LOCALLY SPARSIFIED, GLOBALLY NEAR-OPTIMAL: MATCHING UNDER INDEPENDENT VERTEX ARRIVALS

Anonymous authors
Paper under double-blind review

ABSTRACT

Resource allocation systems often restrict each request to a short list of options before coordinating assignments globally. We study this separation in stochastic bipartite matching under independent vertex arrivals. Each request draws a state from its own known distribution, determining its compatible resources, and independently retains a menu of at most k edges. A maximum matching is then computed on the retained graph.

We show that bounded local menus universally suffice for near-optimal matching. For every $\varepsilon > 0$, there is a menu size k_ε depending only on ε that preserves at least a $(1 - \varepsilon)$ fraction of the expected maximum-matching size of the full realized graph. Earlier guarantees required additional assumptions on how matching mass is distributed across edges; our result resolves the unrestricted case.

Moreover, the menus are simple to generate from any benchmark matching rule, either by weighted sampling according to the benchmark's edge marginals, or by applying the benchmark to sampled realizations and retaining the resulting partners.

Our proof constructs a near-optimal certificate inside the sparsifier by combining a *locally computable* surrogate for the large-marginal edges with a fractional completion from sampled light edges. The surrogate nearly preserves the benchmark's value and endpoint loads while controlling dependencies, which makes the statistical light-edge completion possible.

1 INTRODUCTION

Resource allocation often separates local candidate selection from global coordination. In a batched robotaxi dispatch system, each request can identify a short list of compatible vehicles before the platform jointly assigns vehicles to requests. Similarly, a cloud scheduler can receive a small set of eligible servers for each job before resolving competition for server capacity. These examples motivate a basic question: how much information must each request retain for the final allocation to remain nearly optimal?

The local sparsification model of Ahmadian et al. (2026) captures this separation. Resources form one side of a bipartite graph, and requests form the other. Each request observes its compatible resources and retains a *menu* of at most k incident edges. Its choice may use statistical knowledge of the instance and private randomness, but cannot depend on the realized requests elsewhere in the system. A central coordinator then computes a maximum matching using the retained edges. The parameter k measures the amount of information passed from each local decision to the global optimization: n requests communicate at most nk resource identifiers.

We study this model under *independent vertex arrivals*. Each request independently draws its compatible neighborhood from its own known distribution. The model includes both i.i.d. request types and independent presence or absence of vertices with fixed neighborhoods.

Statistical knowledge can help a request select options that will remain useful when other requests compete for the same resources. However, each menu must be chosen without observing that competition. An option that is dispensable in one realization may be essential in another. The question is whether a bounded number of local choices can accommodate this uncertainty. Ahmadian et al. (2026) established near-optimal preservation under a spread condition on the fractional plan guiding the menus, and demonstrated strong empirical performance on ride-hailing data and synthetic

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

instances, including instances outside the regime covered by their spread-based analysis. Their work left open whether the spread condition reflects a genuine barrier to near-optimal preservation or only a limitation of the analysis. We ask this question for the more general independent-arrival model.

Question 1.1 (Near-optimal matching from bounded local menus)**.** For every $\varepsilon > 0$, is there a menu bound k_ε, depending only on ε, such that every instance with independent vertex arrivals admits a local menu rule preserving at least a $(1-\varepsilon)$ fraction of the expected maximum-matching size of the full realized graph?

Our results. We answer Question 1.1 affirmatively.

Theorem 1.2 (Informal; see Theorem 3.1)**.** *For every $\varepsilon > 0$, there is a finite k_ε such that every instance with independent vertex arrivals admits a local menu rule of size at most k_ε whose retained graph H satisfies*

$$\mathbb{E}[\nu(H)] \geq (1-\varepsilon)\mathbb{E}[\nu(\mathcal{G})],$$

where $\mathcal{G}$ is the full realized graph and $\nu(G)$ is the maximum-matching size of a graph G. The expectations are over the arrivals and the menu randomness.

Note that the number of retained choices per request needed to preserve the expected global matching value depends only on the desired accuracy. In particular, it is independent of the number of requests, resources, possible states, or neighborhood sizes.

The menu construction is simple: weighted sampling according to the conditional edge marginals of any randomized benchmark matching rule $M^\star$ preserves $(1-\varepsilon)\mathbb{E}[|M^\star|]$ for $k \geq k_\varepsilon$. Alternatively, only sample access is needed: for each request, fix its observed state, independently simulate k realizations of the other requests, run $M^\star$ on each, and retain its at most k distinct assigned partners. Taking $M^\star$ to compute a maximum matching gives our near-optimality guarantee. These constructions extend the Monte Carlo approach explored empirically for i.i.d. arrivals by Ahmadian et al. (2026) to general independent arrivals.

Our proof gives the quantitative bound

$$k_\varepsilon \leq \exp\exp\left(O\left(\varepsilon^{-1}\log\frac{1}{\varepsilon}\right)\right).$$

We show that dependence on the menu size is unavoidable. We establish in Section H a lower bound that applies *for every local menu rule*, exhibiting independent-presence instances on which the expected loss is $\Omega(1/k)$ of the full maximum-matching value. Consequently, any universal $(1-\varepsilon)$ guarantee requires $k_\varepsilon = \Omega(1/\varepsilon)$. The gap between this lower bound and our upper bound is large, and our result is foundational and not intended to predict menu sizes needed in practice. Our experiments in Section 4 examine the dependence on k of hub and expanded instance families. They indicate that good preservation can already occur for small menu sizes, consistent with the strong empirical performance observed by Ahmadian et al. (2026) beyond the regime covered by their theory.

Related work. In classical online bipartite matching, each arriving request must be matched or discarded immediately. The RANKING algorithm of Karp et al. (1990) fixes a uniformly random priority order on the offline vertices and matches each request to its highest-priority available neighbor. It achieves the optimal $1 - 1/e$ competitive ratio under adversarial arrivals. Random-order arrivals allow improved guarantees without knowledge of a demand distribution (Karande et al., 2011; Mahdian & Yan, 2011; Peng & Tang, 2025). A complementary line of work assumes requests are drawn independently from a known distribution (Feldman et al., 2009; Bahmani & Kapralov, 2010; Manshadi et al., 2012; Jaillet & Lu, 2014; Huang & Shu, 2021; Huang et al., 2022). In all three models, however, the worst-case competitive ratio is bounded away from one: this limitation persists even with known i.i.d. arrivals (Manshadi et al., 2012; Chierichetti et al., 2025).

Local sparsification permits the final matching to be deferred: the irrevocable decision is which options to retain. Unlike an online matching algorithm, each request chooses its menu without observing earlier requests or current resource availability. Despite this restriction, our result shows that a menu size depending only on ε suffices to preserve a $(1-\varepsilon)$ fraction of the expected offline optimum under independent arrivals with known distributions.

Streaming matching likewise permits a final computation after processing the input, and studies what can be preserved in limited memory (Goel et al., 2012; Kapralov, 2013; 2021). A streaming algorithm

can use its accumulated memory when processing each arrival. Here each menu depends only on the known arrival distributions, the request's own state, and its private randomness. The statistical assumption also differs from random-order streams: we sample request states independently, rather than randomly permuting a fixed collection of requests. Our formulation isolates the power of distribution-informed local selection followed by global optimization.

A different stochastic matching literature studies sparse queries to a graph whose edges are independently realized (Blum et al., 2015; Assadi et al., 2019; Behnezhad et al., 2019; 2020; Behnezhad & Derakhshan, 2020; Azarmehr et al., 2025). In its nonadaptive formulation, the query subgraph is chosen before the edge realizations are observed. Our menus are chosen after observing each request's realized neighborhood, with a separate budget for each request. At the same time, a single request state can jointly determine many edges, so independence of individual edges is unavailable.

Proof ideas. We prove the guarantee by constructing an auxiliary fractional matching on the retained graph whose expected value nearly matches that of the benchmark. This certificate is used only in the analysis: the algorithm simply forms the menus and computes a maximum matching on the retained graph. The large-marginal edges of the benchmark are retained in the menu. The main challenge is that the large- and small-marginal parts of the benchmark share endpoint capacity. Although the benchmark has the right value and endpoint loads, its unmatched vertices may have global dependencies that obstruct concentration for the sampled light edges.

To control these dependencies, we build on the connection between local computation algorithms and stochastic matching developed by Azarmehr et al. (2025), using the framework of Rubinfeld et al. (2011); Alon et al. (2012). We replace the large-marginal benchmark matching with a locally computable surrogate that nearly preserves its value and endpoint loads, while in–out query bounds control dependencies among unmatched vertices. We adapt this approach from independently realized edges to independent vertex arrivals, where resampling one request can change an entire neighborhood. The resulting residual capacity allows a fractional completion from sampled light edges to recover nearly all of the remaining benchmark mass. Bipartite matching integrality then gives an integral matching of at least the certificate's value. Section 3 gives the formal theorem and proof roadmap.

2 Product-arrival model

Definition 2.1 (Product-arrival instance)**.** A *product-arrival instance* $\mathcal{I}$ consists of a finite set V of offline vertices, an integer n, and, for each online position $i \in [n]$, a finite state space $\mathcal{T}_i$, a distribution p_i with support $\mathcal{T}_i$, and a neighborhood map

$$\Gamma_i : \mathcal{T}_i \to 2^V.$$

A *realized graph* $\mathcal{G} \sim \mathcal{I}$ is sampled by drawing independently $T_i \sim p_i$ for each $i \in [n]$ and setting

$$\mathcal{G} = \big([n], V, \{(i, v) : v \in \Gamma_i(T_i)\}\big).$$

We call a pair (i, t) with $t \in \mathcal{T}_i$ a *state*, and a triple (i, t, v) with $v \in \Gamma_i(t)$ a *state-edge*.

Sparsification via menus. For a graph G, let $\nu(G)$ denote its maximum-matching size.

Definition 2.2 (Menu rule)**.** A *menu rule of size* k for an instance $\mathcal{I}$ assigns to every state (i, t) a distribution $\mathcal{D}_{it}$ over subsets $S \subseteq \Gamma_i(t)$ with $|S| \le k$.

Given a realized graph $\mathcal{G}$, the rule samples a *retained graph* H by drawing, independently for each i,

$$S_i \sim \mathcal{D}_{i,T_i},$$

and setting

$$H := \big([n], V, \{(i, v) : v \in S_i\}\big) \subseteq \mathcal{G}.$$

The value of the menu rule is $\mathbb{E}[\nu(H)]$, where the expectation is over both the product arrivals and the menu randomness.

The distributions $\mathcal{D}_{it}$ may depend arbitrarily on the known instance $\mathcal{I}$; the realized menu draws use independent private randomness.

Benchmark marginals and derived menus. Fix a randomized benchmark matching rule $M^\star$. We represent its randomness by an independent random seed $\omega^\star$ and write

$$M^\star = M^\star(T, \omega^\star),$$

where for every state profile T, $M^\star(T, \omega^\star)$ is a matching of the realized graph $\mathcal{G}(T)$. The seed $\omega^\star$ is independent of the arrival states and of all menu and auxiliary randomness used below.

For every state-edge (i, t, v), define the conditional and unconditional benchmark marginals

$$z_{itv} := \mathsf{Pr}[(i, v) \in M^\star \mid T_i = t], \qquad x_{itv} := p_i(t) z_{itv}.$$

Since $M^\star$ is always a matching,

$$\sum_v z_{itv} \le 1 \quad \text{for every state } (i, t), \qquad \sum_{i,t} x_{itv} \le 1 \quad \text{for every } v \in V.$$

Write

$$Z := \sum_{i,t,v} x_{itv} = \mathbb{E}[|M^\star|],$$

and, for a set F of state-edges,

$$x(F) := \sum_{(i,t,v) \in F} x_{itv}.$$

If $M^\star$ selects a maximum matching of every realized graph, then

$$Z = \mathbb{E}[\nu(\mathcal{G})].$$

Definition 2.3 (Benchmark-derived menu rule)**.** A menu rule of size k is *derived from* $M^\star$ if, for every state (i, t) and $S \sim \mathcal{D}_{it}$:

(D1) *Coverage:* for every $v \in \Gamma_i(t)$,

$$\mathsf{Pr}[v \notin S] \le e^{-k z_{itv}}. \tag{1}$$

(D2) *Nonpositive covariances:* for distinct $v, w \in \Gamma_i(t)$, $\mathrm{Cov}(\mathbf{1}\{v \in S\}, \mathbf{1}\{w \in S\}) \le 0$.

We will use the consequence

$$\Pr_{S \sim \mathcal{D}_{it}}[v \in S] \ge (1 - e^{-1}) \min\{1, k z_{itv}\} \ge \tfrac{5}{8} \min\{1, k z_{itv}\}. \tag{2}$$

More generally, (D1) could be replaced by $\mathsf{Pr}[v \notin S] \le e^{-ckz_{itv}}$ for any fixed constant $c > 0$; this only changes universal constants in our statements and proofs below.

Two standard constructions satisfy these conditions. A fixed-size VarOpt$_k$ sample of $(z_{itv})_v$ (Chao, 1982; Srinivasan, 2001; Cohen et al., 2011) keeps every edge with $z_{itv} \ge 1/k$ and has miss probability at most $(1 - k z_{itv})_+ \le e^{-k z_{itv}}$, with negatively orthant dependent inclusion indicators. Alternatively, draw k independent auxiliary realizations of the environment conditional on $T_i = t$, run $M^\star$ on each, and let S_i be the set of partners assigned to i. Then $|S_i| \le k$, an edge is missed with probability $(1 - z_{itv})^k \le e^{-k z_{itv}}$, and the resulting occupancy indicators are negatively associated.

3 Universal sparsification theorem and proof roadmap

Theorem 3.1 (Universal sparsification for product arrivals)**.** *There are universal constants $C, \varepsilon_0 > 0$ such that, for every $0 < \varepsilon \le \varepsilon_0$, if*

$$k_\varepsilon := \left\lceil \exp\left(\exp\left(C \varepsilon^{-1} \log \frac{C}{\varepsilon} \right) \right) \right\rceil,$$

then for every product-arrival instance, every matching rule $M^\star$, and every size-k_ε menu rule derived from $M^\star$,

$$\mathbb{E}[\nu(H)] \ge (1 - \varepsilon) \mathbb{E}[|M^\star|].$$

Taking $M^\star$ to be a maximum-matching selector gives

$$\mathbb{E}[\nu(H)] \geq (1-\varepsilon)\mathbb{E}[\nu(\mathcal{G})].$$

The remainder of this section gives a high-level roadmap of the proof with pointers to formal statements and technical arguments in the appendices.

Fix the benchmark $M^\star$ and write $Z := \mathbb{E}[|M^\star|]$. Fix a benchmark-derived menu rule of size k_ε, and let H be its retained graph. We carry out the proof using an auxiliary accuracy parameter η. Specifically, we construct a random feasible fractional matching Y supported on H, which we call our *certificate*, such that

$$\mathbb{E}[|Y|] \geq (1 - O(\eta))Z.$$

By integrality of the bipartite matching polytope, $\nu(H) \geq |Y|$ in every realization. At the end of the proof, we set η to a sufficiently small constant multiple of ε.

Observe that we cannot simply use $M^\star \cap H$, the benchmark restricted to retained edges, as our certificate. This is because light edges (those with small conditional marginals) may carry a substantial fraction of Z even though each is retained with only a small probability. Consequently, restricting the particular benchmark matching to H can lose most of this mass. Recovering it requires using retained light edges other than those selected by that realization of the benchmark.

3.1 Crucial, light, and intermediate bands.

We partition the state-edges according to their conditional benchmark marginals using two thresholds

$$0 < \tau^- < \tau \leq \frac{1}{4}.$$

The *crucial*, *intermediate*, and *light* state-edges are:

$$C_\tau := \{z_{itv} \geq \tau\}, \qquad I_{\tau^-,\tau} := \{\tau^- < z_{itv} < \tau\}, \qquad L_{\tau^-} := \{z_{itv} \leq \tau^-\}. \tag{3}$$

The threshold τ is set to be large enough so that each realized crucial edge is in H with probability at least $1-\eta$. The light threshold τ^- is set to be substantially smaller than τ as to facilitate statistical guarantees on the sampled (retained) light edges; the precise requirement appears in Lemma D.2.

Our certificate combines two parts that separately recover the crucial and the light parts of the benchmark mass:

- **Crucial.** We construct an auxiliary matching M_C on the realized crucial graph, which we call the *crucial witness* (Section 3.3). It nearly preserves the crucial benchmark value $x(C_\tau)$ and its endpoint loads are (nearly) upper bounded by those of the crucial benchmark $M^\star \cap C_\tau$. Additionally, the witness also has controlled local dependencies (Section 3.2). Note that we construct M_C independently of the menu randomness. Since each crucial edge is retained with probability at least $1-\eta$, intersecting M_C with H loses at most an η fraction of its expected value.
- **Light.** We then construct a fractional matching on retained light edges whose endpoints are free in M_C (Sections 3.4 and 3.5). Together with $M_C \cap H$, this forms our certificate.
- **Intermediate.** The intermediate band $I_{\tau^-,\tau}$ is left unaccounted for by our certificate. In Section 3.6 we show that the two thresholds can be chosen so that $x(I_{\tau^-,\tau}) = O(\eta Z)$.

