma 7. *For $c \in [C]$ and $t \in [T]$, let $\mathcal{H}_{c,b_c,N}(y) = H_{c,b_c,N}(y) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(y) = H_{c,b_c,N}(y)\}}$, it satisfies,*

$$\forall y \leq Nb_c, |\mathcal{H}_{c,b_c,N}(y)| \leq c(1 + |y|)$$

with $\alpha_c = \ln \left(1 - \frac{\min_{j \in [D]}(a_{c,j})}{N}\right)$, $c = \max(|1 - e^{\alpha_c Nb_c}|, |\alpha_c e^{\alpha_c Nb_c}|)$.

Proof.

$$|\mathcal{H}_{c,b_c,N}(y)| \leq \left| \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - y}\right) \nu(d) \right| \quad (48)$$

$$= \left| 1 - \sum_{d=1}^D e^{\ln\left(1 - \frac{a_{c,d}}{N}\right)(Nb_c - y)} \nu(d) \right| \quad (49)$$

$$\leq \left| 1 - \sum_{d=1}^D e^{\ln\left(1 - \frac{\min_j(a_{c,j})}{N}\right)(Nb_c - y)} \nu(d) \right| \quad (50)$$

$$\leq \left| 1 - e^{\ln\left(1 - \frac{\min_j(a_{c,j})}{N}\right)(Nb_c - y)} \right| \quad (51)$$

$$\leq |1 - e^{\alpha_c Nb_c}(1 + \alpha_c y)| \quad (52)$$

$$\leq |1 - e^{\alpha_c Nb_c}| + |\alpha_c e^{\alpha_c Nb_c} y| \quad (53)$$

$$\leq c(1 + |y|) \quad (54)$$

with $\alpha_c = \ln \left(1 - \frac{\min_j(a_{c,j})}{N}\right)$, $c = \max(|1 - e^{\alpha_c Nb_c}|, |\alpha_c e^{\alpha_c Nb_c}|)$. \square

The following technical lemma provides a bound on the distance between $H_{c,b_c,N}$ and its limit as N becomes large.

Lemma 8. *For $c \in [C]$,*

$$\left| \sum_{d=1}^D \left(1 - e^{-a_{c,d}(b_c - M_c(t)/N)}\right) \nu(d) - \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) \nu(d) \right| \leq \sum_{d=1}^D \frac{a_{c,d}}{N} \nu(d)$$

Proof. Let $A = \left| \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - M_c(t)/N)}) \nu(d) - \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) \nu(d) \right|$,

$$A \leq \sum_{d=1}^D \left| \left(1 - e^{-a_{c,d}(b_c - M_c(t)/N)}\right) \nu(d) - \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)}\right) \nu(d) \right| \quad (55)$$

$$\leq \sum_{d=1}^D \left| e^{-a_{c,d}(b_c - M_c(t)/N)} \nu(d) - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - M_c(t)} \nu(d) \right| \quad (56)$$

$$\leq \sum_{d=1}^D \frac{a_{c,d}}{N} \nu(d) \quad (\text{Lemma 3}) \quad (57)$$

□

The following lemma shows that the function f_{c,b_c} , defined as the limit of $H_{c,b_c,N}$ as $N \rightarrow \infty$ and given by $f_{c,b_c}(x) = \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - x)}) \nu(d)$, is L_c -Lipschitz continuous.

Lemma 9. (*Lipschitz condition*) For $c \in [C]$, the function $f_{c,b_c}(x) = \sum_{d=1}^D (1 - e^{-a_{c,d}(b_c - x)}) \nu(d)$ is Lipschitz continuous with constant $L_c = \sum_{d=1}^D a_{c,d} \nu(d)$.

Proof. Let x, y such that $x \leq b_c$ and $y \leq b_c$ for all $c \in [C]$,

$$|f_{c,b_c}(x) - f_{c,b_c}(y)| = \left| \sum_{d=1}^D (e^{-a_{c,d}(b_c - y)} - e^{-a_{c,d}(b_c - x)}) \nu(d) \right| \quad (58)$$

By mean value Theorem, there exists $\xi \in (x, y)$ such that we have,

$$|f_{c,b_c}(x) - f_{c,b_c}(y)| \leq |x - y| \sum_{d=1}^D e^{-a_{c,d}(b_c - \xi)} a_{c,d} \nu(d) \quad (59)$$

$$\leq |x - y| \sum_{d=1}^D a_{c,d} \nu(d) \quad (60)$$

□

The two following lemma provides a technical bound essential for deriving the explicit rate of convergence in Theorem 2. For $t \in [T]$ let,

$$V_c(t) = \frac{1}{N} \sum_{k=0}^t Q_c(k+1) \quad (61)$$

and

$$\frac{M_c(t+1)}{N} = \frac{M_c(0)}{N} + \frac{1}{N} \sum_{l=0}^t \mathcal{H}_{c,b_c,N}(M_c(t)) + \frac{1}{N} \sum_{l=0}^t Q_c(l+1) \quad (62)$$

$$= \frac{1}{N} \sum_{l=0}^t \mathcal{H}_{c,b_c,N}(M_c(t)) + \frac{1}{N} \sum_{l=0}^t Q_c(l+1) \quad (63)$$

Lemma 10. For all $T, N > 0$, and for all $\epsilon > 0$,

$$\mathbb{P} \left(\sup_{0 \leq k \leq T} |V_c(k)| \geq \epsilon \right) \leq \frac{Tb}{N^2 \epsilon^2}$$

Proof. Since $\mathbb{E}[Q_c(t+1)|\mathbf{M}(t), B] = 0$ and $\mathbb{E}[|Q_c(t+1)|^2|\mathbf{M}(t), B] \leq b$, we have $\mathbb{E}[|V_c(t)|^2] \leq \frac{tb}{N^2} \leq \frac{Tb}{N^2}$ for all $t \leq T$. Applying Kolmogorov's inequality (maximal inequality) for martingales to the martingale V leads to the bound of the lemma. □

Lemma 11. For $c \in [C]$, let M_c be defined as in Equation (62) with $|\mathcal{H}_{c,b_c,N}(y)| \leq c(1 + |y|)$. Let y denote the solution of the differential equation associated with F that is defined in Equation (68).

If we denote by $\epsilon := \sup_{l \leq t} |V_c(l)|$, then

$$\max \left\{ \sup_{0 \leq t \leq T} |M_c(t)|, \sum_{0 \leq \tau \leq T/n} |m(\tau)| \right\} \leq K_\alpha$$

with $K_\alpha = (c\alpha + \epsilon) e^{c\alpha}/c$ with c as defined in Lemma 7.