3.2 Local computation as a dependency-control tool.

To recover the light mass from the sampled menus, we use inverse-probability weights that preserve the target masses in expectation. Variance bounds then control the loss from reducing overloaded endpoint weights to obtain a feasible fractional matching. Additionally, we must also restrict to vertices left free by M_C and compensate for their availability. For the weighted sums to concentrate, we need to control the variance by accounting for the dependencies among the *free-status indicators* of the crucial witness M_C, which indicate whether each vertex is unmatched in M_C.

Note that we cannot simply take $M^\star \cap C_\tau$ as the crucial witness. Although it has the desired value and endpoint loads, its free-status indicators may inherit arbitrary global correlations from $M^\star$. We

therefore construct M_C using *local computation*, which lets us control dependencies among its free-status indicators while preserving the required value and endpoint-load guarantees. We describe the local computation tools used for this purpose. See Section A for details.

A local computation algorithm (LCA) (Rubinfeld et al., 2011; Alon et al., 2012) represents a global randomized object through consistent local queries. For example, rather than explicitly constructing an entire matching, a query at a vertex determines its matched/free status by probing only a small connected portion of the input graph. Queries use shared randomness so that all answers are consistent with one global matching. We use the following two primitives.

In–out bounds to variance bounds. On a fixed input graph, each query reads randomness from only a bounded set of locations. Overlap between the locations read by different queries controls the covariances of their outputs. Bounding this overlap therefore yields variance bounds for the weighted sums of free-status indicators. We measure locality in two directions. The *out* complexity R^+ bounds how many input vertices a single query can probe, while the *in* complexity R^- bounds, for any fixed input vertex, the expected number of query roots whose executions probe that vertex. Informally, out locality asks how much one output can depend on, whereas in locality asks how many outputs can depend on one input. Together these bounds control aggregate dependence: an in–out profile (R^+, R^-) bounds the total overlap among the information read by different queries, yielding covariance and variance bounds for sums of local outputs (Lemma A.3). This in–out viewpoint was developed for (a different model of) stochastic matching by Azarmehr et al. (2025).

Local Weighted Matching We construct the crucial matching using a near-optimal local weighted-matching rule (Theorem A.4) with two additional properties: (i) in–out bounds that are independent of graph size (depending only on the maximum degree and the accuracy parameter), and (ii) edge-inclusion probabilities in the output matching are continuous in the weights on each fixed graph.

3.3 A CRUCIAL WITNESS: FROM BENCHMARK LOADS TO A LOCAL MATCHING.

For an arrival realization T, let

$$\mathcal{G}_C(T) := \big([n], V, \{(i, v) : (i, T_i, v) \in C_\tau\}\big)$$

be the realized crucial graph. We seek a matching on this graph that nearly preserves the crucial benchmark value, does not substantially exceed its endpoint loads, and has sufficiently localized dependencies. We refer to this matching as the *crucial witness*.

The construction (Section B) proceeds in two stages. First, we cap the realized crucial graph to obtain degree bounds, which are needed to bound the in–out complexity of the local matching rule. Every online vertex has at most $1/\tau$ crucial neighbors, but an offline vertex can have arbitrarily large realized crucial degree. We delete all edges incident to an offline vertex whose crucial degree exceeds

$$D = \Lambda_\eta(\tau) = O\left(\frac{1}{\tau} + \log\frac{1}{\eta}\right),$$

and denote the resulting graph by $\mathcal{G}_C^D$.

Let $M_C^{\star,D}$ be the restriction of the crucial part of $M^\star$ to $\mathcal{G}_C^D$, let

$$Q := \mathbb{E}[|M_C^{\star,D}|],$$

and let q denote its endpoint marginals, indexed by online state-events and offline vertices. We show (Lemma B.1) that capping preserves almost all of the crucial benchmark value:

$$Q \geq (1 - O(\eta))x(C_\tau) \qquad \text{and} \qquad \Delta(\mathcal{G}_C^D) \leq D.$$

Capping also makes the graph *stable under resampling*: changing one arrival state modifies only a controlled portion of $\mathcal{G}_C^D$. This property lets us extend the fixed-input correlation bounds supplied by the in–out framework to the randomness of the product arrivals. Combining capping stability with the locality of the matching construction yields a decorrelation bound (Theorem D.1), which controls the variance of weighted sums of free-status indicators.

Note that capping alone does not remove the benchmark's global correlations: $M_C^{\star,D}$ may still inherit arbitrary dependencies from $M^\star$. We use $M_C^{\star,D}$ only as a reference for its value and endpoint marginals. In Section B.2, we replace $M_C^{\star,D}$ by an auxiliary matching M_C on $\mathcal{G}_C^D$, constructed with fresh randomness and controlled in–out locality.

The construction uses weighted matching to discourage endpoint loads from exceeding the capped-benchmark loads q. For a candidate vector y of endpoint marginals, we assign weights to the crucial state-edges as a function of y and the target loads q, and apply our local weighted-matching primitive to the realized capped crucial graph. Let $\mathcal{R}(y)$ denote the resulting vector of endpoint marginals.

We choose the edge weights so that the induced map $\mathcal{R}$ is continuous, and Brouwer's theorem supplies a fixed point $y^\star = \mathcal{R}(y^\star)$. The crucial witness M_C is the matching produced at this fixed point.

The weights penalize overloaded endpoints increasingly strongly. In particular, once the candidate load at an endpoint u reaches $(1+\eta^2)q_u$, its penalty saturates and no incident edge is selected by the response. Such an overload therefore cannot occur at a fixed point. Together with near-optimality of the weighted-matching response, this yields a witness that nearly preserves the capped crucial value while controlling every endpoint load. Writing $y := y^\star$ for its endpoint marginals,

$$\mathbb{E}[|M_C|] \geq (1-\eta)Q, \qquad y_u \leq (1+\eta^2)q_u \quad \text{for every endpoint } u. \tag{4}$$

Its free-status indicators also have the controlled in–out locality needed for the light-completion argument (Section 3.5).

Note that M_C is constructed without looking at the sampled menus and need not itself be contained in H. Only afterward do we intersect it with H. Every edge of M_C is crucial, so, conditional on the arrival states and on M_C, the coverage property (D1) gives $\mathsf{Pr}[e \notin H] \leq e^{-k\tau}$ for every $e \in M_C$. Hence, once $k\tau \geq \log(1/\eta)$, $\mathbb{E}[|M_C \cap H|] \geq (1-\eta)\mathbb{E}[|M_C|]$. Combining this with the witness value guarantee and $Q \geq (1-\eta)x(C_\tau)$ gives $\mathbb{E}[|M_C \cap H|] \geq x(C_\tau) - O(\eta Z)$.

3.4 Fitting the light benchmark to the witness.

We next prepare the light benchmark mass for completion on endpoints left free by M_C. This step, formalized in Section C, uses only the marginal probabilities that endpoints are free in M_C; it does not depend on which endpoints are free in a particular realization.

The initial light *plan* is the restriction of the benchmark marginal vector x to L_{τ^-}. Since the light part and $M_C^{\star,D}$ are disjoint subsets of the benchmark matching, this plan is feasible in the residual marginal capacities left by $M_C^{\star,D}$: at most $p_i(t) - q_{it}$ at each online state (i,t) and $1 - q_v$ at each offline vertex v. The witness M_C instead leaves residual capacities $p_i(t) - y_{it}$ and $1 - y_v$.

The pointwise load guarantee (4) implies that the benchmark residual capacity can exceed the residual capacity left by M_C by at most an η^2 fraction of the full endpoint capacity. Consequently, at every endpoint whose residual availability under M_C is at least η, the benchmark residual capacity is at most a $1 + \eta^2/\eta = 1 + \eta$ factor larger than the witness residual capacity.

We therefore discard light edges incident to endpoints whose residual availability is below η. Such endpoints carry only $O(\eta Z)$ benchmark light mass in total. On the remaining endpoints, dividing the light plan by $1 + \eta$ makes it feasible in the residual capacities left by M_C.

The resulting subplan x^0 is supported on L_{τ^-}, satisfies $0 \leq x^0_{itv} \leq x_{itv}$, has

$$|x^0| \geq x(L_{\tau^-}) - O(\eta Z), \qquad \sum_v x^0_{itv} \leq p_i(t) - y_{it}, \qquad \sum_{i,t} x^0_{itv} \leq 1 - y_v.$$

Moreover, every endpoint in the support of x^0 has residual availability at least η. Write $z^0_{itv} := \frac{x^0_{itv}}{p_i(t)}$ for its corresponding conditional masses. In the next step we show how to recover nearly all of x^0 from the retained light edges.

3.5 Realizing the light residual.

We now use the light edges retained by the original menu rule to realize the fitted analysis plan x^0 on endpoints left free by M_C. See Section D for full details. This requires compensating for both menu sampling and endpoint availability.

First consider menu sampling. Conditional on $T_i = t$, assign the Horvitz–Thompson weight (Horvitz & Thompson, 1952)

$$W_{iv} := z^0_{itv} \frac{\mathbf{1}\{v \in S_i\}}{\mathsf{Pr}[v \in S_i \mid T_i = t]}.$$

Thus

$$\mathbb{E}[W_{iv} \mid T_i = t] = z^0_{itv}.$$

These weights recover the fitted light masses in expectation. When $1/k \le \tau^-$, the coverage property (D1) also bounds their conditional second moments by $O(\tau^-)z^0_{itv}$, while (D2) gives nonpositive covariances between distinct weights from the same menu.

Next we restrict to endpoints free in M_C. Let F_i and F_v be the indicators that online vertex i and offline vertex v are free, and write

$$a_{it} := \mathsf{Pr}[F_i = 1 \mid T_i = t] = 1 - \frac{y_{it}}{p_i(t)}, \qquad a_v := \mathsf{Pr}[F_v = 1] = 1 - y_v.$$

Simply multiplying W_{iv} by F_iF_v would lose light mass: even if the free statuses were independent, this would introduce an additional factor $a_{it}a_v$ in the expected weight. We therefore start with the availability-normalized weights

$$U_{iv} := W_{iv} \frac{F_iF_v}{a_{iT_i}a_v}.$$

Each availability in these denominators is at least η on the support of x^0, by the residual-fitting step.

The normalization compensates for endpoint availability, but it does not make these weights exactly unbiased. Conditional on $T_i = t$, the indicators F_i and F_v may be correlated, and the free probability of v may differ from its unconditional value a_v. Moreover, feasibility requires control of the total weight incident to each free endpoint, not just the expected weight of each edge.

This is where the local construction of M_C is used. The product-arrival decorrelation bound of Theorem D.1 controls variances of weighted sums of free-status indicators and the aggregate bias caused by conditioning on an arrival state. Taking τ^- sufficiently small relative to the locality bounds at threshold τ makes the resulting losses $O(\eta Z)$.

The light-completion argument in Lemma D.2 divides these weights by $1 + \eta$ to create slack, then scales down overloaded rows and columns in each realization. The variance and conditioning-bias bounds show that the expected mass lost is only $O(\eta Z)$. To control the initial value error, we also use the pointwise witness bound, which bounds the total offline matched mass $\sum_v y_v$ by $(1 + \eta^2)Q$. The result is a feasible fractional matching $\widehat{Y}$ supported on retained light edges with endpoints free in M_C, such that

$$\mathbb{E}[|\widehat{Y}|] \ge |x^0| - O(\eta Z) \ge x(L_{\tau^-}) - O(\eta Z).$$

Since $\widehat{Y}$ uses only vertices free in M_C, it is compatible with the surviving crucial matching $M_C \cap H$. Their union is our fractional matching certificate Y. Bipartite matching integrality therefore gives

$$\mathbb{E}[\nu(H)] \ge x(C_\tau) + x(L_{\tau^-}) - O(\eta Z) = Z - x(I_{\tau^-,\tau}) - O(\eta Z). \tag{5}$$

3.6 Choosing the band thresholds.

It remains to choose the thresholds τ and τ^-. See Section E for full details. For $m = \Theta(1/\eta)$, we construct $\rho_0 > \rho_1 > \cdots > \rho_m$, so that every consecutive pair $(\tau, \tau^-) = (\rho_j, \rho_{j+1})$ has enough scale separation for the light-completion argument. The corresponding intermediate bands are disjoint. Since their total benchmark mass is at most Z, one of the m bands has mass $O(\eta Z)$ by pigeonhole. We use that pair in (5).

The scale separation required at each step makes the thresholds shrink by a large power. Iterating this recurrence for $\Theta(1/\eta)$ levels gives $\rho_m^{-1} \le \exp\exp\left(O\left(\eta^{-1}\log\frac{1}{\eta}\right)\right)$. Taking $k \ge \rho_m^{-1}$ also guarantees the menu-coverage condition needed for the crucial witness and the sampling scale needed for the light completion. Finally setting $\eta = \Theta(\varepsilon)$ proves Theorem 3.1.

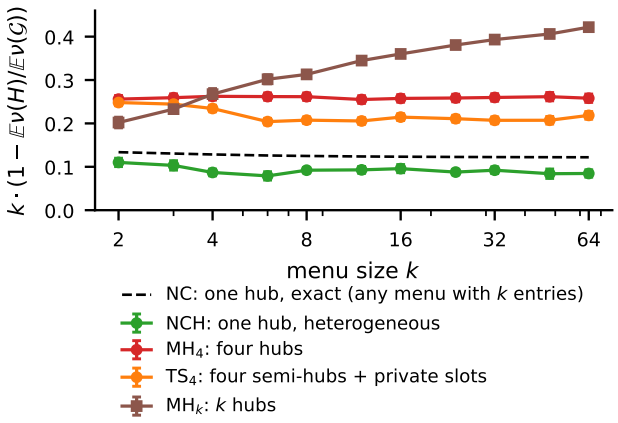

Figure 1: Loss multiplied by k for VarOpt$_k$ menus, maximized over the tested scale grid. Bars show twice the estimated evaluation standard error. Dashed: exact loss for the symmetric one-hub family. Family definitions and grids are in Section I.

4 Experiments: Loss at Moderate Menu Sizes

We examine the loss $1 - \mathbb{E}[\nu(H)]/\mathbb{E}[\nu(\mathcal{G})]$ at moderate menu sizes. Our main experiments use VarOpt$_k$ menus constructed from estimated conditional marginals of a maximum-matching benchmark with random tie-breaking. The estimates and evaluation use independent arrival profiles.

Rates and constants. Figure 1 considers growing hub families, in which hubs compete for resources left free by dedicated arrivals, and a variant with contested private resources. We vary the number of hubs and the presence probabilities; instance sizes grow with k. Each point reports the largest estimated loss over the tested scale grid. Across these families, the loss is below 1.3% at $k = 32$ and 0.7% at $k = 64$, with k times the loss between 0.08 and 0.42 over the tested range. These structured families already admit an $O(1/k)$ guarantee (Proposition H.2 and Section I); the plots assess constants for the implemented sampling rule.

Sampling choices. The comparisons in Table 1 illustrate the value of marginal information and of spreading choices across resources. At $k = 4$, the reported loss on random heterogeneous arrivals is 8.1% for uniform menus versus 0.29% for VarOpt; on the four-hub family, top-k menus lose 11.4% versus 6.6% for VarOpt. Additional tests cover nested neighborhoods and random presence graphs, and show sensitivity to the benchmark selector. Section I gives the families, scale grids, estimation procedure, and further comparisons.

5 Conclusion

Bounded local menus preserve nearly all of the expected maximum-matching value under independent vertex arrivals, with menu size depending only on the desired accuracy.

The main open problem is to improve this dependence. We conjecture that the optimal worst-case loss is $\Theta(1/k)$, so menus of size $O(1/\varepsilon)$ suffice for a $(1-\varepsilon)$ guarantee. The hub class already admits this rate (Proposition H.2).

A second direction is to remove knowledge of the arrival distributions, for example under unknown i.i.d. arrivals or random-order vertex streams (Karande et al., 2011; Mahdian & Yan, 2011). Can a single pass learn enough from an initial sample to retain small menus for a near-optimal final matching? The challenge is to learn useful statistics within the memory budget and, for random-order streams, account for sampling without replacement.

AI USE STATEMENT

We used generative AI tools, primarily ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google), to explore proof strategies, formulate and critically examine intermediate mathematical claims, identify potential gaps, and assist with drafting and revising proofs. We also used these tools for literature searches and improvements to implement and run experiments, the manuscript's organization, notation, and exposition. The authors take responsibility for the mathematical claims, proofs, references, and final content of this work.

REPRODUCIBILITY STATEMENT

The model and sampling assumptions are specified in Section 2, and the proof of the main theorem is developed in Sections A to G. The lower and upper bounds for hub instances are proved in Section H. Section I describes the experimental families, sampling rules, marginal estimation, evaluation procedure, and uncertainty estimates.

REFERENCES

Sara Ahmadian, Edith Cohen, and Mohammad Roghani. Stochastic matching via local sparsification. In *Advances in Neural Information Processing Systems 39 (NeurIPS 2026)*, 2026. URL https://arxiv.org/abs/2605.14195. To appear.

Noga Alon, Ronitt Rubinfeld, Shai Vardi, and Ning Xie. Space-efficient local computation algorithms. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1132–1139. SIAM, 2012. doi: 10.1137/1.9781611973099.89. URL https://doi.org/10.1137/1.9781611973099.89.

Sepehr Assadi, Sanjeev Khanna, and Yang Li. The stochastic matching problem with (very) few queries. *ACM Transactions on Economics and Computation*, 7(3):1–19, 2019. doi: 10.1145/3355903. URL https://doi.org/10.1145/3355903.

Amir Azarmehr, Soheil Behnezhad, Alma Ghafari, and Ronitt Rubinfeld. Stochastic matching via in-n-out local computation algorithms. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 1055–1066. ACM, 2025. doi: 10.1145/3717823.3718279. URL https://arxiv.org/abs/2411.08805.