Proof. By definition of M_c in Equation (62) and Lemma 10, we have,

$$|M_c(t+1)/N| \leq \frac{1}{N} \sum_{l=0}^t c(1 + |M_c(l)|) + \epsilon \quad (64)$$

$$= \frac{tc}{N} + \epsilon + \frac{c}{N} \sum_{l=1}^t |M_c(l)| \quad (65)$$

$$\leq \left(\frac{Tc}{N} + \epsilon \right) e^{cT/N} / c \quad (66)$$

The final inequality follows from the discrete Gronwall's lemma. Substituting $T = \alpha N$ then yields the desired result. \square

The next lemma proves that F defined in Theorem 2 is upper semicontinuous,

Lemma 12. (Upper semicontinuous) Let $f_{c,b_c} : \mathbb{R} \rightarrow \mathbb{R}$ be continuous for each $c \in [C]$, and define the set-valued map $F(m) = \text{conv} (f_{c,b_c}(m_c)e_c \mid c \in \arg \max_{j \in [C]} f_{j,b_j}(m_j))$ where e_c is the c -th standard basis vector in \mathbb{R}^C . Then F is upper semicontinuous as a set-valued map from \mathbb{R}^C to subsets of \mathbb{R}^C .

Proof. Let $(m^{(n)})_{n \in \mathbb{N}} \subset \mathbb{R}^C$ be a sequence such that $m^{(n)} \rightarrow m$ as $n \rightarrow \infty$, meaning:

$$\forall c \in [C], \quad m_c^{(n)} \rightarrow m_c.$$

Let $x_n \in F(m^{(n)})$ be such that $x_n \rightarrow x \in \mathbb{R}^C$. We aim to show that $x \in F(m)$.

Each $x_n \in F(m^{(n)})$ belongs to the convex hull:

$$F(m^{(n)}) = \text{conv} \left(f_{c,b_c}(m_c^{(n)})e_c \mid c \in \arg \max_{j \in [C]} f_{j,b_j}(m_j^{(n)}) \right).$$

Because the index set $[C]$ is finite, there are only finitely many possible argmax sets. Thus, we may extract a subsequence (still denoted by n for simplicity) such that for some fixed $A \subseteq [C]$,

$$\arg \max_{j \in [C]} f_{j,b_j}(m_j^{(n)}) = A \quad \text{for all large } n.$$

Since each f_{j,b_j} is continuous and $m_j^{(n)} \rightarrow m_j$, we have $f_{j,b_j}(m_j^{(n)}) \rightarrow f_{j,b_j}(m_j)$ for all j , and hence:

$$\max_{j \in [C]} f_{j,b_j}(m_j^{(n)}) \rightarrow \max_{j \in [C]} f_{j,b_j}(m_j).$$

Therefore, for every $i \in A$,

$$f_{c,b_c}(m_c) = \max_{j \in [C]} f_{j,b_j}(m_j),$$

i.e., $A \subseteq \arg \max_{j \in [C]} f_{j,b_j}(m_j)$.

Now, each x_n is a convex combination of the vectors $f_{c,b_c}(m_c^{(n)})e_c$ with $c \in A$, and by continuity:

$$f_{c,b_c}(m_c^{(n)}) \rightarrow f_{c,b_c}(m_c), \quad \text{so} \quad f_{c,b_c}(m_c^{(n)})e_c \rightarrow f_{c,b_c}(m_c)e_c.$$

Thus, $x_n \rightarrow x$ implies that x lies in the convex hull of the limit points:

$$x \in \text{conv} (f_{c,b_c}(m_c)e_c \mid c \in A) \subseteq F(m),$$

since $A \subseteq \arg \max_{j \in [C]} f_{j,b_j}(m_j)$.

Hence, every limit point of a converging sequence (x_n) with $x_n \in F(m^{(n)})$ lies in $F(m)$, which proves that F is upper semicontinuous. \square

With all the preparatory results established — in particular, Lemmas 6 and 7, which shows that $M_c(t)$ satisfies the assumptions of Theorem 1 in [16] — we are now ready to prove Theorem 2.

Proof. For all $c \in [C]$ and $\tau \in [T/n]$, we consider the process $\tilde{M}_c(\tau)$ defined by,

$$\tilde{M}_c(\tau + 1/n) = \tilde{M}_c(\tau) + \frac{1}{N} \mathcal{H}_{c,b_c,N}(\tilde{M}_c(\tau)) + \tilde{Q}_c(\tau + 1/n) \quad (67)$$

where $\tilde{Q}_c(\tau) = \frac{Q_c(\tau n)}{n}$ with Q_c as defined in Lemma 6, $\mathcal{H}_{c,b_c,N}$ as defined in Lemma 7, and let $\tilde{\mathbf{M}}(\tau) = (\tilde{M}_1(\tau), \dots, \tilde{M}_C(\tau))$.

When $N \rightarrow \infty$, and $t \in [T]$, the function $H_{c,b_c,N}$ converges to:

$$H_{c,b_c,N}(M_c(t)) \rightarrow \sum_{d=1}^D \left(1 - e^{-a_{c,d}(b_c - \frac{M_c(t)}{N})}\right) \nu(d) = f_{c,b_c}(M_c(t)/N).$$

Let $g(\tilde{\mathbf{M}}) = \left(f_{c,b_c}(\tilde{M}_c) \mathbf{1}_{\{\max_{k \in [C]} f_{k,b_k}(\tilde{M}_k) = f_{c,b_c}(\tilde{M}_c)\}} \right)_{c \in [C]}$ be the drift vector. According to Lemma 7, each element of the drift vector satisfies the first assumption of Theorem 1 in [16]. Moreover, according to Lemma 6, $\tilde{Q}_c(\tau + 1/n)$ satisfies the second assumption of Theorem 1 in [16]. Since $\tilde{M}_c(0) = 0$, based on Theorem 1 in [16], for all $\tau \in [T/n]$, $\tilde{\mathbf{M}}(\tau)$ converges to $m(\tau)$, where m is the solution of the following differential inclusion:

$$\dot{m}(\tau) \in F(m(\tau)) \quad (68)$$

where $F(m) = \text{conv} (f_{c,b_c}(m_c)e_c \mid c \in \arg \max_{j \in [C]} f_{j,b_j}(m_j))$, and conv denotes the convex hull.