Bahman Bahmani and Michael Kapralov. Improved bounds for online stochastic matching. In *Algorithms – ESA 2010*, pp. 170–181. Springer, 2010. doi: 10.1007/978-3-642-15775-2_15. URL https://doi.org/10.1007/978-3-642-15775-2_15.

Soheil Behnezhad and Mahsa Derakhshan. Stochastic weighted matching:(1-ϵ) approximation. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 1392–1403. IEEE, 2020.

Soheil Behnezhad, Alireza Farhadi, MohammadTaghi Hajiaghayi, and Nima Reyhani. Stochastic matching with few queries: New algorithms and tools. In *Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2855–2874. SIAM, 2019. doi: 10.1137/1.9781611975482.177. URL https://arxiv.org/abs/1811.03224.

Soheil Behnezhad, Mahsa Derakhshan, and MohammadTaghi Hajiaghayi. Stochastic matching with few queries: $(1 - varepsilon)$ approximation. In *Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 1111–1124. ACM, 2020. doi: 10.1145/3357713.3384340. URL https://arxiv.org/abs/2002.11880.

Avrim Blum, John P. Dickerson, Nika Haghtalab, Ariel D. Procaccia, Tuomas Sandholm, and Ankit Sharma. Ignorance is almost bliss: Near-optimal stochastic matching with few queries. In *Proceedings of the 16th ACM Conference on Economics and Computation (EC)*, pp. 325–342. ACM, 2015. doi: 10.1145/2764468.2764479. URL https://doi.org/10.1145/2764468.2764479.

M. T. Chao. A general purpose unequal probability sampling plan. *Biometrika*, 69(3):653–656, 1982. doi: 10.1093/biomet/69.3.653. URL https://doi.org/10.1093/biomet/69.3.653.

Flavio Chierichetti, Mirko Giacchini, Alessandro Panconesi, and Andrea Vattani. A new impossibility result for online bipartite matching problems. In *52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025)*, volume 334 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 58:1–58:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025. doi: 10.4230/LIPIcs.ICALP.2025.58. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2025.58.

Edith Cohen, Nick G. Duffield, Haim Kaplan, Carsten Lund, and Mikkel Thorup. Efficient stream sampling for variance-optimal estimation of subset sums. *SIAM Journal on Computing*, 40(5):1402–1431, 2011. doi: 10.1137/10079817X. URL https://doi.org/10.1137/10079817X.

Guy Even, Moti Medina, and Dana Ron. Best of two local models: Centralized local and distributed local algorithms. *Information and Computation*, 262:69–89, 2018. doi: 10.1016/j.ic.2018.07.001. URL https://arxiv.org/abs/1402.3796.

Jon Feldman, Aranyak Mehta, Vahab Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating $1 - 1/e$. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 117–126. IEEE, 2009. URL https://arxiv.org/abs/0905.4100.

Ashish Goel, Michael Kapralov, and Sanjeev Khanna. On the communication and streaming complexity of maximum bipartite matching. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 468–485. SIAM, 2012. doi: 10.1137/1.9781611973099.41. URL https://doi.org/10.1137/1.9781611973099.41.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685, 1952. doi: 10.1080/01621459.1952.10483446. URL https://doi.org/10.1080/01621459.1952.10483446.

Zhiyi Huang and Xinkai Shu. Online stochastic matching, Poisson arrivals, and the natural linear program. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 682–693. ACM, 2021. doi: 10.1145/3406325.3451079. URL https://arxiv.org/abs/2103.13024.

Zhiyi Huang, Xinkai Shu, and Shuyi Yan. The power of multiple choices in online stochastic matching. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pp. 91–103. ACM, 2022. doi: 10.1145/3519935.3520046. URL https://arxiv.org/abs/2203.02883.

Patrick Jaillet and Xin Lu. Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, 39(3):624–646, 2014. doi: 10.1287/moor.2013.0621. URL https://doi.org/10.1287/moor.2013.0621.

Michael Kapralov. Better bounds for matchings in the streaming model. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1679–1697. SIAM, 2013. doi: 10.1137/1.9781611973105.121. URL https://arxiv.org/abs/1206.2269.

Michael Kapralov. Space lower bounds for approximating maximum matching in the edge arrival model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1874–1893. SIAM, 2021.

Chinmay Karande, Aranyak Mehta, and Pushkar Tripathi. Online bipartite matching with unknown distributions. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 587–596. ACM, 2011. doi: 10.1145/1993636.1993715. URL https://doi.org/10.1145/1993636.1993715.

Richard M. Karp, Umesh V. Vazirani, and Vijay V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 352–358. ACM, 1990. doi: 10.1145/100216.100262. URL https://doi.org/10.1145/100216.100262.

Mohammad Mahdian and Qiqi Yan. Online bipartite matching with random arrivals: An approach based on strongly factor-revealing LPs. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pp. 597–606. ACM, 2011. doi: 10.1145/1993636.1993716. URL https://doi.org/10.1145/1993636.1993716.

Vahideh H. Manshadi, Shayan Oveis Gharan, and Amin Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012. doi: 10.1287/moor.1120.0551. URL https://doi.org/10.1287/moor.1120.0551.

Huy N. Nguyen and Krzysztof Onak. Constant-time approximation algorithms via local improvements. In *Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 327–336. IEEE, 2008. doi: 10.1109/FOCS.2008.81. URL https://doi.org/10.1109/FOCS.2008.81.

Bo Peng and Zhihao Gavin Tang. Revisiting ranking for online bipartite matching with random arrivals: the primal-dual analysis. In *Proceedings of the 26th ACM Conference on Economics and Computation*, pp. 820–835, 2025.

Seth Pettie and Peter Sanders. A simpler linear time $2/3 - \varepsilon$ approximation for maximum weight matching. *Information Processing Letters*, 91(6):271–276, 2004. doi: 10.1016/j.ipl.2004.05.007. URL https://doi.org/10.1016/j.ipl.2004.05.007.

Ronitt Rubinfeld, Gil Tamir, Shai Vardi, and Ning Xie. Fast local computation algorithms. In *Innovations in Computer Science (ICS)*, pp. 223–238. Tsinghua University Press, 2011. URL https://arxiv.org/abs/1104.1377.

Aravind Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 588–597, 2001. doi: 10.1109/SFCS.2001.959935. URL https://doi.org/10.1109/SFCS.2001.959935.

A LOCAL COMPUTATION PRIMITIVES

The LCA model was introduced by Rubinfeld et al. (2011) and further developed by Alon et al. (2012); our analysis uses the in–out framework of Azarmehr et al. (2025).

Definition A.1 (Probe access and shared randomness)**.** An execution rooted at a vertex u starts by probing u and learns the input graph by probing vertices. Probing a vertex reveals its neighbors and local input data, and subsequent probes may only follow already revealed adjacencies; in particular the probed set is connected to u. We call such an execution *natural*.

Shared randomness π is represented by independent *dictionary marks*. Each mark is assigned to a vertex in its local support, called its *owner*, and may be read only after that owner has been probed. Thus all randomness used by an execution is revealed locally along its probe exploration.

Definition A.2 (Local rule and query profile)**.** A *local rule* answers a query at every root u by a natural execution, returning an output $X_u = X_u(G, \pi)$. The answers $(X_u)_u$ jointly represent the global object computed by the rule. Let $\mathrm{Read}(u)$ be the set of vertices probed by the execution rooted at u. Its *query profile* is (R^+, R^-) if

$$|\mathrm{Read}(u)| \le R^+$$

pointwise for every root and every realization of the shared randomness, while for every vertex w,

$$\mathbb{E}\big[\big|\{u : w \in \mathrm{Read}(u)\}\big|\big] \le R^-.$$

We refer to R^+ and R^- as the out-query and expected in-query complexities, respectively.

We use two LCA tools, stated next, with proofs deferred to Section F.

The first tool (Lemma A.3) is a fixed-input correlation bound for local outputs. For a fixed input graph and each vertex u, the expected number of other queries whose read sets intersect $\mathrm{Read}(u)$ is at most R^+R^-, where the expectation is over the shared randomness π. These intersection probabilities control the absolute covariances between the indicators X_u, yielding a bound on the variance of weighted sums of the form $\sum_u a_u X_u$. We apply this bound to the free-status indicators of the crucial matching. After accounting for the arrival randomness, this gives the variance bounds needed for the light-completion analysis in Lemma D.2.

Lemma A.3 (Fixed-input in–out correlation (Azarmehr et al., 2025, Lemmas 3.7–3.8))**.** *Fix a graph G and a natural local rule whose randomness is grouped into independent private tapes at their owner vertices. Suppose its query profile is (R^+, R^-). For roots u, u', let*

$$\delta(u, u') := \mathsf{Pr}[\mathrm{Read}(u) \cap \mathrm{Read}(u') \neq \varnothing].$$

Then

$$\sum_{u' \neq u} \delta(u, u') \le R^+R^- \tag{6}$$

for every u. Moreover, if the query outputs are binary $X_u \in \{0, 1\}$, then

$$|\operatorname{Cov}(X_u, X_{u'})| \le C\,\delta(u, u') \tag{7}$$

for a universal constant C. Consequently, for deterministic coefficients (a_u),

$$\mathsf{Var}\left(\sum_u a_u X_u\right) \le CR^+R^- \sum_u a_u^2. \tag{8}$$

The second tool, Theorem A.4, is a near-optimal local weighted-matching rule. Local approximation schemes for maximum-weight matching were developed by Even et al. (2018), building on the local-improvement framework of Nguyen & Onak (2008). Our rule has bounds on both out-query and expected in-query complexity that are independent of graph size, and its output-edge marginals are continuous in the edge weights on each fixed graph. Continuity is needed for the fixed-point argument in Section B. The proof is deferred to Section F and combines bounded weighted augmentations (Pettie & Sanders, 2004) with truncated random-greedy MIS (Azarmehr et al., 2025).

We use auxiliary edge weights in our construction of the crucial witness, as a way to penalize overused endpoints. For edge weights $w = (w_e)_{e \in E(G)}$, write

$$w(F) := \sum_{e \in F} w_e \qquad \text{and} \qquad \mathrm{OPT}_w(G) := \max\{w(M) : M \text{ is a matching of } G\}.$$

When all edge weights equal one, these quantities reduce to $|F|$ and $\nu(G)$, respectively.

Theorem A.4 (Local weighted matching rule)**.** *There are universal constants $c_{\mathrm{lca}}, C_{\mathrm{lca}} > 0$ such that the following holds. Fix $\eta \in (0, 1/4)$ and $\gamma \in (0, 1]$ with $\gamma \geq c_{\mathrm{lca}}\eta$. There is a natural local rule Ψ with the following properties for every finite graph G of maximum degree at most $\Delta \geq 2$ and every edge-weight vector $w \in [\gamma, 1]^{E(G)}$. The rule computes a matching*

$$M = \Psi(G, \pi; w),$$

where π comprises all internal randomness, represented by independent dictionary marks with local owners as in Definition A.1.

(i) Approximation and locality. *The output satisfies*

$$\mathbb{E}_\pi[w(M)] \geq (1 - \eta)\,\mathrm{OPT}_w(G).$$

Its query profile satisfies

$$R^+, R^- \leq \Delta^{a_\eta}, \qquad a_\eta := C_{\mathrm{lca}}\eta^{-2}\log^2\frac{e}{\eta}.$$

(ii) Continuous edge marginals. *For fixed G, η, γ and every $e \in E(G)$, the map*

$$w \longmapsto \Pr_\pi[e \in \Psi(G, \pi; w)]$$

is continuous on $[\gamma, 1]^{E(G)}$. The expectation and probability are over π, with the graph and input weights fixed.

B THE CRUCIAL WITNESS

B.1 CAPPING THE CRUCIAL GRAPH

Fix $\tau \in (0, \frac{1}{4}]$ and consider the crucial state-edges C_τ defined in Equation (3). We first cap their realized degree. Define the projected benchmark endpoint marginals

$$c_{it} := \sum_{v:(i,t,v)\in C_\tau} z_{itv}, \qquad c_v := \sum_{i,t:(i,t,v)\in C_\tau} x_{itv}.$$

Then

$$\sum_{i,t} p_i(t)c_{it} = \sum_v c_v = x(C_\tau).$$

Every state (i, t) has at most τ^{-1} crucial neighbors, since $\sum_v z_{itv} \leq 1$ and every crucial edge has $z_{itv} \geq \tau$. Write

$$d := \lceil \tau^{-1} \rceil.$$

For an offline vertex v, let

$$N_v := \sum_i \mathbf{1}\{(i, T_i, v) \in C_\tau\}$$

be its realized crucial degree. The summands are independent Bernoulli variables, and

$$\mathbb{E}[N_v] \leq \frac{c_v}{\tau} \leq \frac{1}{\tau}. \tag{9}$$

Unlike the online degree, N_v has no upper bound by a function of τ. We show that we can cap the degree while losing only an $O(\eta)$ fraction of the benchmark crucial value.

Lemma B.1 (Capping)**.** *Fix $\eta \in (0, \frac{1}{2})$ and set*

$$D = \Lambda_\eta(\tau) := \left\lceil \frac{2e}{\tau} \right\rceil + \left\lceil \log_2 \frac{1}{\eta} \right\rceil.$$

Let $\mathcal{G}_C^D$ be the capped *crucial graph, that is, $\mathcal{G}_C$ after deleting all edges incident to an offline vertex with $N_v > D$, and define*

$$M_C^{\star,D} := (M^\star \cap C_\tau) \cap E(\mathcal{G}_C^D), \qquad Q := \mathbb{E}[|M_C^{\star,D}|],$$

with endpoint marginals

$$q_{it} := \Pr[T_i = t,\ i \in V(M_C^{\star,D})], \qquad q_v := \Pr[v \in V(M_C^{\star,D})].$$

Then $\mathcal{G}_C^D$ is a deterministic function of T, and $M_C^{\star,D}$ is always a matching of $\mathcal{G}_C^D$. Moreover:

(i) in every realization, the maximum degree is bounded by D

$$\Delta(\mathcal{G}_C^D) \le D = \Lambda_\eta(\tau); \tag{10}$$

(ii) if T, T' differ only in coordinate i, then every changed capped edge is incident to one of the at most $2d$ offline vertices adjacent to i in one of the two realized crucial graphs, and

$$|E(\mathcal{G}_C^D(T)) \triangle E(\mathcal{G}_C^D(T'))| \le 2dD. \tag{11}$$

If $X_i(T, T')$ is the set of vertices whose capped neighborhood changes, then

$$|X_i(T, T') \cup \{i\}| \le 1 + 2d(D+1), \tag{12}$$

and $X_i(T, T')$ lies within distance two of i in the union of the two realized crucial graphs;

(iii) the retained benchmark value satisfies

$$Q \ge (1-\eta)x(C_\tau); \tag{13}$$

(iv) the endpoint marginals satisfy

$$\sum_{i,t} q_{it} = \sum_v q_v = Q, \qquad q_{it} \le p_i(t)c_{it}, \qquad q_v \le c_v.$$

Proof. The graph $\mathcal{G}_C^D$ depends only on the realized states. An offline vertex either retains its $N_v \le D$ crucial edges or is isolated, while every arrival has at most $d \le D$ crucial neighbors; this proves (10). Since $M_C^{\star,D}$ is obtained by intersecting $M^\star$ with this graph, it is always a matching.

For the value bound, let A_v be the event that $M^\star$ contains a crucial edge incident to v. Since $M^\star$ is a matching,

$$x(C_\tau) = \sum_v \mathsf{Pr}[A_v].$$

The capping rule retains all crucial edges at v when $N_v \le D$ and isolates v when $N_v > D$. Thus

$$x(C_\tau) - Q = \sum_v \mathsf{Pr}[A_v \cap \{N_v > D\}] \le \sum_v \mathsf{Pr}[N_v > D].$$

For every integer $r \ge 1$, independence of the Bernoulli summands in N_v and (9) give

$$\mathsf{Pr}[N_v \ge r] \le \frac{\mathbb{E}[(N_v)_r]}{r!} \le \frac{c_v}{r!\tau^r},$$

where we used $c_v \le 1$. Taking $r = D+1$ and summing over v yields

$$x(C_\tau) - Q \le \frac{x(C_\tau)}{(D+1)!\tau^{D+1}} \le \eta x(C_\tau),$$

because $D+1 \ge 2e/\tau$ and $D+1 \ge \log_2(1/\eta)$. This proves (13).

For stability, changing one state modifies only edges incident to the arrival i, so only the at most $2d$ corresponding offline neighbors can change their count. At such a vertex, if the cap status does not change then at most the edge to i changes; if it flips, the two counts are D and $D+1$, so exactly the D capped edges on the uncrossed side can change. This proves (11). The endpoints of these changed edges are i, the affected offline vertices, and at most D additional arrivals per affected offline vertex, giving (12) and the radius-two claim.

Finally, the two sums of the q's count the two endpoints of the edges of $M_C^{\star,D}$. Conditional on $T_i = t$, the probability that $M^\star$ uses a crucial edge at i is c_{it}, so $q_{it} \le p_i(t)c_{it}$; similarly $q_v \le c_v$. □

From now on set $\eta := c_0\varepsilon$, where the universal constant $c_0 > 0$ is chosen sufficiently small at the end, after fixing the constants in the witness and light-completion bounds. We choose ε_0 sufficiently small that η satisfies the universal upper bounds required below.

B.2 CONSTRUCTING THE LOCAL CRUCIAL WITNESS

We store each arrival state T_i at vertex i. The scores of its incident state-edges are evaluated there from T_i and the deterministic instance parameters. Auxiliary marks have fixed local owners, as in Definition A.1, and are sampled independently of the arrival states and the private randomness used to draw menus. This convention will also be used when coupling executions under a change of one arrival state in Section G.

Lemma B.2 (Crucial witness)**.** *There is a universal constant $C_{\mathrm{loc}} > 0$ such that the following holds for every $0 < \eta \le 1/4$ and $\tau \in (0, \frac{1}{4}]$. Let q_{it}, q_v, and Q be the capped-benchmark marginals and value of Lemma B.1, with capping parameter η. There is a random matching M_C of the realized capped crucial graph $\mathcal{G}_C^D$, a function of the arrival states and of a shared dictionary π independent of both the arrival states and the menu randomness, whose endpoint marginals*

$$y_{it} := \mathsf{Pr}[T_i = t,\ i \in V(M_C)], \qquad y_v := \mathsf{Pr}[v \in V(M_C)]$$

satisfy:

(a) value:

$$\mathbb{E}[|M_C|] \ge (1-\eta)Q;$$

(b) endpoint-load control:

$$y_{it} \le (1+\eta^2)q_{it}, \qquad y_v \le (1+\eta^2)q_v$$

for every state (i,t) and offline vertex v;

(c) locality: *the free-status queries $u \mapsto \mathbf{1}\{u \notin V(M_C)\}$, over arrivals and offline vertices alike, are answered by a natural local rule in the probe model of Definitions A.1 and A.2, with out-query and expected in-query complexity at most*

$$\Lambda_\eta(\tau)^{\widehat{a}_\eta}, \qquad \widehat{a}_\eta := C_{\mathrm{loc}}\eta^{-2}\log^2\frac{e}{\eta}.$$

The rule may depend on the fixed instance, capped-benchmark marginals, and parameters η, τ.

Proof. Set

$$\alpha := \frac{\eta}{4(1+c_{\mathrm{lca}})}, \qquad \delta := \eta^2, \qquad \gamma := c_{\mathrm{lca}}\alpha.$$

Thus $\alpha \le \eta$, $\gamma \ge c_{\mathrm{lca}}\alpha$, and

$$2\gamma = \frac{c_{\mathrm{lca}}}{2(1+c_{\mathrm{lca}})}\eta < 1.$$

We will invoke Theorem A.4 with accuracy parameter α.

If $Q = 0$, then (13) gives $x(C_\tau) = 0$, and $M_C = \varnothing$ satisfies the lemma. Assume henceforth that $Q > 0$. The instance, capped-benchmark marginals, and parameters are fixed throughout the proof.

Let

$$\mathcal{U} := \{(i,t) : t \in \mathcal{T}_i\} \sqcup V$$

be the endpoint index set: an online endpoint is indexed by a state (i,t) and an offline endpoint by a vertex v. Thus q_u and, below, y_u denote unconditional endpoint marginals on either side.

Write

$$\mathcal{U}_+ := \{u \in \mathcal{U} : q_u > 0\}.$$

We work on the endpoint-marginal feasibility box

$$K := \prod_{u \in \mathcal{U}_+} [0, \bar{y}_u], \qquad \bar{y}_{it} = p_i(t), \quad \bar{y}_v = 1,$$

and extend vectors in K by zero outside $\mathcal{U}_+$.

We preview the construction of the crucial witness. Given a candidate endpoint-marginal vector $y \in K$, we use y to assign scores to the edges of each realization of $\mathcal{G}_C^D$, and then use independent random thresholds to retain a weighted subgraph. Denote by $M(y)$ the random matching obtained by

running the local weighted-matching rule Ψ of Theorem A.4 on this weighted graph. Let $\mathcal{R}(y)$ be the vector of unconditional endpoint marginals of $M(y)$, averaged over the arrival realization and all auxiliary randomness.

Thus $\mathcal{R}$ maps a candidate marginal vector y to the marginals actually produced by the matching rule in response to the weights induced by y. We will show that $\mathcal{R}$ is a continuous self-map of K. Since K is nonempty, compact, and convex, Brouwer's theorem then gives a fixed point

$$y^\star = \mathcal{R}(y^\star).$$

We define the crucial witness by $M_C := M(y^\star)$.

Penalized response map. We restrict the response graph to state-edges whose two endpoint indices lie in $\mathcal{U}_+$. This removes no edge of $M_C^{\star,D}$ with positive probability.

For $y \in K$, define

$$\mathrm{val}(y) := \frac{1}{2} \sum_{u \in \mathcal{U}_+} y_u.$$

When y is the endpoint-marginal vector of a random matching, $\mathrm{val}(y)$ is its expected size.

Define

$$h_\delta(z) := \begin{cases} 0, & z \le 0, \\ \dfrac{z^2}{2\delta}, & 0 < z < \delta, \\ z - \dfrac{\delta}{2}, & z \ge \delta. \end{cases}$$

This function is convex and continuously differentiable, with

$$h'_\delta(z) = \begin{cases} 0, & z \le 0, \\ \dfrac{z}{\delta}, & 0 < z < \delta, \\ 1, & z \ge \delta. \end{cases}$$

In particular, $0 \le h'_\delta(z) \le 1$.

Define the one-sided overuse penalty

$$P(y) := \sum_{u \in \mathcal{U}_+} q_u h_\delta\left(\frac{y_u - q_u}{q_u}\right),$$

and the penalized objective

$$\Phi(y) := \mathrm{val}(y) - P(y).$$

The endpoint price is

$$s_u(y) := \frac{\partial P}{\partial y_u}(y) = h'_\delta\left(\frac{y_u - q_u}{q_u}\right), \tag{14}$$

so

$$0 \le s_u(y) \le 1.$$

Assign an edge with endpoints u, v the score

$$g_{uv}(y) := 1 - s_u(y) - s_v(y).$$

Thus $g_{uv}(y) \le 1$, although it may be negative.

For every potential crucial edge e, independently draw a locally owned threshold

$$\Theta_e \sim \mathrm{Unif}[\gamma, 2\gamma].$$

For $y \in K$ and a realization of the arrivals and thresholds, let G_y be the subgraph of $\mathcal{G}_C^D$ obtained by retaining exactly those edges e for which

$$g_e(y) \ge \Theta_e.$$

Every edge of G_y therefore has score in $[\gamma, 1]$. On G_y, with edge weights $g_e(y)$, apply the local weighted-matching rule Ψ of Theorem A.4, with accuracy parameter α. Let

$$M(y) := \Psi(G_y, \pi_\Psi; g(y)),$$

where π_Ψ denotes the internal randomness of Ψ. The threshold marks Θ and π_Ψ are mutually independent and independent of the arrivals and menu randomness.

Define the response map $\mathcal{R}$ to be the unconditional endpoint-marginal vector of $M(y)$:

$$\mathcal{R}_{it}(y) := \mathsf{Pr}[T_i = t,\ i \in V(M(y))], \qquad \mathcal{R}_v(y) := \mathsf{Pr}[v \in V(M(y))].$$

The probabilities average over the arrivals, threshold marks, and internal randomness of Ψ, so $\mathcal{R}$ is deterministic once the instance and capped-benchmark data are fixed.

We verify the hypotheses of Brouwer's fixed-point theorem. First, K is nonempty, compact, and convex. Second, $\mathcal{R}(K) \subseteq K$: every output of Ψ is a matching, so

$$\mathcal{R}_{it}(y) \le p_i(t), \qquad \mathcal{R}_v(y) \le 1.$$

It remains to prove continuity. Fix $y^0 \in K$. The prices $s_u(y)$, and hence the edge scores $g_e(y)$, are continuous in y. Since there are finitely many potential edges and every Θ_e has a continuous distribution, with probability one

$$\Theta_e \neq g_e(y^0) \qquad \text{for every potential edge } e.$$

For every realization outside this null event, the thresholded graph G_y is therefore constant throughout some neighborhood of y^0. On this fixed graph the surviving weights $g_e(y)$ vary continuously with y and remain in $[\gamma, 1]$. Part (ii) of Theorem A.4 implies that the inclusion probability of every edge in the output matching is continuous in these weights. Endpoint marginals are finite sums of such edge-inclusion probabilities and are therefore continuous as well. Since they are bounded by one, dominated convergence over the arrivals and threshold marks shows that the unconditional response map $\mathcal{R}$ is continuous.

Thus $\mathcal{R}$ is a continuous self-map of the nonempty compact convex set K. Brouwer's theorem gives a fixed point

$$y^\star = \mathcal{R}(y^\star).$$

Fix one such $y^\star$ and define

$$M_C := M(y^\star).$$

The fixed point $y^\star$ is deterministic; all randomness of M_C comes from the arrivals and the auxiliary dictionary

$$\pi := (\Theta, \pi_\Psi).$$

By definition, $\mathcal{R}(y^\star)$ is exactly the endpoint-marginal vector of M_C. Since $y^\star = \mathcal{R}(y^\star)$, the coordinates of $y^\star$ are therefore precisely the endpoint marginals appearing in the statement of the lemma. Consequently,

$$\mathrm{val}(y^\star) = \frac{1}{2} \sum_{u \in \mathcal{U}_+} y^\star_u = \mathbb{E}[|M_C|], \tag{15}$$

because every matching edge contributes exactly two matched endpoints.

Endpoint-load control. Suppose that, for some $u \in \mathcal{U}_+$,

$$y^\star_u \ge (1 + \delta) q_u.$$

Then (14) gives $s_u(y^\star) = 1$. Since every endpoint price is nonnegative, every edge e incident to u has

$$g_e(y^\star) \le 0.$$

But every threshold satisfies $\Theta_e \ge \gamma > 0$, so no edge incident to u survives in $G_{y^\star}$. Hence u is never matched by $M(y^\star)$ and

$$\mathcal{R}_u(y^\star) = 0,$$

contradicting

$$y^\star_u = \mathcal{R}_u(y^\star) \ge (1 + \delta) q_u > 0.$$

Thus

$$y_u^\star < (1+\delta)q_u \qquad \text{for every } u \in \mathcal{U}_+.$$

For $u \notin \mathcal{U}_+$, both sides are zero by construction. Since $\delta = \eta^2$,

$$y_u^\star \le (1+\eta^2)q_u \qquad \text{for every endpoint } u. \tag{16}$$

This proves part (b).

Summing (16) over all endpoints and using $\sum_u q_u = 2Q$ gives

$$\mathbb{E}[|M_C|] = \mathrm{val}(y^\star) \le (1+\eta^2)Q. \tag{17}$$

Value. For this paragraph, write

$$g_e := g_e(y^\star), \qquad g(M) := \sum_{e \in M} g_e.$$

Since P is convex, Φ is concave, with

$$\frac{\partial \Phi}{\partial y_u}(y) = \frac{1}{2} - s_u(y).$$

The capped-benchmark endpoint vector q belongs to K, and

$$P(q) = 0, \qquad \mathrm{val}(q) = \frac{1}{2}\sum_u q_u = Q.$$

Hence $\Phi(q) = Q$. Applying the tangent inequality for the concave function Φ at $y^\star$, evaluated at q, gives

$$Q - \Phi(y^\star) \le \sum_u \left(\frac{1}{2} - s_u(y^\star)\right)(q_u - y_u^\star).$$

For any random matching M supported on endpoint indices in $\mathcal{U}_+$, with endpoint marginals r_u, its expected score at the prices determined by $y^\star$ is

$$\mathbb{E}[g(M)] = \sum_u \left(\frac{1}{2} - s_u(y^\star)\right) r_u. \tag{18}$$

Applying (18) to $M_C^{\star,D}$ and M_C gives

$$\mathbb{E}[g(M_C^{\star,D})] = \sum_u \left(\frac{1}{2} - s_u(y^\star)\right) q_u$$

and

$$\mathbb{E}[g(M_C)] = \sum_u \left(\frac{1}{2} - s_u(y^\star)\right) y_u^\star.$$

Therefore

$$Q - \Phi(y^\star) \le \mathbb{E}[g(M_C^{\star,D})] - \mathbb{E}[g(M_C)]. \tag{19}$$

For a realization of the arrivals and threshold marks, let $W^\star$ be the maximum matching weight of $G_{y^\star}$ under the score weights g_e. For each realization of the arrivals, benchmark randomness, and threshold marks, the retained edges of $M_C^{\star,D}$ form a feasible matching in $G_{y^\star}$ and hence have total score at most $W^\star$. Every benchmark edge removed by thresholding has score below 2γ. Therefore, pointwise,

$$g(M_C^{\star,D}) \le W^\star + 2\gamma |M_C^{\star,D}|.$$

Taking expectations and using $\mathbb{E}[|M_C^{\star,D}|] = Q$ gives

$$\mathbb{E}[g(M_C^{\star,D})] \le \mathbb{E}[W^\star] + 2\gamma Q. \tag{20}$$

Conditional on the arrivals and threshold marks, $G_{y^\star}$ and its score weights are fixed and lie in $[\gamma, 1]$. Applying part (i) of Theorem A.4 with accuracy parameter α gives

$$\mathbb{E}_{\pi_\Psi}[g(M_C) \mid T, \Theta] \ge (1-\alpha)W^\star.$$

Averaging yields

$$\mathbb{E}[g(M_C)] \geq (1-\alpha)\mathbb{E}[W^\star]. \tag{21}$$

Every edge used by M_C has score at most 1, so

$$\mathbb{E}[g(M_C)] \leq \mathbb{E}[|M_C|] \leq (1+\eta^2)Q$$

by (17). Hence

$$\mathbb{E}[W^\star] \leq \frac{1+\eta^2}{1-\alpha}Q \leq 2Q,$$

where the last inequality uses $\eta \leq 1/4$.

Substituting (20) and (21) into (19) gives

$$\begin{aligned} Q - \Phi(y^\star) &\leq \alpha\,\mathbb{E}[W^\star] + 2\gamma Q \\ &\leq 2\alpha Q + 2c_{\mathrm{lca}}\alpha Q \\ &= 2(1+c_{\mathrm{lca}})\alpha Q = \frac{\eta}{2}Q. \end{aligned}$$

Thus

$$\Phi(y^\star) \geq (1-\eta/2)Q.$$

Since $P(y^\star) \geq 0$, (15) gives

$$\mathbb{E}[|M_C|] = \mathrm{val}(y^\star) \geq \Phi(y^\star) \geq (1-\eta)Q,$$

proving part (a).

Combining this with $Q \geq (1-\eta)x(C_\tau)$ from Lemma B.1 also gives

$$\mathbb{E}[|M_C|] \geq (1-2\eta)x(C_\tau). \tag{22}$$

Locality. The witness is computed from the capped graph using the locally owned threshold marks Θ and the internal dictionary π_Ψ, all independent of the arrivals and menu randomness. The fixed point $y^\star$ and the capped-benchmark marginals are deterministic parameters of the rule.

The thresholded graph has maximum degree at most

$$D = \Lambda_\eta(\tau).$$

The weighted-matching rule is invoked with accuracy α, so Theorem A.4 gives query profile

$$D^{a_\alpha}, \qquad a_\alpha = C_{\mathrm{lca}}\alpha^{-2}\log^2\frac{e}{\alpha}.$$

Simulating the threshold-filtering layer from the capped graph costs at most one additional factor D. Since α is a fixed universal constant multiple of η, choosing C_{loc} sufficiently large gives

$$a_\alpha + 1 \leq \widehat{a}_\eta = C_{\mathrm{loc}}\eta^{-2}\log^2\frac{e}{\eta}.$$

Consequently

$$R^+, R^- \leq \Lambda_\eta(\tau)^{\widehat{a}_\eta}. \tag{23}$$

This proves part (c) and completes the proof. □

Presence of the crucial witness. Every edge of M_C is crucial. Conditional on the arrival profile and on M_C, the menu randomness is independent of the witness, so (D1) gives

$$\mathsf{Pr}[e \notin H \mid T, M_C] \leq e^{-k\tau} \qquad \text{for every } e \in M_C.$$

Hence, whenever $k\tau \geq \log(1/\eta)$,

$$\mathbb{E}[|M_C \cap H| \mid T, M_C] \geq (1-\eta)|M_C|.$$

Averaging and applying Lemma B.2,

$$\mathbb{E}[|M_C \cap H|] \geq (1-\eta)\mathbb{E}[|M_C|] \geq (1-\eta)^2 Q. \tag{24}$$

Together with $Q \geq (1-\eta)x(C_\tau)$, this also gives

$$\mathbb{E}[|M_C \cap H|] \geq (1-3\eta)x(C_\tau).$$

For VarOpt menus, the same threshold condition implies $1/k \leq \tau$, so every crucial edge, and hence every edge of M_C, is retained deterministically.

Interface to light completion. From this point on, we use only four properties of M_C: its near-optimal value, the pointwise endpoint-load control

$$y_u \le (1+\eta^2)q_u,$$

independence from the menus conditional on the arrival profile, and the free-status query profile (23). The pointwise load bound lets us transfer the benchmark light plan to the marginal endpoint capacities left by M_C with only $O(\eta Z)$ loss, while the independence and locality properties provide the variance and bias bounds needed to recover that mass from sampled light edges.

C Fitting the light benchmark to the witness

Fix $0 < \tau^- < \tau$ and recall the decomposition $C_\tau, I_{\tau^-,\tau}, L_{\tau^-}$ from Equation (3). Let

$$a_{it} := 1 - \frac{y_{it}}{p_i(t)} \quad (p_i(t) > 0), \qquad a_v := 1 - y_v$$

be the marginal endpoint availabilities left by M_C.