According to Lemma 12 and Lemma 7, the differential inclusion Equation (68) admits at least one solution m . To establish the uniqueness of this solution, it is sufficient to show that the set-valued map F is one-sided Lipschitz. By Lemma 9, each function f_{c,b_c} is L_c -Lipschitz continuous for all $c \in [C]$, and let $L = \max_{c \in [C]} L_c$. Let $s, s' \in \mathbb{R}^C$ and suppose $z \in F(s)$, $z' \in F(s')$. Then:

$$\langle z - z', s - s' \rangle = \sum_{i=1}^C (s_i - s'_i) (f_{i,b_i}(s_i) - f_{i,b_i}(s'_i)) \leq \sum_{i=1}^C L_i (s_i - s'_i)^2 \leq L \|s - s'\|^2. \quad (69)$$

Thus F is one-sided Lipschitz with constant L , which guarantees the uniqueness of the solution m to the differential inclusion Equation (68). To get the explicit rate of the convergence of \tilde{M}_c to m_c , we use Theorem 4 of [16]. According to Lemma 6, for $t \in [T]$, $\mathbb{E}[|Q_c(t+1)|^2 | \mathbf{M}(t), B] \leq b$ and F is one sided Lipschitz with constant $L = \max_{c \in [C]} L_c$ where L_c is defined in Lemma 9. Thus according to theorem 4 in [16], taking $\delta_c = \sum_{d=1}^D \frac{a_{c,d}}{N e} \nu(d)$, K_α as defined in Lemma 11 and $U_c = \sup_{0 \leq t \leq T} f_{c,b_c}(M_c(t))$, taking $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d} b_c}) \nu(d)$ we define $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha$. We have,

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \frac{M_c(t)}{N} - m_c(t/N) \right| \geq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \right) \leq \frac{bT}{N^2 \epsilon^2} \quad (70)$$

\square

C.2 Proof of Theorem 3

Theorem 3. For any $c \in \mathcal{C}$, the function

$$m_c^* : t \mapsto (b_c - \beta_c^{(k_t)}) + \left(f_{c, \beta^{(k_t)}}^{-1} (F_{k_t, \beta^{(k_t)}} (\mu_{k_t, \beta^{(k_t)}} (t - t_{k_t}))) \right)_+ \quad (6)$$

where $k_t = \max\{k \in [C] : t > t_k\}$, is the solution of the differential inclusion defined in Theorem 2.

We need a few lemmas before going into the proof of the Theorem 3.

Lemma 13. $\forall t \in \mathbb{R}_+, \forall c \in \mathcal{C}$,

$$\mu_c(t) = \begin{cases} b_c - \beta_c^{(k_t)} + f_{c, \beta^{(k_t)}}^{-1} (F_{k_t, \beta^{(k_t)}} (\mu_{k_t, \beta^{(k_t)}} (t - t_{k_t}))) & \text{if } c \leq k_t \\ 0 & \text{otherwise} \end{cases} \quad (71)$$

Proof. For $t \geq 0$ and $c > k_t$, $(b_c - \beta_c^{(k_t)}) = 0$ by def of $\beta_c^{(k_t)}$. Then

By definition of k_t , $t \in [t_{k_t}, t_{k_t+1})$.

$$t < t_{k_t+1} \Leftrightarrow t < t_{k_t} + \mu_{k_t, \beta^{(k_t)}}^{-1} \left(F_{k_t+1, \beta^{(k_t)}}^{-1} (f_{k_t+1, \beta^{(k_t)}}(0)) \right) \quad (72)$$

$$\Leftrightarrow \mu_{k_t, \beta^{(k_t)}}(t - t_{k_t}) < F_{k_t, \beta^{(k_t)}}^{-1} (f_{k_t+1, \beta^{(k_t)}}(0)) \quad (\mu_{k_t, \beta^{(k_t)}} \text{ is increasing}) \quad (73)$$

$$\Leftrightarrow F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})) > f_{k_t+1, \beta^{(k_t)}}(0) \quad (F_{k_t, \beta^{(k_t)}} \text{ is decreasing}) \quad (74)$$

$$\Leftrightarrow \forall c > k_t, F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})) > f_{c, \beta^{(k_t)}}(0) \quad (F_{k_t, \beta^{(k_t)}} \text{ is decreasing}) \quad (75)$$

$$\Leftrightarrow \forall c > k_t, f_{c, \beta^{(k_t)}}^{-1}(F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t}))) < 0 \quad (f_{c, \beta^{(k_t)}}^{-1} \text{ is decreasing}) \quad (76)$$

□

Lemma 14. $\forall t \in \mathbb{R}_+, \sum_{c \in \mathcal{C}} \mu_c(t) = \|b - \beta^{(k_t)}\|_1 + \mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})$

Proof. By definition, for any $t \geq 0$, $\beta_c^{(k_t)} \leq b_c$, so $\sum_{c \in \mathcal{C}} (b_c - \beta_c^{(k_t)}) = \|b - \beta^{(k_t)}\|_1$. Let $k \in [C]$ be a phase.

$$\forall t \in [t_k, t_{k+1}), \sum_{c \in \mathcal{C}} \left(f_{c, \beta^{(k)}}^{-1} (F_{k, \beta^{(k)}} (\mu_{k, \beta^{(k)}} (t - t_k))) \right)_+ = \mu_{k, \beta^{(k)}}(t - t_k) \quad (77)$$

$$\Leftrightarrow \forall t \in [0, t_{k+1} - t_k], \sum_{c \leq k} f_{c, \beta^{(k)}}^{-1} (F_{k, \beta^{(k)}} (\mu_{k, \beta^{(k)}} (t))) = \mu_{k, \beta^{(k)}}(t) \quad (78)$$

$$\Leftrightarrow \forall s \in [\mu_{k, \beta^{(k)}}^{-1}(0), \mu_{k, \beta^{(k)}}^{-1}(t_{k+1} - t_k)], \sum_{c \leq k} f_{c, \beta^{(k)}}^{-1} (F_{k, \beta^{(k)}} (s)) = s \quad (79)$$

$$\Leftrightarrow \forall u \in [F_{k, \beta^{(k)}}^{-1}(\mu_{k, \beta^{(k)}}^{-1}(t_{k+1} - t_k)), F_{k, \beta^{(k)}}^{-1}(\mu_{k, \beta^{(k)}}^{-1}(0))], \sum_{c \leq k} f_{c, \beta^{(k)}}^{-1}(u) = F_{k, \beta^{(k)}}^{-1}(u) \quad (80)$$

The last line is TRUE by definition of $F_{k, \beta^{(k)}}$. □

Lemma 15. $\forall t \in \mathbb{R}_+, \forall c \in \mathcal{C}$

$$f_c(\mu_c(t)) = \begin{cases} F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})) & \text{if } c \leq k_t \\ f_{c, N b_c}(0) & \text{otherwise} \end{cases} \quad (81)$$

and $f_c(\mu_c(t)) \leq f_{k_t}(\mu_{k_t}(t))$.

Proof. For any $t \in \mathbb{R}_+$ and $c \in \mathcal{C}$,

$$f_c(\mu_c(t)) = f_{c, b_c}(\mu_c(t)) \quad (82)$$

$$= f_{c, \beta_c^{(k_t)}} \left((\beta_c^{(k_t)} - b_c + \mu_c(t)) \right) \quad (83)$$

$$= f_{c, \beta_c^{(k_t)}} \left(\left(f_{c, \beta^{(k_t)}}^{-1} (F_{k_t, \beta^{(k_t)}} (\mu_{k_t, \beta^{(k_t)}} (t - t_{k_t}))) \right)_+ \right) \quad (84)$$

Lemma 13 allows to conclude. □

With all the preparatory lemmas established, we now proceed to the proof.

Proof. The result is proven by induction. For $k \in [C]$, the induction hypothesis is

(A_k) $\forall t \leq t_k$, μ is such that $\dot{\mu}(t) \in F(\mu(t))$.

For $t \in [t_k, t_{k+1})$, we have $k = k_t$. Thus, restricted to the interval $[t_k, t_{k+1})$,

$$\dot{\mu} \in F(\mu) \Leftrightarrow \dot{\mu} \in \text{conv} \left(f_c(\mu_c) e_c | c \in \arg \max_{k \in [C]} f_k(\mu_k) \right) \quad (85)$$

$$\Leftrightarrow \dot{\mu} \in F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})) \text{conv} \left(e_c | c \in \arg \max_{k \in [C]} f_k(\mu_k) \right) \quad (86)$$

$$\Leftrightarrow \left(\sum_{c \leq k} \dot{\mu}_c \leq F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})) \right) \text{ and } (\forall c > k, \dot{\mu}_c = 0) \quad (87)$$

$$\Leftrightarrow (\dot{\mu}_{k_t, \beta^{(k_t)}}(t - t_{k_t}) \leq F_{k_t, \beta^{(k_t)}}(\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t}))) \text{ and } (\forall c > k, \dot{\mu}_c = 0) \quad (88)$$

Going from (85) to (86) comes from Lemma 15. Going from (86) to (87) comes from the fact that the convex hull of a set of basis vectors is the L_1 ball restricted to the corresponding subspace. Going from (87) to (88) comes from applying Lemma 14. Equation (88) is TRUE by definition of $\mu_{k_t, \beta^{(k_t)}}(t - t_{k_t})$ for the first term and by applying Lemma 13 for the second term. \square

C.3 Ex-post Balance

While the Ex-ante Balance policy improves upon the purely compatibility-driven Myopic approach by considering the probability of successful matches, it still relies on an expected availability model—it selects the class that likely has unmatched nodes, without verifying their presence. This can lead to wasted opportunities when the chosen class ends up being empty. To address this limitation, we introduce the Ex-post Balance policy, which takes an additional, more grounded step. Instead of only maximizing the expected match probability, Ex-post Balance ensures that the selected class actually contains at least one available node at the time of decision. This additional check prevents the algorithm from targeting unavailable options, making it more efficient. The formal procedure is detailed in Algorithm 6.

Algorithm 6: Ex-post Balance

Output: Updated matching $M(t)$

```

1 for  $t \in [T]$  do
2   Choose  $c_t = \arg \max_{c \in C(t)} \sum_{d=1}^D (1 - (1 - \frac{a_{c,d}}{N})^{Nb_c - M_c(t)}) \nu(d)$ 
3   if  $\mathcal{F}_{c_t}(t) = \emptyset$  then
4      $M(t) = M(t-1)$ 
5   else
6      $M(t) = M(t-1) \cup \{(n_t, t)\}$  for  $n_t \sim \text{unif}(\mathcal{F}_{c_t}(t))$ 

```

The Ex-post Balance policy introduces an additional layer of selectivity compared to Ex-ante Balance by explicitly verifying the presence of available nodes in the selected class before committing to a match. While this refinement improves matching efficiency, it also increases the non-smoothness of the system's dynamics. In particular, the policy induces more discontinuities—not only due to abrupt switches in the maximization rule, but also because the feasibility of a class now depends on the cardinality of its unmatched nodes at each step. As a result, the evolution of the matching process under Ex-post Balance cannot be captured by the same differential inclusion used for Ex-ante Balance. Instead, it follows a more constrained inclusion, where the feasible directions of evolution depend explicitly on whether a class still has unmatched capacity. The following theorem formalizes this behavior, showing that, with high probability, the normalized matching sizes converge to the solution of a new differential inclusion that incorporates both compatibility and real-time availability constraints.

Theorem 5. For all $c \in [C]$, the matching size built by *Ex-post Balance* satisfies for all $t \in [T]$ with probability $1 - \frac{b\alpha}{N\epsilon}$,

$$\left| \frac{M_c(t)}{N} - m_c(t/N) \right| \leq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \quad (89)$$

where m_c is the c -th coordinate of m the unique solution of the differential inclusion $\dot{m} \in G(m)$ and $G(m) = \text{conv} \left(f_{c,b_c}(m_c) e_c | c = \arg \max_{k \in [1,C]} f_{k,b_k}(m_k), b_c > m_c \right)$ with e_c a basis vector of \mathbb{R}^C , $L = \max_{c \in [1,C]} \sum_{d=1}^D a_{c,d} \nu(d)$, $\delta_c = \sum_{d=1}^D \frac{a_{c,d}}{N\epsilon} \nu(d)$, $K_\alpha = (c\alpha + \epsilon)e^{c\alpha}/c$ with ϵ as defined in Lemma 11 and c in Lemma 7, $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d} b_c}) \nu(d)$, $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta_c + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha$.

As in Appendix C.1, our objective here is to approximate the matching size M_c produced by *Ex-post Balance* for each class $c \in [C]$. Following the same approach as before, the first step involves computing the drift of the process M_c .

Lemma 16. For $c \in [C]$,

$$\mathbb{E}[M_c(t+1) - M_c(t) | B, \mathbf{M}(t)] = H_{c,b_c,N}(M_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(M_k(t)) = H_{c,b_c,N}(M_c(t))\}} \mathbb{1}_{\{Nb_c > M_c(t)\}}$$

Proof.

$$\mathbb{E}[M_c(t+1) - M_c(t) | B, \mathbf{M}(t)] = \mathbb{P}(\exists n \in \mathcal{N}_c(t) \setminus \mathcal{M}_c(t), c^* = c, m_n(t+1) = 1 | B, \mathbf{M}(t)) \quad (90)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t) \setminus \mathcal{M}_c(t), c^* = c, m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d) \nu(d) \quad (91)$$

$$= \sum_{d=1}^D \mathbb{P}(\exists n \in \mathcal{N}_c(t) \setminus \mathcal{M}_c(t), m_n(t+1) = 1 | \mathbf{M}(t), B, d(t+1) = d, c^* = c) \nu(d) \quad (92)$$

$$= H_{c,b_c,N}(M_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(M_k(t)) = H_{c,b_c,N}(M_c(t))\}} \mathbb{1}_{\{Nb_c > M_c(t)\}} \quad (93)$$

where $H_{c,b_c,N}(x) = \sum_{d=1}^D \left(1 - \left(1 - \frac{a_{c,d}}{N}\right)^{Nb_c - x}\right) \nu(d)$. \square