The benchmark light plan is feasible in the residual capacities left by the capped benchmark $M_C^{\star,D}$, whereas we need a plan feasible in the capacities left by M_C. The pointwise load control of Lemma B.2 shows that these two sets of capacities are close, except possibly at endpoints with very small residual availability. We discard such endpoints and then use one common rescaling.

Lemma C.1 (Residual fitting)**.** *There is a deterministic subplan x^0 supported on L_{τ^-} such that*

$$|x^0| \ge x(L_{\tau^-}) - C\eta Z, \tag{25}$$

and, writing $z^0_{itv} := x^0_{itv}/p_i(t)$,

$$0 \le z^0_{itv} \le \tau^-, \qquad \sum_v z^0_{itv} \le a_{it}, \qquad \sum_{i,t} x^0_{itv} \le a_v.$$

Every endpoint in the support of x^0 has residual availability at least η.

Proof. For an endpoint u, write $b_u = p_i(t)$ if $u = (i,t)$ and $b_u = 1$ if $u = v$, so $b_u - y_u = a_u b_u$. Let ℓ_u be the load of the benchmark light plan at u. The light benchmark and the capped crucial benchmark are disjoint parts of the same matching, so

$$\ell_u + q_u \le b_u. \tag{26}$$

Delete all light edges incident to endpoints with $a_u < \eta$. At each such endpoint, the witness bound gives

$$(1-\eta)b_u < y_u \le (1+\eta^2)q_u, \qquad \ell_u \le b_u - q_u \le \frac{\eta+\eta^2}{1-\eta}\, q_u.$$

Since the q-loads sum to Q on each side, the total deleted mass is at most $2(\eta+\eta^2)Q/(1-\eta) = O(\eta Z)$.

Let $\bar{x}$ be the remaining plan. On its support, $b_u - y_u \ge \eta b_u$, and hence

$$\begin{aligned} b_u - q_u &= (b_u - y_u) + (y_u - q_u) \\ &\le (b_u - y_u) + \eta^2 b_u \le (1+\eta)(b_u - y_u). \end{aligned}$$

Thus $x^0 := \bar{x}/(1+\eta)$ fits every residual marginal capacity $b_u - y_u$. This scaling loses at most ηZ additional mass. Dividing the online-state constraints by $p_i(t)$ gives the claimed row bounds; the offline constraints give the column bounds. Finally, deletion and scaling preserve $0 \le z^0_{itv} \le z_{itv} \le \tau^-$, and every remaining endpoint has availability at least η. □

D SAMPLING THE LIGHT RESIDUAL

The fitted light plan x^0 from Lemma C.1 satisfies the marginal capacity constraints left by M_C. We now recover its mass, up to $O(\eta Z)$ expected loss, as a feasible fractional matching $\widehat{Y}$ on retained light edges whose endpoints are free in M_C. Combining $\widehat{Y}$ with $M_C \cap H$ completes the fractional matching certificate supported on H.

We begin with moment bounds for the menu weights. We then use the locality of M_C to bound variances and conditioning bias for weighted sums of free-status indicators. These estimates allow us to control the mass lost when making the light-edge loads feasible.

Menu weights. The analysis of the given derived menu rule uses only (D1) and the nonpositive covariances in (D2). For a state-edge in the support of x^0, the Horvitz–Thompson weight

$$W_{iv} := z^0_{itv} \frac{\mathbf{1}\{v \in S_i\}}{\mathsf{Pr}[v \in S_i \mid T_i = t]}$$

has conditional mean z^0_{itv}. Whenever $1/k \le \tau^-$, using (2), $z^0_{itv} \le z_{itv}$, and $\max\{z^0_{itv}, 1/k\} \le \tau^-$, we obtain

$$\mathbb{E}[W^2_{iv} \mid T_i = t] = \frac{(z^0_{itv})^2}{\mathsf{Pr}[v \in S_i \mid T_i = t]} \le \frac{8}{5} z^0_{itv} \max\{z^0_{itv}, 1/k\} \le 4\tau^- z^0_{itv}.$$

Conditional on $T_i = t$, distinct weights from the same menu have nonpositive covariance. Conditional on the arrival profile T, menus of different arrivals are independent and are independent of M_C.

D.1 DECORRELATION FROM LOCALITY.

The fixed-input correlation bound (Lemma A.3) controls dependence among free-status indicators conditional on the arrival states. The following theorem uses capping stability to account also for the arrival randomness, bounding variances of weighted sums and the bias caused by conditioning on a single arrival state.

Theorem D.1 (In–out decorrelation transfer)**.** *Let M_C be a matching of $\mathcal{G}^D_C(T)$ computed by a natural local rule whose free-status queries have query profile (R^+, R^-) for every arrival realization T. The input consists of the capped graph with each arrival vertex i labelled by T_i, together with deterministic instance data. The label T_i is revealed only when i is probed. The shared dictionary π is independent of T and of the menu randomness. Put*

$$F_v := \mathbf{1}\{v \notin V(M_C)\}, \qquad F_i := \mathbf{1}\{i \notin V(M_C)\}.$$

There is a universal constant $C_{\mathrm{dec}} \ge 1$ such that, with

$$\Gamma := C_{\mathrm{dec}}\big(1 + d^2 D(D + d + 1)\big) R^+ R^-,$$

the following bounds hold for all deterministic coefficient families $(h_v)_v$ and $(h_{it})_{i,t}$:

$$\mathsf{Var}\left(\sum_v h_v F_v\right) \le \Gamma \sum_v h_v^2, \qquad \mathsf{Var}\left(\sum_{i,t} h_{it} \mathbf{1}\{T_i = t\} F_i\right) \le \Gamma \sum_{i,t} p_i(t) h_{it}^2.$$

The first inequality remains valid after conditioning on any single state $T_i = t$ with $p_i(t) > 0$. In addition, for every deterministic coefficient array satisfying $0 \le h_{itv} \le h_{\max}$,

$$\sum_{i,t} p_i(t) \left| \sum_v h_{itv} \big(\mathsf{Pr}[F_v = 1 \mid T_i = t] - \mathsf{Pr}[F_v = 1]\big) \right| \le C_{\mathrm{dec}} d^2 D R^- h_{\max} x(C_\tau).$$

The proof is given in Section G. We apply the theorem in Lemma D.2. Together with the menu moment bounds above, its variance bounds control fluctuations in the sampled light-edge loads, while its bias bound controls the error caused by conditioning on an arrival state.

For the witness of Lemma B.2,

$$R^+, R^- \le \Lambda_\eta(\tau)^{\widehat{a}_\eta}, \qquad d \le D = \Lambda_\eta(\tau).$$

Consequently, choosing a sufficiently large universal constant C_4 and defining

$$A_\eta := C_4 \eta^{-2} \log^2 \frac{C_4}{\eta}, \tag{27}$$

we obtain

$$\Gamma \le \Lambda_\eta(\tau)^{A_\eta}. \tag{28}$$

D.2 Light completion

We now combine the menu moment bounds with Theorem D.1 to construct the light completion. For this construction, we require

$$4\tau^- \le \eta^{C_{\text{light}}} \Lambda_\eta(\tau)^{-(A_\eta + C_{\text{light}})}, \tag{29}$$

where $C_{\text{light}} \ge 6$ is a universal constant. We choose thresholds satisfying this condition in Section E.

Lemma D.2 (Light completion)**.** *Assume $1/k \le \tau^-$ and (29). Let x^0 be the deterministic light subplan given by Lemma C.1, so that*

$$0 \le z^0_{itv} := \frac{x^0_{itv}}{p_i(t)} \le \tau^-, \qquad \sum_v z^0_{itv} \le a_{it}, \qquad \sum_{i,t} x^0_{itv} \le a_v,$$

and every endpoint in its support has residual availability at least η. Then there is a random feasible fractional matching $\widehat{Y}$, supported on retained edges from the support of x^0 and using only vertices free in M_C, such that

$$\mathbb{E}[|\widehat{Y}|] \ge |x^0| - 3\eta Z.$$

Proof. Set

$$\ell_{it} := \sum_v z^0_{itv} \le a_{it}, \qquad \ell_v := \sum_{i,t} x^0_{itv} \le a_v.$$

Both $\sum_{i,t} p_i(t)\ell_{it}$ and $\sum_v \ell_v$ equal $|x^0| \le Z$. Throughout the proof, sums and quotients involving availability denominators are restricted to the support of x^0; there all availabilities are at least η. All other edge weights and endpoint loads are zero.

Preload moments. Recall the menu weights W_{iv} defined above and set $b := 4\tau^-$. Their moment bounds are

$$\mathbb{E}[W_{iv} \mid T_i = t] = z^0_{itv}, \qquad \mathbb{E}[W^2_{iv} \mid T_i = t] \le b z^0_{itv}.$$

Conditional on $T_i = t$, distinct weights from the same menu have nonpositive covariance by (D2), and the menu of i is independent of the witness. Define the preloads

$$P_i := \sum_v W_{iv} \frac{F_v}{a_v}, \qquad P_v := \sum_i W_{iv} \frac{F_i}{a_{iT_i}}.$$

For a row, condition first on $T_i = t$ and its menu. The conditional offline variance bound of Theorem D.1 gives

$$\mathsf{Var}(P_i \mid T_i = t, \text{menu}_i) \le \Gamma \sum_v \frac{W^2_{iv}}{a^2_v}.$$

The expectation of this bound over the menu is at most $\Gamma b \ell_{it}/\eta^2$. The conditional mean is $\sum_v W_{iv} \Pr[F_v = 1 \mid T_i = t]/a_v$; its menu variance is at most $b\ell_{it}/\eta^2$ by (D2).

For a column, independently pre-sample a latent menu S_{it} for every state (i,t), independently of T and the witness dictionary, and use S_{iT_i} as the realized menu. Let W_{itv} denote its latent Horvitz–Thompson weight. Conditional on this table,

$$P_v = \sum_{i,t} \frac{W_{itv}}{a_{it}} \mathbf{1}\{T_i = t\} F_i, \qquad \mathbb{E}[P_v \mid \text{menu table}] = \sum_{i,t} p_i(t) W_{itv}.$$

The mean identity uses $\Pr[F_i = 1 \mid T_i = t] = a_{it}$ and independence of the table from T and the witness. The online-state variance bound of Theorem D.1, averaged over the table, is at most

$\Gamma b\ell_v/\eta^2$. Independence of the latent menus bounds the variance of the displayed conditional mean by $\sum_{i,t} p_i(t)^2 \mathbb{E}[W_{itv}^2] \le b\ell_v$. Thus the law of total variance, $\Gamma \ge 1$, and $\eta \le 1$ give

$$\mathsf{Var}(P_i \mid T_i = t) \le \xi \ell_{it}, \qquad \mathsf{Var}(P_v) \le \xi \ell_v, \qquad \xi := \frac{2\Gamma b}{\eta^2}. \tag{30}$$

The column mean is exact: $\mathbb{E}[P_v] = \ell_v$. For the row means, write

$$\delta_{it} := \left|\mathbb{E}[P_i \mid T_i = t] - \ell_{it}\right|, \qquad B_{\text{bias}} := \sum_{i,t} p_i(t)\delta_{it}.$$

Applying the bias bound of Theorem D.1 with $h_{itv} = z^0_{itv}/a_v \le \tau^-/\eta$ gives

$$B_{\text{bias}} \le C_{\text{dec}} d^2 D R^- \frac{\tau^-}{\eta} x(C_\tau) \le \frac{\Gamma b}{4\eta} Z = \frac{\xi\eta}{8} Z. \tag{31}$$

Here the definition of Γ and $R^+ \ge 1$ justify the second inequality. By (28), (29), and $C_{\text{light}} \ge 6$, using $D = \Lambda_\eta(\tau) \ge 2$, we have

$$\Gamma b \le \frac{\eta^6}{2}, \qquad \xi \le \eta^4, \qquad B_{\text{bias}} \le \frac{\eta^5}{8} Z. \tag{32}$$

Value before removing overload. Define the availability-normalized weights

$$U_{iv} := W_{iv} \frac{F_i F_v}{a_{iT_i} a_v}.$$

Reading their value by columns and using the exact column means gives

$$\mathbb{E}[|U|] - |x^0| = \sum_v \frac{\mathrm{Cov}(F_v, P_v)}{a_v}.$$

Since $\mathsf{Var}(F_v) = a_v(1 - a_v) = a_v y_v$, Cauchy–Schwarz gives

$$\begin{aligned} \left|\mathbb{E}[|U|] - |x^0|\right| &\le \sum_v \sqrt{\frac{\xi \ell_v y_v}{a_v}} \\ &\le \sqrt{\frac{\xi}{\eta}} \sqrt{\left(\sum_v \ell_v\right)\left(\sum_v y_v\right)} \\ &\le \sqrt{\frac{(1+\eta^2)\xi}{\eta}}\, Z \le \eta Z. \end{aligned}$$

We used $\sum_v y_v \le (1+\eta^2)Q \le (1+\eta^2)Z$ from Lemma B.2, followed by (32) and $\eta \le 1/4$. Now create slack by setting

$$\widetilde{Y}_{iv} := \frac{U_{iv}}{1+\eta}, \qquad R_i := \sum_v \widetilde{Y}_{iv}, \qquad L_v := \sum_i \widetilde{Y}_{iv}.$$

Then

$$\mathbb{E}[|\widetilde{Y}|] \ge |x^0| - 2\eta Z. \tag{33}$$

Direct control of overload. We use a scalar estimate that requires no independence between a preload and its free-status indicator. Let F be Bernoulli with mean $a > 0$, let P have mean μ and variance σ^2, and let $\ell \le a$. With $D_P := P - \mu$,

$$\begin{aligned} \left(\frac{FP}{(1+\eta)a} - 1\right)_+ &= \frac{F}{(1+\eta)a}\big(P - (1+\eta)a\big)_+ \\ &\le \frac{F}{(1+\eta)a}\big((D_P - \eta a)_+ + (\mu - \ell)_+\big). \end{aligned}$$

Using $(x-s)_+ \le x^2/(4s)$ for $s>0$, $F \le 1$, and $\mathbb{E}[F]=a$, we obtain

$$\mathbb{E}\left[\left(\frac{FP}{(1+\eta)a}-1\right)_+\right] \le (\mu-\ell)_+ + \frac{\sigma^2}{4\eta a^2}. \tag{34}$$

For a row, apply this conditionally on $T_i=t$, with $F=F_i$, $P=P_i$, $a=a_{it}$, and $\ell=\ell_{it}$. For a column, apply it unconditionally with $F=F_v$, $P=P_v$, $a=a_v$, and $\ell=\ell_v$; its mean error is zero. Summing over row states with weights $p_i(t)$ and over columns, (30) and (31) give

$$\begin{aligned}\mathbb{E}\left[\sum_i (R_i-1)_+ + \sum_v (L_v-1)_+\right] &\le B_{\text{bias}} + \frac{\xi}{2\eta^3}|x^0| \\ &\le \left(\frac{\eta^5}{8}+\frac{\eta}{2}\right) Z \le \eta Z.\end{aligned}$$

In each realization, scale every overloaded row of $\widetilde{Y}$ down to load one, then do the same for the remaining overloaded columns. The first step loses exactly $\sum_i (R_i-1)_+$ and can only reduce column loads, so the second loses at most $\sum_v (L_v-1)_+$. Column scaling preserves row feasibility. The resulting $\widehat{Y}$ is feasible, is supported on retained light edges, and uses only vertices free in M_C. Together with (33), this gives

$$\mathbb{E}[|\widehat{Y}|] \ge |x^0| - 3\eta Z.$$

□

E CHOOSING THRESHOLDS AND COMPLETING THE PROOF

We now choose $\tau^- < \tau$ so that (29) holds and the intermediate band has small benchmark mass. Let

$$m := \lceil 1/\eta \rceil, \qquad C_3 := C_{\text{light}} + 1, \qquad B_\eta := A_\eta + C_3.$$

Starting from $\rho_0 = 1/4$, define

$$\rho_{j+1} := \tfrac{1}{2}\eta^{C_3}\Lambda_\eta(\rho_j)^{-B_\eta}, \qquad 0 \le j < m. \tag{35}$$

Since $\Lambda_\eta(\rho) \ge 2e/\rho$ and $B_\eta \ge 1$, we have

$$\rho_{j+1} \le \frac{\eta^{C_3}}{4e}\rho_j \le \eta\rho_j.$$

Thus the bands

$$\mathcal{B}_j := \{(i,t,v) : \rho_{j+1} < z_{itv} < \rho_j\}, \qquad 0 \le j < m,$$

are disjoint. Their total benchmark mass is at most Z, so for some $j < m$,

$$x(\mathcal{B}_j) \le Z/m \le \eta Z.$$

Fix such a j and set

$$\tau := \rho_j, \qquad \tau^- := \rho_{j+1}.$$

Then

$$4\tau^- = 2\eta^{C_3}\Lambda_\eta(\tau)^{-B_\eta} \le \eta^{C_{\text{light}}}\Lambda_\eta(\tau)^{-(A_\eta+C_{\text{light}})},$$

where the inequality uses $2\eta/\Lambda_\eta(\tau) \le 1$. Thus (29) holds. Moreover $\rho_{j+1} \le \eta\rho_j$ gives

$$\frac{\tau}{\tau^-} \ge \frac{1}{\eta} \ge \log\frac{1}{\eta}. \tag{36}$$

Hence, whenever $1/k \le \tau^-$, we also have $k\tau \ge \log(1/\eta)$, as required for (24).