Proof. For all $c \in [C]$ and $\tau \in [T/N]$, we consider the process $\tilde{M}_c(\tau)$ defined by,

$$\tilde{M}_c(\tau + 1/N) = \tilde{M}_c(\tau) + \frac{1}{N} H_{i,N}(\tilde{M}_c(\tau)) + \tilde{Q}_c(\tau + 1/N)$$

where $\tilde{Q}_c(\tau) = \frac{Q_c(\tau N)}{N}$, $Q_c(t+1) = M_c(t+1) - M_c(t) - \mathbb{E}[M_c(t+1) - M_c(t) | \mathbf{M}(t), B]$, $H_{c,b_c,N}(y) = H_{c,b_c,N}(M_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(M_k(t)) = H_{c,b_c,N}(M_c(t))\}} \mathbb{1}_{\{Nb_c > M_c(t)\}}$, and let $\tilde{\mathbf{M}}(\tau) = (\tilde{M}_1(\tau), \dots, \tilde{M}_C(\tau))$.

When $N \rightarrow \infty$, and $t \in [T]$, the function $H_{c,b_c,N}(M_c(t))$ converges to:

$$H_{c,b_c,N}(M_c(t)) \rightarrow \sum_{d=1}^D \left(1 - e^{-a_{c,d}(b_c - \frac{M_c(t)}{N})}\right) \nu(d) = f_{c,b_c}(M_c(t)/N).$$

Let $g(\tilde{M}) = \left(f_{c,b_c}(\tilde{M}_c) \mathbb{1}_{\{\max_{k \in [C]} f_{k,b_k}(\tilde{M}_k) = f_{c,b_c}(\tilde{M}_c)\}} \mathbb{1}_{\{b_c > \tilde{M}_c\}} \right)_{i \in [C]}$ be the drift vector. According

to Lemma 7, each element of the drift vector satisfies the first assumption of Theorem 1 in [16]. Moreover, according to Lemma 6, $\tilde{Q}_c(\tau + 1/n)$ satisfies the second assumption of Theorem 1 in [16]. Since $\tilde{M}_c(0) = 0$, based on Theorem 1 in [16], for all $\tau \in [T/N]$, $\tilde{M}(\tau)$ converges to $m(\tau)$, where m is the solution of the following differential inclusion:

$$\dot{m}(\tau) \in G(m(\tau))$$

where $G(m) = \text{conv} (f_{c,b_c}(m_c)e_c \mid c \in \arg \max_{j \in [C]} f_{j,b_j}(m_j), b_c > m_c)$, and conv denotes the convex hull.

Following the same results as in Appendix C.1, one can prove that G is upper semicontinuous and L -one-sided Lipschitz, where $L = \max_{c \in [C]} L_c$. This ensures uniqueness of the solution m . To get the explicit rate of the convergence of \tilde{M}_c to m_c , we use Theorem 4 of [16]. According to Lemma 6, for $t \in [T]$, $\mathbb{E}[|Q_c(t+1)|^2 | \mathbf{M}(t), B] \leq b$ and F is one sided Lipschitz with constant $L = \max_{c \in [C]} L_c$ where L_c is defined in Lemma 9. Thus according to theorem 4 in [16], taking $\delta_c = \sum_{d=1}^D \frac{a_{c,d}}{Ne} \nu(d)$, K_α as defined in Lemma 11 and $U_c = \sup_{0 \leq t \leq T} f_{c,d_c}(M_c(t))$, taking $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d} b_c}) \nu(d)$ we define $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha$. We have,

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} \left| \frac{M_c(t)}{N} - m_c(t/N) \right| \geq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \right) \leq \frac{bT}{N^2 \epsilon} \quad (94)$$

□

D Proof of Theorem 4

In this section, we provide the proof of Theorem 4.

Theorem 4. *Let $R(T) = \sum_{i \in \mathcal{C}} M_i(T) - \hat{M}_i(T)$ denote the regret of ETC – balance. Suppose the exploration phase lasts for $T_{\text{explore}} = T^{\frac{q+3}{4}}$, for some $0 < q < 1$. Then the regret satisfies*

$$R(T) = \mathcal{O}(T^{\frac{q+3}{4}}).$$

The proof of Theorem 4 is structured in two main steps:

- We begin by establishing a concentration result for an estimator of $(1 - \frac{a_{c,d}}{N})^{Nb_c - M_c(t)}$.
- Next, we decompose the regret and derive bounds for each resulting term.

Let $m', m \in [0, Nb_c)$ and $\mathcal{V}_m = \{m' \in [0, Nb_c) \mid 1/2 \leq \frac{Nb_c - m'}{Nb_c - m} \leq 2\}$, we consider a Bernoulli random variable $Y_{c,d,m'}(t)$ with parameter $1 - D_{c,d}(m') := 1 - (1 - a_{c,d}/N)^{Nb_c - m'}$, whenever $c(t) = c$, $d(t) = d$, and $M_c(t) = m'$. Let the number of observations defined as $T_{c,d,m'} := \sum_{t=1}^T \mathbb{1}_{\{c(t)=c, d(t)=d, M_c(t)=m'\}}$ and $T_{\text{total}} = \sum_{m' \in \mathcal{V}_m} T_{c,d,m'}$. We consider the following estimator:

$$\Theta(m) := \frac{1}{T_{\text{total}}} \sum_{m' \in \mathcal{V}_m} \sum_{t=1}^T \mathbb{1}_{\{c(t)=c, d(t)=d, M_c(t)=m'\}} (1 - Y_{c,d,m'}(t)). \quad (95)$$

$$\mathbb{E}[\Theta(m)] = \frac{1}{T_{\text{total}}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} D_{c,d}(m') \quad (96)$$

$$\mathbb{E}[\Theta(m)] = \frac{1}{T_{\text{total}}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} D_{c,d}(m) \frac{Nb_c - m'}{Nb_c - m} = g(D_{c,d}(m)) \quad (97)$$

with $D_{c,d}(m) = (1 - a_{c,d}/N)^{Nb_c - m}$. The next lemma shows that g^{-1} is Lipschitz continuous.

Lemma 17. g^{-1} is $2e^a$ Lipschitz.

Proof. Let

$$g(x) = \frac{1}{T_{\text{total}}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} x^{\frac{Nb_c - m'}{Nb_c - m}},$$

defined on the interval $x \in [(1 - \frac{a}{N})^{Nb_c}, 1]$, where $a_{c,d} \leq a < N$. The function g is continuously differentiable and strictly increasing on this interval, as it is a finite sum of positive-coefficient power functions. Hence, g is invertible, and its inverse is differentiable with

$$\frac{dg^{-1}}{dy}(y) = \frac{1}{g'(g^{-1}(y))}.$$

We compute

$$g'(x) = \frac{1}{T_{total}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} \cdot \frac{Nb_c - m'}{Nb_c - m} \cdot x^{\frac{Nb_c - m'}{Nb_c - m} - 1},$$

so that

$$\frac{dg^{-1}}{dy}(y) = \frac{1}{\frac{1}{T_{total}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} \cdot \frac{Nb_c - m'}{Nb_c - m} \cdot (g^{-1}(y))^{\frac{Nb_c - m'}{Nb_c - m} - 1}}.$$

To bound this derivative, we lower bound $g'(x)$ over the domain. Let

$$\Delta := \min \left\{ 1, \left(1 - \frac{a}{N} \right)^{Nb_c \cdot \max_{m' \in \mathcal{V}_m} \left(\frac{Nb_c - m'}{Nb_c - m} - 1 \right)} \right\},$$

then for all $x \in \left[\left(1 - \frac{a}{N} \right)^{Nb_c}, 1 \right]$,

$$g'(x) \geq \frac{\Delta}{T_{total}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} \cdot \frac{Nb_c - m'}{Nb_c - m}.$$

This yields

$$\frac{dg^{-1}}{dy}(y) \leq \frac{1}{\frac{\Delta}{T_{total}} \sum_{m' \in \mathcal{V}_m} T_{c,d,m'} \cdot \frac{Nb_c - m'}{Nb_c - m}}.$$

Using the assumption $\mathcal{V}_m = \left\{ m' \in [0, Nb_c) \mid \frac{1}{2} \leq \frac{Nb_c - m'}{Nb_c - m} \leq 2 \right\}$ we get,

$$\frac{dg^{-1}}{dy}(y) \leq \frac{\left(1 - \frac{a}{N} \right)^{-Nb_c}}{1/2} = 2 \left(1 - \frac{a}{N} \right)^{-Nb_c} \leq 2e^a,$$

So g^{-1} is $2e^a$ -Lipschitz. □

The next result establishes a concentration result on an estimator of $\left(1 - \frac{a_{c,d}}{N} \right)^{Nb_c - M_c(t)}$.

Lemma 18. *With probability $1 - \delta$, $\hat{D}_{c,d}(m) = g^{-1}(\Theta(m))$ satisfies,*

$$|\hat{D}_{c,d}(m) - D_{c,d}(m)| \leq 2e^a \sqrt{\frac{\log(2/\delta)}{2T_{total}}} \quad (98)$$

Proof. Let $\Theta(m) = \frac{1}{T_{total}} \sum_{m' \in \mathcal{V}_m} \sum_{t=1}^T \mathbb{1}_{\{c(t)=c, d(t)=d, M_c(t)=m'\}} (1 - Y_{c,d,m'}(t))$, it is a sum of independent Bernoulli random variables, then using Hoeffding inequality, with probability $1 - \delta$, it satisfies,

$$|\Theta(m) - \mathbb{E}(\Theta(m))| \leq \sqrt{\frac{\log(2/\delta)}{2T_{total}}}$$

As proved in Lemma 17, g^{-1} is $2e^a$ Lipschitz, $\hat{D}_{c,d}(m) = g^{-1}(\Theta(m))$ satisfies with probability $1 - \delta$,

$$|\hat{D}_{c,d}(m) - D_{c,d}(m)| \leq 2e^a \sqrt{\frac{\log(2/\delta)}{2T_{total}}} \quad (99)$$

□

Let $R(T)$ be the cumulative regret defined by,

$$\begin{aligned} R(T) &= \sum_{c \in \mathcal{C}} M_c(T) - \hat{M}_c(T) \\ &= R_{\text{explore}}(T_{\text{explore}}) + R_{\text{exploit}}(T_{\text{exploit}}) \end{aligned}$$

We consider $T_{\text{explore}} = T^\omega$ with $0 < \omega < 1$ and $T_{\text{exploit}} = T - T_{\text{explore}}$, and M_c is the matching size in the class c created by **Ex-ante Balance** and \hat{M}_c is the matching size in the class c created by **ETC – balance**. Since **ETC – balance** may not be making optimal decision during exploration, we can bound $R_{\text{explore}}(T_{\text{explore}})$, as $R_{\text{explore}}(T_{\text{explore}}) \leq CT^\omega$. For exploitation phase, let m_c be the solution of differential inclusion defined in Theorem 2. $R_{\text{exploit}}(T_{\text{exploit}})$ satisfies,

$$R_{\text{exploit}}(T_{\text{exploit}}) = \sum_{c \in \mathcal{C}} \underbrace{\sum_{t=T_{\text{explore}}}^{T-1} \mathbb{1}_{\{\exists n \in \mathcal{N}_c(t), c^*=c, m_n(t+1)=1\}}}_{M_c^{\text{exploit}}} - \sum_{t=T_{\text{explore}}}^{T-1} \underbrace{\mathbb{1}_{\{\exists n \in \mathcal{N}_c(t), d^*=c, m_n(t+1)=1\}}}_{\hat{M}_c^{\text{exploit}}} \quad (100)$$

c^* is the class chosen by **Ex-ante Balance** and d^* is the class chosen by **ETC – balance**.

$$|R_{\text{exploit}}(T_{\text{exploit}})| \quad (101)$$

$$\leq \sum_{c \in \mathcal{C}} |M_c^{\text{exploit}} - Nm_c(T/N) + Nm_c(T/N) - \hat{M}_c^{\text{exploit}}| \quad (102)$$

$$\leq \sum_{c \in \mathcal{C}} |M_c(T) - M_c(T_{\text{explore}}) - Nm_c(T/N) + Nm_c(T/N) - \hat{M}_c(T) + \hat{M}_c(T_{\text{explore}})| \quad (103)$$

$$\leq \sum_{c \in \mathcal{C}} \underbrace{|M_c(T) - Nm_c(T/N)|}_{w_1} + \underbrace{|\hat{M}_c(T_{\text{explore}}) - M_c(T_{\text{explore}})|}_{w_2} + \underbrace{|Nm_c(T/N) - \hat{M}_c(T)|}_{w_3} \quad (104)$$

To bound w_2 , we use the previous bound of CT^ω . To bound w_1 , we leverage the result from Theorem 2. The next lemma, built upon Theorem 2, quantifies the discrepancy between M_c , the matching size produced by the **Ex-ante Balance** algorithm, and m_c , the solution to the differential inclusion described therein.