The complete certificate is the fractional matching

$$Y := \mathbf{1}_{M_C \cap H} + \widehat{Y},$$

where $\mathbf{1}_{M_C \cap H}$ is the edge-incidence vector of the retained crucial witness. Since $\widehat{Y}$ uses only retained light edges and vertices free in M_C, Y is feasible and supported on H. Bipartite matching integrality gives $\nu(H) \geq |Y|$ in every realization. Using Lemmas B.2, C.1 and D.2 and (24),

$$\begin{aligned}\mathbb{E}[\nu(H)] &\geq \mathbb{E}[|M_C \cap H|] + \mathbb{E}[|\widehat{Y}|] \\ &\geq x(C_\tau) + x(L_{\tau^-}) - C\eta Z \\ &= Z - x(I_{\tau^-,\tau}) - C\eta Z \\ &\geq (1 - C'\eta)Z.\end{aligned}$$

Choosing the constant c_0 in $\eta = c_0\varepsilon$ so that $C'c_0 \leq 1$ proves the $(1-\varepsilon)$ value guarantee.

It remains only to bound the menu size. Since the chosen $\tau^- = \rho_{j+1} \geq \rho_m$, it suffices that

$$k \geq \left\lceil \rho_m^{-1} \right\rceil.$$

We show that the value of k_ε in Theorem 3.1 meets this requirement. Put $u_j := \log(1/\rho_j)$. From (35) and

$$\Lambda_\eta(\rho) \leq C\left(\rho^{-1} + \log\frac{1}{\eta}\right)$$

we obtain

$$u_{j+1} \leq B_\eta\left(u_j + C\log\frac{C}{\eta}\right).$$

Thus

$$u_m \leq CB_\eta^m \log\frac{C}{\eta},$$

and, since

$$B_\eta = O\left(\eta^{-2}\log^2\frac{1}{\eta}\right), \qquad m = O(\eta^{-1}),$$

we have

$$\log u_m \leq C\eta^{-1}\log\frac{C}{\eta}.$$

Consequently, for a sufficiently large universal C,

$$\left\lceil \rho_m^{-1} \right\rceil \leq \left\lceil \exp\left(\exp\left(C\varepsilon^{-1}\log\frac{C}{\varepsilon}\right)\right)\right\rceil = k_\varepsilon,$$

as required.

Where the double exponential comes from. The locality/decorrelation argument costs a polynomial $\Lambda_\eta(\tau)^{A_\eta}$ with $A_\eta = O(\eta^{-2}\log^2(1/\eta))$. Threshold selection then descends through $m = \Theta(1/\eta)$ separated bands, each transition raising the previous scale to a power $B_\eta = A_\eta + O(1)$. Hence $\log(1/\rho_m)$ is exponential in $\eta^{-1}\log(1/\eta)$, and converting back from the logarithmic scale produces the double exponential. Improving the universal rate therefore seems to require either milder dependence of the light completion on the crucial threshold or a way to avoid the $\Theta(1/\eta)$-level threshold ladder.

F LOCAL COMPUTATION MACHINERY

This appendix supplies the technical LCA arguments behind Lemma A.3 and Theorem A.4. We first state an elementary composition rule and prove the fixed-input correlation bound. We then state the truncated random-greedy MIS primitive and use it to construct the weighted matching rule. Continuous edge marginals are obtained by an internal perturbation of the weights.

Lemma F.1 (Composition of query profiles)**.** *Let $\mathcal{R}$ be a natural local rule with query profile (P^+, P^-) and dictionary π_0. Suppose an outer natural local rule, using an independent fresh dictionary π_1, accesses $\mathcal{R}$ only by issuing root queries. Conditional on every realization of π_0, assume that*

- *each output-root execution issues at most B^+ queries to $\mathcal{R}$;*

- *for every root x of $\mathcal{R}$, the expected number, over π_1, of queries to $\mathcal{R}$ at x issued by all output-root executions is at most B^-.*

Then the composed local rule has query profile at most

$$(B^+P^+,\ B^-P^-).$$

The same conclusion holds, up to adding the corresponding direct-probe bounds, if the outer rule also probes the base graph directly.

Proof. For an output root u, let $C_{u,x}$ be the number of queries that its execution issues to $\mathcal{R}$ at root x. Pointwise,

$$|\mathrm{Read}_{\mathrm{comp}}(u)| \le \sum_x C_{u,x}\,|\mathrm{Read}_{\mathcal{R}}(x)| \le B^+P^+.$$

For the reverse direction, fix a base vertex w and condition on π_0. Then the read sets of $\mathcal{R}$ are fixed, and

$$\begin{aligned}\mathbb{E}_{\pi_1}\left[\left|\{u : w \in \mathrm{Read}_{\mathrm{comp}}(u)\}\right|\right] &\le \sum_x \mathbb{E}_{\pi_1}\left[\sum_u C_{u,x}\right]\mathbf{1}\{w \in \mathrm{Read}_{\mathcal{R}}(x)\}\\ &\le B^- \sum_x \mathbf{1}\{w \in \mathrm{Read}_{\mathcal{R}}(x)\}.\end{aligned}$$

Averaging over π_0 gives the bound B^-P^-. □

Lemma A.3 (Fixed-input in–out correlation (Azarmehr et al., 2025, Lemmas 3.7–3.8))**.** *Fix a graph G and a natural local rule whose randomness is grouped into independent private tapes at their owner vertices. Suppose its query profile is (R^+, R^-). For roots u, u', let*

$$\delta(u,u') := \mathsf{Pr}[\mathrm{Read}(u) \cap \mathrm{Read}(u') \neq \varnothing].$$

Then

$$\sum_{u' \neq u} \delta(u,u') \le R^+R^- \tag{6}$$

for every u. Moreover, if the query outputs are binary $X_u \in \{0,1\}$, then

$$|\,\mathrm{Cov}(X_u, X_{u'})| \le C\,\delta(u,u') \tag{7}$$

for a universal constant C. Consequently, for deterministic coefficients (a_u),

$$\mathsf{Var}\left(\sum_u a_u X_u\right) \le CR^+R^- \sum_u a_u^2. \tag{8}$$

Proof of Lemma A.3. Group all independent dictionary keys owned by one base vertex into that vertex's private tape. The hypotheses then match the private-tape LCA model of Azarmehr et al. (2025). Their Lemma 3.8 bounds the expected size of the correlated set of any root by the product of the expected out- and in-query bounds. Since our out-query bound is pointwise, this gives (6). Their Lemma 3.7, specialized to two roots, couples their two outputs to independent variables with total-variation error at most a universal constant times $\delta(u,u')$; for Bernoulli outputs this implies (7).

Finally, expand the variance. Using $\mathsf{Var}(X_u) \le 1$ and $2|a_ua_{u'}| \le a_u^2 + a_{u'}^2$, and then (6), gives

$$\begin{aligned}\mathsf{Var}_\pi\left(\sum_u a_u X_u\right) &\le \sum_u a_u^2 + C\sum_{u \neq u'} |a_ua_{u'}|\delta(u,u')\\ &\le C(1 + R^+R^-)\sum_u a_u^2.\end{aligned}$$

Every query reads its own root, so $R^+, R^- \ge 1$; absorbing the first term proves (8). □

Greedy independent set. Assign independent uniform priorities to the vertices, breaking ties by a fixed ordering of their identifiers. Write $w \prec_\pi u$ for the resulting total order. The *random-greedy* independent set $I = I(G, \pi)$ processes vertices in this order and inserts a vertex exactly when none of its previously processed neighbors was inserted. It is maximal, and

$$u \in I \iff \text{no neighbor } w \prec_\pi u \text{ of } u \text{ belongs to } I.$$

The membership recursion visits earlier neighbors in increasing priority order and stops at the first affirmative answer. It follows strictly decreasing paths and hence terminates, but its depth need not be bounded independently of graph size. The following primitive truncates this recursion to obtain a deterministic out-query bound while deleting only a small expected fraction of the exact random-greedy independent set.

Theorem F.2 (Truncated random-greedy MIS (Azarmehr et al., 2025, Definition 6.7 and Claim 6.8))**.** *For every $\widehat{\Delta} \geq 1$ and $\zeta \in (0, 1)$ there is a natural local rule on graphs of maximum degree at most $\widehat{\Delta}$, computing an independent set $I_\zeta = I_\zeta(G, \pi)$ such that*

(i) for every dictionary realization π,

$$I_\zeta(G, \pi) \subseteq I(G, \pi),$$

where $I(G, \pi)$ is the exact random-greedy maximal independent set using the same priorities;

(ii) for every G of maximum degree at most $\widehat{\Delta}$,

$$\mathbb{E}_\pi[|I(G, \pi) \setminus I_\zeta(G, \pi)|] \leq \zeta\, \mathbb{E}_\pi[|I(G, \pi)|];$$

(iii) the out-query complexity is

$$R^+ = O(\widehat{\Delta}^3/\zeta);$$

(iv) the expected in-query complexity is

$$R^- = O(\widehat{\Delta}^2).$$

In particular, $I_\zeta(G, \pi)$ is an independent set for every G and every π.

Proof. Simulate the exact membership recursion from the query root, aborting the entire query and returning *false* if it exceeds $C_{\mathrm{mis}}\widehat{\Delta}^2/\zeta$ recursive calls. Otherwise return its exact answer. Here C_{mis} is a sufficiently large universal constant. This is the truncation in Azarmehr et al. (2025, Definition 6.7). Every affirmative answer is correct, so $I_\zeta \subseteq I$ for every dictionary realization, proving (i).

By Azarmehr et al. (2025, Lemma 6.6), querying the exact recursion once from every root makes $O(n\widehat{\Delta})$ recursive calls in expectation. Consequently, choosing C_{mis} sufficiently large gives

$$\mathbb{E}_\pi[|\{u : \text{the query at } u \text{ truncates}\}|] \leq \frac{\zeta n}{\widehat{\Delta} + 1} \leq \zeta\, \mathbb{E}_\pi[|I|],$$

where the last inequality uses maximality of I. Every vertex of $I \setminus I_\zeta$ is a truncating root, proving (ii).

Each recursive call requires at most $O(\widehat{\Delta})$ vertex probes to read the current vertex and its neighbors' priorities. The call budget therefore gives (iii). For (iv), distinguish these probes from recursive calls. Lemma 6.6 also bounds the expected total number of calls to any fixed vertex by $O(\widehat{\Delta})$, over one execution from every root. A vertex w is probed only during a call to w or one of its neighbors. Summing the call bound over these at most $\widehat{\Delta} + 1$ vertices gives $O(\widehat{\Delta}^2)$. Truncation can only remove calls and probes. □

Bounded augmentations. For a matching M, an *augmentation* is an M-alternating path or cycle P such that $M \triangle P$ is again a matching. Its gain is

$$g_M(P) := w(P \setminus M) - w(P \cap M).$$

A k_0-augmentation contains at most k_0 nonmatching edges. It therefore contains $O(k_0)$ vertices.

Lemma F.3 (Comparison lemma (Pettie & Sanders, 2004, Theorem 2.1))**.** *Let M be a matching of a weighted graph and $k_0 \geq 1$. There is a vertex-disjoint family $\mathcal{A}$ of k_0-augmentations such that*

$$g_M(\mathcal{A}) \geq \frac{k_0+1}{2k_0+1}\left(\frac{k_0}{k_0+1}\,\mathrm{OPT}_w - w(M)\right).$$

Theorem A.4 (Local weighted matching rule)**.** *There are universal constants $c_{\mathrm{lca}}, C_{\mathrm{lca}} > 0$ such that the following holds. Fix $\eta \in (0, 1/4)$ and $\gamma \in (0, 1]$ with $\gamma \geq c_{\mathrm{lca}}\eta$. There is a natural local rule Ψ with the following properties for every finite graph G of maximum degree at most $\Delta \geq 2$ and every edge-weight vector $w \in [\gamma, 1]^{E(G)}$. The rule computes a matching*

$$M = \Psi(G, \pi; w),$$

where π comprises all internal randomness, represented by independent dictionary marks with local owners as in Definition A.1.

(i) Approximation and locality. *The output satisfies*

$$\mathbb{E}_\pi[w(M)] \geq (1-\eta)\,\mathrm{OPT}_w(G).$$

Its query profile satisfies

$$R^+, R^- \leq \Delta^{a_\eta}, \qquad a_\eta := C_{\mathrm{lca}}\eta^{-2}\log^2\frac{e}{\eta}.$$

(ii) Continuous edge marginals. *For fixed G, η, γ and every $e \in E(G)$, the map*

$$w \longmapsto \Pr_\pi[e \in \Psi(G, \pi; w)]$$

is continuous on $[\gamma, 1]^{E(G)}$. The expectation and probability are over π, with the graph and input weights fixed.

Proof. We first analyze the construction with a working accuracy $\alpha \in (0, 1/4)$ and a lower weight bound b satisfying

$$\alpha \leq b \leq 1.$$

This gives a base rule $\Psi^{\mathrm{base}}_{\alpha,b}$. We will then choose the internal accuracy and perturb the input weights to obtain a single rule with both properties of the theorem. Throughout the analysis of the base rule, G and w are fixed, $w_e \in [b, 1]$, and all randomness belongs to the priority dictionary π_0.

Write

$$W := \mathrm{OPT}_w(G).$$

If G has no edges, the empty matching satisfies all claims. Henceforth assume $W > 0$.

The construction. Fix

$$k_0 = \left\lceil \frac{C_0}{\alpha} \right\rceil, \qquad \zeta = c_{\mathrm{mis}}\alpha \leq \frac{1}{4},$$

where C_0 is a sufficiently large universal constant and $c_{\mathrm{mis}} > 0$ is a sufficiently small universal constant. Recall that a k_0-augmentation has at most k_0 nonmatching edges and therefore has $O(k_0)$ vertices. Its gain is at most k_0.

The algorithm starts from the empty matching and performs a fixed number of sweeps. Each sweep processes the dyadic gain buckets

$$[2^j\alpha b,\ 2^{j+1}\alpha b), \qquad j = 0, \ldots, J-1,$$

in decreasing order of j, where

$$J := 1 + \left\lfloor \log_2 \frac{k_0}{\alpha b} \right\rfloor = O\Big(\log\frac{e}{\alpha}\Big).$$

The last bound uses $b \geq \alpha$.

At a bucket stage, consider all currently valid k_0-augmentations whose gains lie in that bucket. Form their conflict graph, joining two candidates when they share a base-graph vertex. Using fresh independent priorities, let I be the exact random-greedy maximal independent set of this graph, and let $I_\zeta \subseteq I$ be the truncated set supplied by Theorem F.2. Apply all augmentations in I_ζ simultaneously. They are vertex-disjoint, so the result is a matching. All applied augmentations have positive gain.

Progress in one sweep. Let M_s be the matching at the start of sweep s, and define

$$\mathrm{gap}(M) := \frac{k_0}{k_0+1}W - w(M).$$

By Lemma F.3, when $\mathrm{gap}(M_s) > 0$ there is a vertex-disjoint family $\mathcal{A}^\star$ of k_0-augmentations with

$$g_{M_s}(\mathcal{A}^\star) \geq \frac{1}{2}\,\mathrm{gap}(M_s).$$

When $\mathrm{gap}(M_s) \leq 0$, the same inequality holds with $\mathcal{A}^\star = \varnothing$. Discarding members of nonpositive gain can only increase the total gain, so assume every member has positive gain.

Each member then contains a nonmatching edge. Choosing one such edge from each member gives a matching, and therefore

$$|\mathcal{A}^\star| \leq \nu(G) \leq \frac{W}{b}.$$

Discard from this comparison family all augmentations of gain below αb. Their total gain is at most

$$\alpha b\,|\mathcal{A}^\star| \leq \alpha W.$$

Thus the surviving comparison family has total gain at least

$$\frac{1}{2}\,\mathrm{gap}(M_s) - \alpha W.$$

This family is used only in the analysis; the algorithm considers all valid augmentations in each bucket.

Fix all randomness in the sweep. For a surviving comparison augmentation A, if an applied augmentation from an earlier bucket meets A, charge A to the first such augmentation. Otherwise no previous update has touched a vertex of A. Consequently, when the bucket containing $g_{M_s}(A)$ is processed, A is still valid and has the same gain. By maximality of the exact greedy set I in that bucket, A either belongs to I or meets a member of I. Charge A to such a member.

In either case, the augmentation receiving the charge has gain at least $g_{M_s}(A)/2$. Moreover, each candidate contains $O(k_0)$ vertices, whereas the comparison augmentations are vertex-disjoint. Hence each candidate receives at most $O(k_0)$ charges.