Lemma 19. For $c \in \mathcal{C}$ and $0 < q < 1$,

$$|M_c(T) - Nm_c(T/N)| \sim \mathcal{O}(T^{\frac{q+3}{4}})$$

Proof. From Theorem 2, with probability $1 - \frac{b\alpha}{N\epsilon^2}$,

$$\left| \frac{M_c(T)}{N} - m_c(T/N) \right| \quad (105)$$

$$\leq \min(\alpha, e^{L\alpha}/\sqrt{2L}) \sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \quad (106)$$

Taking $\epsilon^2 = \frac{1}{N^{1-q}}$ with $0 < q < 1$, $\delta_c = \sum_{d=1}^D \frac{a_{i,j}}{Ne} \nu(d)$, $K_\alpha = (c\alpha + \epsilon)e^{c\alpha}/c$ with c defined in Lemma 7, $U_c = \sum_{d=1}^D (1 - e^{-a_{c,d}b_c})\nu(d) \leq 1$, $A_{\alpha,c} = U_c(U_c^2 + \frac{14U_c}{3} + 2K_\alpha) \leq (\frac{17}{3} + 2K_\alpha)$, $B_{\alpha,c} = 2U_c^2 + 4L\delta_c + 12K_\alpha \leq 2 + 4L\delta_c + 12K_\alpha$, $C_{\alpha,c} = 2U_c^2 + 4L\epsilon + 8K_\alpha \leq 2 + 4L\epsilon + 8K_\alpha$.

$$\sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \leq \sqrt{\frac{(\frac{17}{3} + 2K_\alpha)}{N} + \delta_c(2 + 4L\delta_c + 12K_\alpha) + \epsilon(2 + 4L\epsilon + 8K_\alpha)} \quad (107)$$

$$\leq \sqrt{\frac{17/3 + \alpha e^{c\alpha} + 2}{N} + \frac{e^{c\alpha}/c}{N^{\frac{3-q}{2}}} + \frac{4L}{N^2} + \frac{2 + \alpha e^{c\alpha}}{N^{\frac{1-q}{2}}} + \frac{4L + e^{c\alpha}/c}{N^{1-q}}} \quad (108)$$

Thus $\sqrt{A_{\alpha,c}/N + \delta_c B_{\alpha,c} + \epsilon C_{\alpha,c}} \sim \mathcal{O}(\frac{1}{N^{\frac{1-q}{4}}})$. Since $T = \alpha N$ with $\alpha > 1$, this leads to $|M_c(T) - m_c(T/N)| \sim \mathcal{O}(T^{\frac{q+3}{4}})$. \square

The next lemma is for bounding w_3 ,

Lemma 20.

$$|Nm_c(T/N) - \hat{M}_c(T)| \sim \mathcal{O}(T^{\frac{q+3}{4}})$$

Proof. The goal here is to apply the differential inclusion approximation, to bound $|\hat{M}_c(T_{exploit}) - Nm_c(T_{exploit}/N)|$. As defined previously, \hat{M}_c is the matching size in the class $c \in \mathcal{C}$, built by a policy that considers the estimator $\hat{D}_{c,d}(m)$. The process \hat{M}_c satisfies then for $t \in [T_{explore} + 1, T]$,

$$\hat{M}_c(t+1) = \hat{M}_c(t) + \mathbb{1}_{\{\exists n \in \mathcal{N}_c(t), c^* = c, m_n(t+1)=1\}}$$

Here c^* is the class chosen by ETC – balance. Let us compute the expected one step change of the process \hat{M}_c ,

$$\mathbb{E}[\hat{M}_c(t+1) - \hat{M}_c(t) | \hat{M}_c(t), B] = \hat{H}_{c,b_c,N}(\hat{M}_c(t)) \mathbb{1}_{\{\max_{k \in [C]} \hat{H}_{k,b_k,N}(\hat{M}_c(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t))\}} \quad (109)$$

$$= \hat{\mathcal{H}}_{c,b_c,N}(\hat{M}_c(t)) \quad (110)$$

where $\hat{H}_{c,b_c,N}(\hat{M}_c(t)) = \sum_{d=1}^D (1 - \hat{D}_{c,d}(\hat{M}_c(t))) \nu(d)$. Let $\hat{Q}_c(t+1) = \hat{M}_c(t+1) - \hat{M}_c(t) - \mathbb{E}[\hat{M}_c(t+1) - \hat{M}_c(t) | B, \hat{M}_c(t)]$. Here Q_c is a martingale difference sequences that satisfies the same assumptions in Lemma 6. for $t \in [T_{explore} + 1, T]$

$$\hat{M}_c(t+1) = \hat{M}_c(t) + \hat{\mathcal{H}}_{c,b_c,N}(\hat{M}_c(t)) + \hat{Q}_c(t+1) \quad (111)$$

$$= \hat{M}_c(t) + \mathcal{H}_{c,b_c,N}(\hat{M}_c(t)) + \Delta_{c,b_c,N}(\hat{M}_c(t)) + \hat{Q}_c(t+1) \quad (112)$$

Note that the function $\mathcal{H}_{c,b_c,N}(\hat{M}_c(t))$ is the same as the one defined in Equation (67). The goal of the proof is to show that $\frac{1}{N} \sum_{l=1}^t \Delta_{c,b_c,N}(\hat{M}_c(l))$ converges to 0 with high probability. This is important because, according to the proofs of Theorems 1 and 4 in [16], establishing this convergence implies that the process \hat{M}_c converges to the same solution as the differential inclusion introduced in Theorem 2. This follows from the fact that we can write the evolution of \hat{M}_c as in Equation (111) where \hat{Q}_c is a martingale difference term and $\mathcal{H}_{c,b_c,N}(\hat{M}_c(t))$ is the drift of the process M_c defined in Equation (67). Thus, showing that the average of the perturbation term $\Delta_{c,b_c,N}(\hat{M}_c(t))$ vanishes ensures that \hat{M}_c asymptotically follows the same differential inclusion introduced in Theorem 2.