Let

$$G_s := w(M_{s+1}) - w(M_s)$$

be the actual gain of the sweep. Let D_s be the sum, over its bucket stages, of the gains of candidates in $I \setminus I_\zeta$. The charging argument gives the pointwise bound

$$\frac{1}{2}\,\mathrm{gap}(M_s) - \alpha W \leq Ck_0(G_s + D_s).$$

Equivalently, for universal constants $c, C > 0$,

$$G_s \geq \frac{c}{k_0}\,\mathrm{gap}(M_s) - \frac{C\alpha}{k_0}W - D_s. \tag{37}$$

Loss from truncation. Consider a single bucket stage, conditional on its history before the fresh priority dictionary is drawn. The candidate graph and all candidate gains are then fixed. Write

$$G_j^\star := \sum_{A\in I} g(A), \qquad G_j := \sum_{A\in I_\zeta} g(A), \qquad D_j := \sum_{A\in I\setminus I_\zeta} g(A).$$

Pointwise, $G_j^\star = G_j + D_j$. Since gains within a bucket differ by at most a factor two, Theorem F.2(ii) implies

$$\mathbb{E}_{\pi_0}[D_j \mid \mathrm{past}] \leq 2\zeta\,\mathbb{E}_{\pi_0}[G_j^\star \mid \mathrm{past}].$$

Using $\zeta \leq 1/4$ and rearranging gives

$$\mathbb{E}_{\pi_0}[D_j \mid \mathrm{past}] \leq 4\zeta\,\mathbb{E}_{\pi_0}[G_j \mid \mathrm{past}].$$

Let $D_{\mathrm{tot}} := \sum_s D_s$ and $G_{\mathrm{tot}} := \sum_s G_s$. Summing over all stages and using the tower property yields

$$\mathbb{E}_{\pi_0}[D_{\mathrm{tot}}] \leq 4\zeta\,\mathbb{E}_{\pi_0}[G_{\mathrm{tot}}] \leq 4\zeta W = O(\alpha)W. \tag{38}$$

Here $G_{\mathrm{tot}} \leq W$ pointwise because the algorithm starts from the empty matching and applies only positive-gain augmentations.

Number of sweeps and approximation guarantee. Since

$$\mathrm{gap}(M_{s+1}) = \mathrm{gap}(M_s) - G_s,$$

(37) implies

$$\mathrm{gap}(M_{s+1}) \le \left(1 - \frac{c}{k_0}\right)\mathrm{gap}(M_s) + \frac{C\alpha}{k_0}W + D_s.$$

Choose

$$L = \left\lceil C_L k_0 \log \frac{e}{\alpha} \right\rceil$$

sweeps, with C_L sufficiently large. Unrolling the recurrence, the initial term contributes $O(\alpha)W$, and the geometric sum of the deterministic additive terms contributes $O(\alpha)W$. The expected contribution of the D_s terms is at most $\mathbb{E}_{\pi_0}[D_{\mathrm{tot}}] = O(\alpha)W$ by (38). Thus

$$\mathbb{E}_{\pi_0}[\mathrm{gap}(M_L)] \le O(\alpha)W.$$

Since $k_0/(k_0+1) = 1 - O(\alpha)$, there is a universal constant $K \ge 1$ such that

$$\mathbb{E}_{\pi_0}[w(M_L)] \ge (1 - K\alpha)W.$$

The total number of sequential bucket stages is

$$S = LJ = O\left(\alpha^{-1}\log^2 \frac{e}{\alpha}\right). \tag{39}$$

Local implementation. For fixed k_0, the number of potential augmentations containing any given base-graph vertex is $\Delta^{O(k_0)}$. The same bound holds for the number of potential augmentations intersecting a given augmentation. Consequently every candidate conflict graph has maximum degree at most

$$\widehat{\Delta} = \Delta^{O(k_0)}.$$

Priority marks are keyed by the stage and a canonical encoding of the augmentation, and are owned by a canonical vertex of its support. The key universe can be fixed in advance using stage-tagged vertex sequences of length $O(k_0)$. Invalid encodings are never used as candidates. All priority marks are independent, and their collection is the shared randomness π_0.

Represent each intermediate matching by a local rule that, when queried at a vertex, returns its matched partner or declares it free. To answer an updated query at a root u, enumerate the potential augmentations containing u, determine which are valid candidates, and query their membership in I_ζ. Their validity and gains are determined by local edge data and queries to the preceding matching rule at their support vertices. Since selected augmentations are vertex-disjoint, these answers, together with the preceding matching at u, determine the updated partner of u.

A probe of the candidate conflict graph is simulated by enumerating intersecting potential augmentations and checking their validity. All these enumerations and local validations have multiplicity $\Delta^{O(k_0)}$. Within each base-root execution, cache simulated candidate probes and answers from the preceding matching rule, so that repeated requests do not repeat their underlying probes. The cache is private to that execution and does not change the MIS recursion or its call budget. In particular, a fixed base location can be inspected by the direct local validation of only $\Delta^{O(k_0)}$ potential candidates.

Query-profile accounting. Suppose the preceding matching rule has query profile (P^+, P^-). We apply Lemma F.1 to one bucket stage.

For the out-query direction, a base-root execution launches at most $\Delta^{O(k_0)}$ truncated-MIS root queries. By Theorem F.2(iii), each makes at most $O(\widehat{\Delta}^3/\zeta)$ conflict-graph probes. Each such probe requires at most $\Delta^{O(k_0)}$ queries to the preceding matching rule. Including the direct candidate validations gives

$$B^+ \le \frac{\Delta^{O(k_0)}}{\zeta}.$$

For the reverse direction, condition on all preceding-stage dictionaries. The current candidate graph and the lists of candidate roots queried from each base root are now fixed. Only the fresh priorities of the current stage remain random.

Each candidate appears as a truncated-MIS query root for at most $\Delta^{O(k_0)}$ base output roots. If every candidate root were queried once, Theorem F.2(iv) would bound the expected number of executions probing any fixed candidate by $O(\widehat{\Delta}^2)$. Accounting for the bounded repetition of candidate roots therefore gives at most

$$\Delta^{O(k_0)}\, O(\widehat{\Delta}^2)$$

expected probes of any fixed candidate.

Fix a root x of the preceding matching rule. A query at x can arise either from a direct candidate validation at a base output root, or while simulating a conflict-graph probe. The first source has multiplicity $\Delta^{O(k_0)}$. For the second source, only $\Delta^{O(k_0)}$ conflict candidates have local probe simulations that can reach x, and each such simulation issues at most $\Delta^{O(k_0)}$ queries to the preceding matching rule. Thus, over one execution from every base output root, the expected number of preceding-rule queries at x is

$$B^- \le \Delta^{O(k_0)}.$$

This bound holds conditional on every preceding-stage dictionary, as required by Lemma F.1.

Direct base-graph probes, including probes needed to read locally owned marks, satisfy the same bounds. By the additive clause of Lemma F.1, each stage therefore multiplies both profile coordinates, after adding one to each, by at most

$$B := \frac{\Delta^{O(k_0)}}{\zeta}.$$

Since $\Delta \ge 2$, $k_0 = \Theta(1/\alpha)$, and $\zeta^{-1} = O(1/\alpha)$, we can write $B = \Delta^{O(k_0)}$. Starting from the empty matching and composing the S stages gives

$$R^+, R^- \le \Delta^{O(k_0 S)} = \Delta^{O\left(\alpha^{-2}\log^2 \frac{e}{\alpha}\right)}.$$

The construction is natural. Candidate enumeration follows bounded paths from the query root, conflict-graph steps move between intersecting augmentations, and every preceding-rule query is rooted at a vertex exposed by this connected exploration. A priority mark is read only after its owner has been probed.

The final rule: internal perturbations and accuracy. Let $K \ge 1$ be the universal constant in the base approximation guarantee above. Fix

$$\alpha := \frac{\eta}{2(K+1)}, \qquad \varpi := \frac{\eta}{2},$$

and take $c_{\mathrm{lca}} \ge 2$. For every edge e, draw an independent mark

$$\rho_e \sim \mathrm{Uniform}[0, \varpi],$$

independently of the priority dictionary π_0. Each mark is owned by a fixed endpoint of its edge and is read only after that owner is probed. Define

$$\widetilde{w}_e := \frac{w_e(1+\rho_e)}{1+\varpi}, \qquad \widetilde{b} := \frac{\gamma}{1+\varpi}.$$

Then $\widetilde{w}_e \in [\widetilde{b}, 1]$, and

$$\varpi \le \frac{\eta}{2} < \frac{1}{4}, \qquad \widetilde{b} \ge \frac{4}{5}\gamma \ge \alpha.$$

Thus the base construction applies at working accuracy α and lower weight bound $\widetilde{b}$. With the full dictionary $\pi := (\pi_0, \rho)$, define the final rule by

$$M = \Psi(G, \pi; w) := \Psi^{\mathrm{base}}_{\alpha,\widetilde{b}}(G, \pi_0; \widetilde{w}).$$

The parameters $\alpha, \varpi, \widetilde{b}$ depend only on η, γ, and all marks have fixed local owners.

Conditional on ρ, the perturbed weights are fixed and π_0 retains its original distribution. The base query-profile bounds therefore hold uniformly in ρ. Reading the perturbation marks adds only bounded local probe overhead. Averaging over ρ preserves the expected in-query bound, while the

out-query bound holds for every full dictionary realization. Since α is a fixed constant multiple of η, a sufficiently large universal C_{lca} gives

$$R^+, R^- \le \Delta^{a_\eta}, \qquad a_\eta = C_{\mathrm{lca}} \eta^{-2} \log^2 \frac{e}{\eta}.$$

This also accounts for the perturbation marks within the local-access model of Definition A.1.

Conditional on ρ, the base approximation guarantee gives

$$\mathbb{E}_{\pi_0}[\widetilde{w}(M) \mid \rho] \ge (1 - K\alpha)\,\mathrm{OPT}_{\widetilde{w}}(G) \ge (1 - \eta/2)\,\mathrm{OPT}_{\widetilde{w}}(G).$$

Since $w_e \ge \widetilde{w}_e \ge w_e/(1+\varpi)$,

$$\begin{aligned}
\mathbb{E}_\pi[w(M)] &\ge \mathbb{E}_\rho[\mathbb{E}_{\pi_0}[\widetilde{w}(M) \mid \rho]] \\
&\ge (1 - \eta/2)\,\mathbb{E}_\rho[\mathrm{OPT}_{\widetilde{w}}(G)] \\
&\ge \frac{1 - \eta/2}{1 + \varpi}\,\mathrm{OPT}_w(G) \\
&\ge (1 - \eta)\,\mathrm{OPT}_w(G),
\end{aligned}$$

where the last inequality uses $\varpi \le \eta/2$. Together with the query-profile bounds, this proves part (i).

Continuity of the output marginals. Fix G, η, γ and a weight vector

$$w^0 \in [\gamma, 1]^{E(G)}.$$

In particular, the working accuracy, perturbation range, lower weight bound $\widetilde{b}$, number of stages, and bucket boundaries are fixed independently of the varying weight vector w.

There are finitely many possible intermediate matchings, bounded augmentations, and gain cutoffs. For any fixed matching N and nonempty augmentation P, its perturbed gain at w^0 is

$$\widetilde{g}_N(P; w^0, \rho) = \frac{1}{1+\varpi} \left(\sum_{e \in P \setminus N} w_e^0 (1 + \rho_e) - \sum_{e \in P \cap N} w_e^0 (1 + \rho_e) \right).$$

Because $w_e^0 > 0$, this is a nonconstant affine function of the independent continuous marks ρ_e. It therefore equals any fixed gain cutoff or bucket boundary with probability zero.

Take a finite union over all possible intermediate matchings, augmentations, and relevant boundaries. With probability one, none of these gain comparisons is tight at w^0. Priority ties also have probability zero. This argument considers all possible discrete histories at once and does not condition on the realized history.

Fix a full dictionary realization $\pi = (\pi_0, \rho)$ outside this null event. All relevant gain comparisons are strict at w^0 and depend continuously on w. Inducting over the finitely many stages, the candidate graphs, truncated-MIS decisions, and resulting matchings remain unchanged throughout some neighborhood of w^0. In particular, for every $e \in E(G)$,

$$\mathbf{1}\{e \in \Psi(G, \pi; w)\}$$

is locally constant at w^0 for almost every π.

The indicator is bounded by one. Bounded convergence therefore shows that

$$w \longmapsto \Pr_\pi[e \in \Psi(G, \pi; w)]$$

is continuous at w^0. Since w^0 was arbitrary, this proves part (ii). □

G Decorrelation under product arrivals

This appendix proves the transfer theorem used in the light completion. The only additional ingredient beyond the fixed-input correlation bound is the stability of the capped graph under resampling one arrival state.

State-resampling stability. For the bounds below, which average over arrivals and auxiliary marks, we separate these two sources of randomness. Conditional on the arrival states T, the capped graph is fixed and only the locally owned dictionary of the witness remains random. To control the remaining variation in T, we resample one arrival state at a time while keeping the entire witness dictionary fixed.

Corollary G.1 (State-resampling stability)**.** Let T' be obtained from T by replacing a single state T_i by an independent copy T_i', while keeping the witness dictionary fixed. Then

$$\left|E(\mathcal{G}_C^D(T))\triangle E(\mathcal{G}_C^D(T'))\right| \le 2dD.$$

Let $X_i(T,T')$ be the set of vertices whose capped neighborhood changes, as in (12), and put

$$\widehat{X}_i(T,T') := X_i(T,T')\cup\{i\}.$$

With the state-edge storage convention of Section B.2,

$$|\widehat{X}_i(T,T')| \le K := 1+2d(D+1), \tag{40}$$

all capped adjacencies and all state-dependent local data are identical outside $\widehat{X}_i(T,T')$, and $\widehat{X}_i(T,T')$ lies within distance two of i in the union of the two realized crucial graphs. Consequently, under the coupling that uses the same dictionary in the two environments, the answer of a natural query can change only if its execution reads a vertex of $\widehat{X}_i(T,T')$.

Proof. The edge bound and the neighborhood-change set are given by Lemma B.1(ii). The proof of that lemma bounds the containing set $\{i\}\cup X_i(T,T')$ by $1+2d(D+1)$, so adjoining i does not increase the displayed bound. State-dependent state-edge data are stored at their arrival endpoint, while all auxiliary marks are held fixed. Hence outside $\widehat{X}_i$ the two executions see identical adjacency, local data, and dictionary values. A first divergence can therefore occur only when a vertex of $\widehat{X}_i$ is read. □

Theorem D.1 (In–out decorrelation transfer)**.** *Let M_C be a matching of $\mathcal{G}_C^D(T)$ computed by a natural local rule whose free-status queries have query profile (R^+,R^-) for every arrival realization T. The input consists of the capped graph with each arrival vertex i labelled by T_i, together with deterministic instance data. The label T_i is revealed only when i is probed. The shared dictionary π is independent of T and of the menu randomness. Put*

$$F_v := \mathbf{1}\{v\notin V(M_C)\}, \qquad F_i := \mathbf{1}\{i\notin V(M_C)\}.$$

There is a universal constant $C_{\mathrm{dec}}\ge 1$ such that, with

$$\Gamma := C_{\mathrm{dec}}\big(1+d^2D(D+d+1)\big)R^+R^-,$$

the following bounds hold for all deterministic coefficient families $(h_v)_v$ and $(h_{it})_{i,t}$:

$$\mathsf{Var}\left(\sum_v h_vF_v\right) \le \Gamma\sum_v h_v^2, \qquad \mathsf{Var}\left(\sum_{i,t} h_{it}\mathbf{1}\{T_i=t\}F_i\right) \le \Gamma\sum_{i,t}p_i(t)h_{it}^2.$$

The first inequality remains valid after conditioning on any single state $T_i=t$ with $p_i(t)>0$. In addition, for every deterministic coefficient array satisfying $0\le h_{itv}\le h_{\max}$,

$$\sum_{i,t}p_i(t)\left|\sum_v h_{itv}\big(\mathsf{Pr}[F_v=1\mid T_i=t]-\mathsf{Pr}[F_v=1]\big)\right| \le C_{\mathrm{dec}}d^2DR^-h_{\max}x(C_\tau).$$

Proof. Let $C_{\mathrm{corr}}\ge 1$ be a universal constant for the variance bound in Lemma A.3. We will show that $C_{\mathrm{dec}} := C_{\mathrm{corr}}+14$ suffices. We separate the witness-dictionary randomness from the arrival-state randomness. Conditional on a realization of T, the capped graph and all state-dependent local data are fixed, and the witness is a private-tape local rule in the sense of Lemma A.3. Hence, for every deterministic coefficient vector (a_u),

$$\mathsf{Var}_\pi\left(\sum_u a_uF_u\,\middle|\,T\right) \le C_{\mathrm{corr}}R^+R^-\sum_u a_u^2. \tag{41}$$

For the transfer over T, fix a realization T and put

$$L^T_{u,w} := \Pr_\pi[w \in \mathrm{Read}_T(u)].$$

By the query-profile bounds,

$$\sum_w L^T_{u,w} \le R^+, \qquad \sum_u L^T_{u,w} \le R^-. \tag{42}$$

Indeed, the first inequality follows from the pointwise out-query bound, and the second is precisely the expected in-query bound on the fixed input T. Schur's test therefore gives, for every deterministic (a_u),

$$\sum_w \left(\sum_u |a_u| L^T_{u,w} \right)^2 \le R^+ R^- \sum_u a_u^2. \tag{43}$$

We next record the multiplicity with which one fixed input location can be affected by a single-coordinate resampling. Let $T^{(j)}$ be obtained by replacing T_j by an independent draw $T'_j \sim p_j$, and let

$$\widehat{X}_j := \widehat{X}_j(T, T^{(j)})$$

be the state-resampling set of Corollary G.1. Besides $|\widehat{X}_j| \le K := 1 + 2d(D+1)$, we have, for every fixed T and input vertex w,

$$\sum_j \Pr[\, w \in \widehat{X}_j \mid T \,] \le M := 1 + d(D + d + 1). \tag{44}$$

For an offline vertex v, a single-coordinate resampling cannot change its capped neighborhood when $N_v(T) \ge D + 2$. Otherwise there are at most $D + 1$ arrivals currently realizing a crucial edge at v, whereas

$$\sum_j \Pr[(j, T'_j, v) \in C_\tau] = \sum_{j,t} p_j(t) \mathbf{1}\{z_{jtv} \ge \tau\} \le \frac{1}{\tau} \sum_{j,t} p_j(t) z_{jtv} \mathbf{1}\{z_{jtv} \ge \tau\} = \frac{c_v}{\tau} \le d.$$

Thus the left side of (44) is at most $D + d + 1$ when $w = v$ is offline. Now let $w = r$ be an arrival. The resampling $j = r$ contributes at most one. For $j \ne r$, the capped neighborhood of r can change only through one of the at most d crucial offline neighbors of its fixed state T_r; applying the preceding offline bound to each such neighbor gives at most $1 + d(D + d + 1) = M$. This proves (44).