We now turn our attention to analyzing $\Delta_{c,b_c,N}(\hat{M}_c(t))$,

$$\Delta_{c,b_c,N}(\hat{M}_c(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t)) \mathbb{1}_{\{\max_{k \in [C]} \hat{H}_{k,b_k,N}(\hat{M}_c(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t))\}} \quad (113)$$

$$- H_{c,b_c,N}(\hat{M}_c(t)) \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(\hat{M}_c(t)) = H_{c,b_c,N}(\hat{M}_c(t))\}} \quad (114)$$

$$|\Delta_{c,b_c,N}(\hat{M}_c(t))| \leq \left| \hat{H}_{c,b_c,N}(\hat{M}_c(t)) - H_{c,b_c,N}(\hat{M}_c(t)) \right| \quad (115)$$

$$+ H_{c,b_c,N}(\hat{M}_c(t)) \left| \mathbb{1}_{\{\max_{k \in [C]} \hat{H}_{k,b_k,N}(\hat{M}_c(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t))\}} - \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(\hat{M}_c(t)) = H_{c,b_c,N}(\hat{M}_c(t))\}} \right| \quad (116)$$

From the concentration inequality in Lemma 18, we have,

$$A = \left| \hat{H}_{c,b_c,N}(\hat{M}_c(t)) - H_{c,b_c,N}(\hat{M}_c(t)) \right| \leq \sum_{d=1}^D |\hat{D}_{c,d}(\hat{M}_c(t)) - D_{c,d}(\hat{M}_c(t))| \nu(d) \quad (117)$$

with probability at least $1 - \delta$,

$$A \leq 2e^a \sqrt{\frac{\log(2/\delta)}{2T_{total}}} \quad (118)$$

Now let us focus on $\left| \mathbb{1}_{\{\max_{k \in [C]} \hat{H}_{k,b_k,N}(\hat{M}_c(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t))\}} - \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(\hat{M}_c(t)) = H_{c,b_c,N}(\hat{M}_c(t))\}} \right|$ we need to bound the mismatch between the indicator functions. These indicators differ only when the maximum changes, i.e., when the "argmax" under $\hat{D}_{c,d}(\hat{M}_c(t))$ and $D_{c,d}(\hat{M}_c(t))$ do not agree. Suppose $H_{c,b_c,N}(\hat{M}_c(t))$ is the largest value among all $H_{k,b_k,N}(\hat{M}_c(t))$ by a margin $\gamma > 0$,

$$H_{c,b_c,N}(\hat{M}_c(t)) > H_{k,b_k,N}(\hat{M}_c(t)) + \psi \quad \text{for } k \neq c$$

To ensure that the estimator $\hat{D}_{c,d}$ does not flip the argmax, we want to control how much each function $H_{k,b_k,N}(\hat{M}_k(t))$ can change under small perturbations in $D_{c,d}$. Suppose that the maximizer changes for the function \hat{H} , this means that for some $j \neq c$,

$$\hat{H}_{j,b_j,N}(\hat{M}_j(t))(\hat{M}_j(t)) \geq \hat{H}_{c,b_c,N}(\hat{M}_c(t)) \quad (119)$$

Thus,

$$\hat{H}_{j,b_j,N}(\hat{M}_j(t)) - \hat{H}_{c,b_c,N}(\hat{M}_c(t)) = \hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t)) - \hat{H}_{c,b_c,N}(\hat{M}_c(t)) \quad (120)$$

$$+ H_{c,b_c,N}(\hat{M}_c(t)) + H_{j,b_j,N}(\hat{M}_j(t)) - H_{c,b_c,N}(\hat{M}_c(t)) \quad (121)$$

$\hat{H}_{j,b_j,N}(\hat{M}_j(t)) \geq \hat{H}_{c,b_c,N}(\hat{M}_c(t))$ implies,

$$J = \hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t)) - \hat{H}_{c,b_c,N}(\hat{M}_c(t)) + H_{c,b_c,N}(\hat{M}_c(t)) \quad (122)$$

$$J \geq H_{c,b_c,N}(\hat{M}_c(t)) - H_{j,b_j,N}(\hat{M}_j(t)) \geq \psi \quad (123)$$

Applying the triangle inequality, we get,

$$B = |\hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t)) - \hat{H}_{c,b_c,N}(\hat{M}_c(t)) + H_{c,b_c,N}(\hat{M}_c(t))| \quad (124)$$

$$B \leq |\hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t))| + |\hat{H}_{c,b_c,N}(\hat{M}_c(t)) - H_{c,b_c,N}(\hat{M}_c(t))| \quad (125)$$

Thus by the margin condition, we have

$$|\hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t))| + |\hat{H}_{c,b_c,N}(\hat{M}_c(t)) - H_{c,b_c,N}(\hat{M}_c(t))| \geq \psi \quad (126)$$

$$\Rightarrow \max\{|\hat{H}_{j,b_j,N}(\hat{M}_j(t)) - H_{j,b_j,N}(\hat{M}_j(t))| + |\hat{H}_{c,b_c,N}(\hat{M}_c(t)) - H_{c,b_c,N}(\hat{M}_c(t))|\} \geq \psi/2 \quad (127)$$

Let $\gamma_t = \left| \mathbb{1}_{\{\max_{k \in [C]} \hat{H}_{k,b_k,N}(\hat{M}_k(t)) = \hat{H}_{c,b_c,N}(\hat{M}_c(t))\}} - \mathbb{1}_{\{\max_{k \in [C]} H_{k,b_k,N}(\hat{M}_k(t)) = H_{c,b_c,N}(\hat{M}_c(t))\}} \right|$, according to previous development, we deduce that,

$$\gamma_t \leq \mathbb{1}_{\{\exists k \in C, |\hat{H}_{k,b_k,N}(\hat{M}_k(t)) - H_{k,b_k,N}(\hat{M}_k(t))| \geq \psi/2\}} \quad (128)$$

But from concentration, we have the following condition on γ ,

$$\psi \geq 4e^a \sqrt{\frac{\log(2/\delta)}{2T_{total}}} \quad (129)$$

Thus choosing δ such that this condition is satisfied, we ensure that with high probability $\frac{1}{N} \sum_{l=1}^t \Delta_{c,b_c,N}(\hat{M}_c(l))$ tends to 0. Having established this convergence, and noting that the process \hat{M}_c can be represented in the form given in Equation (67). Thus according to the proof of Theorem 1 and 4 in [16], \hat{M}_c converges to the solution of the same differential inclusion defined in Theorem 2. and we have $|\hat{M}_c(T) - Nm_c(T/N)| \sim \mathcal{O}(N^{\frac{q+3}{4}})$ for $0 < q < 1$. □

With all the previous result, $R(T) = R_{\text{explore}}(T_{\text{explore}}) + R_{\text{exploit}}(T_{\text{exploit}})$, we showed first that $R_{\text{explore}}(T_{\text{explore}}) \sim \mathcal{O}(T^\omega)$, and $R_{\text{exploit}}(T_{\text{exploit}}) \leq C(2T^{\frac{q+3}{4}} + T^\omega)$, thus choosing $\omega = \frac{q+3}{4}$, we get that $R(T) \sim \mathcal{O}(T^{\frac{q+3}{4}})$.