We first prove the offline variance bound. Put

$$S(T, \pi) := \sum_v h_v F_v, \qquad f(T) := \mathbb{E}_\pi[S(T, \pi)].$$

Couple T and $T^{(j)}$ using the same witness dictionary. By Corollary G.1, if the two answers at an offline root v differ then, before their first divergence, the execution on input T must read a vertex of $\widehat{X}_j$. Hence

$$\Pr_\pi[F_v(T) \ne F_v(T^{(j)})] \le \sum_{w \in \widehat{X}_j} L^T_{v,w},$$

and therefore

$$|f(T) - f(T^{(j)})| \le \sum_{w \in \widehat{X}_j} \sum_v |h_v| L^T_{v,w}. \tag{45}$$

For nonnegative numbers b_w and a set X of size at most K, $(\sum_{w \in X} b_w)^2 \le K \sum_{w \in X} b_w^2$. Squaring (45), summing over j, and taking the conditional expectation over the resampled states gives

$$\begin{aligned} &\sum_j \mathbb{E}[(f(T) - f(T^{(j)}))^2 \mid T] \\ &\quad \le K \sum_w \left(\sum_v |h_v| L^T_{v,w} \right)^2 \sum_j \Pr[w \in \widehat{X}_j \mid T] \\ &\quad \le K M R^+ R^- \sum_v h_v^2, \end{aligned}$$

where the last line uses (44) and (43). Efron–Stein now yields

$$\mathsf{Var}_T f(T) \le \frac{KM}{2} R^+ R^- \sum_v h_v^2.$$

Combining this with the conditional variance bound (41) through the law of total variance gives

$$\mathsf{Var}\left(\sum_v h_v F_v\right) \le (C_{\mathrm{corr}} + KM/2) R^+ R^- \sum_v h_v^2.$$

Since

$$KM \le 10 d^2 D(D+d+1),$$

our choice of C_{dec} proves the offline variance bound in Theorem D.1. If one state $T_i = t$ is fixed, the same argument applies to the remaining independent coordinates and simply omits i from Efron–Stein, so the same bound holds conditionally on $T_i = t$.

For the online form, define

$$S_{\mathrm{on}}(T,\pi) := \sum_i h_{i,T_i} F_i, \qquad f_{\mathrm{on}}(T) := \mathbb{E}_\pi[S_{\mathrm{on}}(T,\pi)].$$

Conditional on T, (41) with $a_i = h_{i,T_i}$ gives

$$\mathsf{Var}_\pi(S_{\mathrm{on}} \mid T) \le C_{\mathrm{corr}} R^+ R^- \sum_i h_{i,T_i}^2.$$

Under the coupling of T and $T^{(j)}$, decompose

$$\begin{aligned} |f_{\mathrm{on}}(T) - f_{\mathrm{on}}(T^{(j)})| \le \sum_i |h_{i,T_i}| \, \mathsf{Pr}_\pi[F_i(T) \ne F_i(T^{(j)})] \\ + |h_{j,T_j} - h_{j,T'_j}|. \end{aligned}$$

The square of the first term is handled exactly as in (45), now with coefficients $|h_{i,T_i}|$. The sum of its expected squares over j is at most

$$KM R^+ R^- \sum_{i,t} p_i(t) h_{it}^2.$$

For the direct coefficient change,

$$\sum_j \mathbb{E}[|h_{j,T_j} - h_{j,T'_j}|^2] \le 2 \sum_j \mathbb{E}[h_{j,T_j}^2 + h_{j,T'_j}^2] = 4 \sum_{j,t} p_j(t) h_{jt}^2.$$

Using $(a+b)^2 \le 2a^2 + 2b^2$, Efron–Stein, and the law of total variance gives

$$\mathsf{Var}(S_{\mathrm{on}}) \le (C_{\mathrm{corr}} + KM + 4) R^+ R^- \sum_{i,t} p_i(t) h_{it}^2,$$

where $R^+ R^- \ge 1$ absorbs the direct coefficient-change term. The bounds on KM and the choice of C_{dec} imply the claimed online-state variance bound with the same Γ.

It remains to prove the aggregate conditioning-bias bound. Fix (i,t) and couple an unconditional profile T to $T^{i \leftarrow t}$, obtained by replacing only T_i by t, using the same dictionary. Put

$$A_{it} := \{c_{i,T_i} > 0 \text{ or } c_{it} > 0\}.$$

Outside this event, i is isolated in both crucial graphs and its state change cannot affect any offline output of a natural rule. The first-divergence argument and the column bound in (42) therefore give

$$\begin{aligned} \left| \sum_v h_{itv} \big(\mathsf{Pr}[F_v = 1 \mid T_i = t] - \mathsf{Pr}[F_v = 1] \big) \right| \\ \le h_{\max} R^- \mathbb{E}\Big[|\widehat{X}_i(T, T^{i \leftarrow t})| \mathbf{1}_{A_{it}} \Big]. \end{aligned} \tag{46}$$

Since $c_{it} > 0$ implies $c_{it} \geq \tau$,

$$\sum_{i,t} p_i(t)\mathbf{1}\{c_{it} > 0\} \leq \frac{x(C_\tau)}{\tau} \leq d\, x(C_\tau).$$

Using $|\widehat{X}_i| \leq K$ and averaging over both the old and target states consequently gives

$$\sum_{i,t} p_i(t)\mathbb{E}\Big[|\widehat{X}_i(T, T^{i\leftarrow t})|\mathbf{1}_{A_{it}}\Big] \leq 2Kd\, x(C_\tau) \leq 10d^2 D\, x(C_\tau).$$

Summing (46) with weights $p_i(t)$ and using $C_{\mathrm{dec}} \geq 10$ proves the bias bound. □

H HUB INSTANCES: THE UNIVERSAL $\Omega(1/k)$ LOSS

For an instance $\mathcal{I}$, let $\mathcal{G}$ be its full realized graph and let H_Φ be the graph retained by a menu rule Φ. When $\mathbb{E}[\nu(\mathcal{G})] > 0$, define

$$\alpha_k(\mathcal{I}) := \sup_{\Phi:\,\mathrm{menu}(\Phi)\leq k} \frac{\mathbb{E}[\nu(H_\Phi)]}{\mathbb{E}[\nu(\mathcal{G})]}, \qquad \alpha^\star(k) := \inf_{\mathcal{I}} \alpha_k(\mathcal{I}),$$

where the infimum is over nontrivial product-arrival instances. Thus $\alpha^\star(k)$ is the best preservation guarantee obtainable uniformly over all such instances. By Theorem 3.1, $\alpha^\star(k) \to 1$; quantitatively, for all sufficiently large k,

$$1 - \alpha^\star(k) \leq O\left(\frac{\log\log\log k}{\log\log k}\right). \tag{47}$$

We show that the universal convergence cannot be faster than order $1/k$. This holds for *independent presence* and applies to all menu rules, not only benchmark-derived ones. The construction uses the following class of instances.

Hub instances. We call $\mathcal{I}$ a *hub instance* if its offline vertices $v_1, \ldots, v_m$ (the *columns*) each have a *dedicated arrival* u_j, present with probability $\sigma_j \in [0, 1]$ and adjacent only to v_j, and its remaining arrivals are $h \geq 1$ *hubs*, hub i being present with probability $\pi_i \in (0, 1]$ and adjacent to every column. All arrivals are independent. A hub adds value only when it finds a free column, one whose dedicated arrival is absent.

Theorem H.1 (Multi-hub lower bound)**.** *Let $k, h \geq 1$ be integers, and let $\mathcal{I}$ be the hub instance with $m = (k+1)h$ columns, $\sigma_j = k/(k+1)$ for all j, and h always-present hubs. Every menu rule of size at most k satisfies*

$$\mathbb{E}[\nu(H)] \leq \left(1 - \frac{(k/(k+1))^k - h^{-1/2}}{k+1}\right)\mathbb{E}[\nu(\mathcal{G})].$$

Consequently, letting $h \to \infty$,

$$1 - \alpha^\star(k) \geq \frac{k^k}{(k+1)^{k+1}} \geq \frac{1}{e(k+1)}, \qquad k \geq 1. \tag{48}$$

The first lower bound is asymptotic to $1/(ek)$.

Proof. We may assume that every dedicated edge is retained, since adding these edges respects the menu budget and can only increase $\nu(H)$. In either graph, there is a maximum matching that matches every present dedicated arrival: if such an arrival is unmatched, its column must be matched to a hub, and we can replace that edge by the dedicated edge.

Let N be the set of columns whose dedicated arrivals are absent. Then

$$|N| \sim \mathrm{Bin}\left(m, \frac{1}{k+1}\right), \qquad \mathbb{E}[|N|] = h, \qquad \mathsf{Var}(|N|) \leq h.$$

Matching the present dedicated arrivals first gives

$$\nu(\mathcal{G}) = m - |N| + \min\{h, |N|\}, \qquad \mathbb{E}[\nu(\mathcal{G})] \leq m = (k+1)h.$$

Let D_i be hub i's menu. In a maximum matching of H that matches all present dedicated arrivals, each matched hub uses a column of N. Consequently,

$$\nu(H) \le m - |N| + \sum_{i=1}^{h} \mathbf{1}\{D_i \cap N \neq \varnothing\}.$$

Each hub has a single state, so its menu is independent of N. Conditional on D_i, independence of the dedicated arrivals and $|D_i| \le k$ give

$$\mathsf{Pr}[D_i \cap N = \varnothing \mid D_i] = \left(\frac{k}{k+1}\right)^{|D_i|} \ge \left(\frac{k}{k+1}\right)^{k}.$$

It follows that

$$\begin{aligned}\mathbb{E}[\nu(\mathcal{G}) - \nu(H)] &\ge \mathbb{E}[\min\{h, |N|\}] - h\left(1 - \left(\frac{k}{k+1}\right)^{k}\right) \\ &= h\left(\frac{k}{k+1}\right)^{k} - \mathbb{E}[(h - |N|)_+] \ge h\left(\frac{k}{k+1}\right)^{k} - \sqrt{h}.\end{aligned}$$

The last inequality uses

$$\mathbb{E}[(h - |N|)_+] \le \mathbb{E}\big[\,\big||N| - h\big|\,\big] \le \sqrt{\mathsf{Var}(|N|)} \le \sqrt{h}.$$

If the resulting lower bound on the loss is nonpositive, the claimed inequality follows from $\nu(H) \le \nu(\mathcal{G})$. Otherwise, dividing by $\mathbb{E}[\nu(\mathcal{G})] \le (k+1)h$ gives

$$\frac{\mathbb{E}[\nu(\mathcal{G}) - \nu(H)]}{\mathbb{E}[\nu(\mathcal{G})]} \ge \frac{(k/(k+1))^k - h^{-1/2}}{k+1},$$

which proves the first claim.

Letting $h \to \infty$ and using the definition of $\alpha^\star(k)$ yields

$$1 - \alpha^\star(k) \ge \frac{k^k}{(k+1)^{k+1}} = \frac{1}{(k+1)(1+1/k)^k} \ge \frac{1}{e(k+1)}.$$

Here $(1+1/k)^k \le e$ for every $k \ge 1$, and its convergence to e shows that the first lower bound is asymptotic to $1/(ek)$. □

The bound stems from a coverage effect: Any menu of k columns is fixed before the arrivals are seen, while each column turns out to be free with probability only $1/k$; so a hub's menu contains no free column with probability $(1-1/k)^k \to 1/e$, however the menus are chosen. The full graph has no such restriction: its h hubs can use the roughly h free columns wherever they happen to fall.

There is also an upper bound that holds for every hub instance: independent uniform menus lose at most $1/(k+1)$. Therefore, hub instances can not yield tighter asymptotic lower bounds.

Proposition H.2 (Hub instances are $\Theta(1/k)$)**.** For every hub instance and every $k \ge 1$, the menu rule that retains every dedicated edge and gives each present hub an independent uniformly random set of k columns (all columns if $k \ge m$) satisfies

$$\mathbb{E}[\nu(H)] \;\ge\; \Big(1 - \frac{1}{k+1}\Big)\mathbb{E}[\nu(\mathcal{G})].$$

Proof. If $k \ge m$, every edge is retained, so assume $k < m$. Fix an arrival profile T. Let f be the number of columns whose dedicated arrivals are absent, and let h' be the number of present hubs. Put

$$s := \min\{h', f\}, \qquad r := m - f + s = \nu(\mathcal{G}).$$

First match every present dedicated arrival to its column. Choose any s present hubs in an order determined independently of their menus. Expose their menus in that order, matching each hub to an unused column in its menu whenever possible.

Before the $(j+1)$st hub is processed, at most $m-f+j$ columns are occupied. Conditional on T and all previously exposed menus, the next menu is an independent uniform k-subset of the columns. Thus the probability that this hub cannot be matched is at most

$$\frac{\binom{m-f+j}{k}}{\binom{m}{k}} \le \left(\frac{m-f+j}{m}\right)^k, \qquad 0 \le j < s,$$

where $\binom{a}{k} = 0$ when $a < k$.

Let L be the number of processed hubs left unmatched. The resulting matching has size $r - L$, and hence

$$\mathbb{E}[\nu(\mathcal{G}) - \nu(H) \mid T] \le \mathbb{E}[L \mid T] \le \sum_{j=0}^{s-1} \left(\frac{m-f+j}{m}\right)^k \le \int_{m-f}^{r} \left(\frac{x}{m}\right)^k dx$$
$$\le \frac{r^{k+1}}{(k+1)m^k} \le \frac{r}{k+1}.$$

The integral bound uses monotonicity of the integrand, and the last inequality uses $r \le m$. Since $r = \nu(\mathcal{G})$, averaging over T proves the claim. □

Combining (47) and (48), for all sufficiently large k,

$$\frac{k^k}{(k+1)^{k+1}} \le 1 - \alpha^\star(k) \le O\left(\frac{\log\log\log k}{\log\log k}\right).$$

On hub instances, Proposition H.2 bounds the worst-case loss between $k^k/(k+1)^{k+1}$ and $1/(k+1)$ for every $k \ge 1$. The ratio of these upper and lower bounds is $(1+1/k)^k < e$, approaching e as $k \to \infty$.

I Experimental details

Scope of the experiments. The hub families model competition for resources whose availability is uncertain. Each offline vertex has a dedicated arrival adjacent only to it, present with a specified probability. Hubs are adjacent to all these vertices and add value by using those whose dedicated arrivals are absent. NC is symmetric with one hub; NCH has heterogeneous presence probabilities; MH_4 has four hubs; and MH_k has k hubs. The private-resource variant TS_4 lets the same arrival have both a contested private resource and many shared alternatives.

NC, NCH, MH_4 and MH_k are hub instances, so Proposition H.2 applies at every scale. The construction of Theorem H.1 with $h = k$ corresponds to scale $c = 1 + 1/k$ in MH_k, which is not generally included in our tested grid. For TS_4, consider the rule that retains the private vertex c_i in the home state and samples $k-1$ uniform columns in either state. Match private vertices first. Their contribution is the same in the full and retained graphs, and the remaining semi-hubs form an independent random set of hubs with menus of size $k-1$. Proposition H.2 therefore bounds this rule's loss by $1/k$. The nested-neighborhood, random-presence, and random-heterogeneous families are outside the scope of that proposition. These analytical guarantees concern specified rules; the main plot evaluates VarOpt with estimated marginals. Its maxima over a finite scale grid are not estimates of $\alpha^\star(k)$.

Families. All hub families have always-present hubs. $\mathrm{NC}(m)$: one hub, $\sigma_j = 1 - 1/m$. $\mathrm{NCH}(m)$: one hub, $1 - \sigma_j \propto$ log-uniform on $[1/4, 4]$ with mean $1/m$. $\mathrm{MH}_h(m)$: h hubs, $1 - \sigma_j = h/m$. $\mathrm{TS}_h(m)$: the light layer of $\mathrm{MH}_h(m)$, whose hubs are replaced by h semi-hubs; semi-hub i has a private offline vertex c_i, contested by an intruder present with probability $1/2$ and adjacent to c_i alone, and two equiprobable states, "home" with neighborhood $\{c_i\} \cup \{v_1, \dots, v_m\}$ and "away" with neighborhood $\{v_1, \dots, v_m\}$. Scaled runs use $m = ck$ for NC and NCH, $m = 4ck$ for MH_4 and TS_4, with $c \in \{1.5, 2, 2.3, 3, 4\}$ and $k \in \{2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 64\}$; MH_k uses $h = k$ and $m = ck^2$ with $c \in \{0.75, 1, 1.5, 2.3\}$ for $k \le 32$ and $c \in \{0.75, 1\}$ for $k \in \{48, 64\}$. In the main figure NC appears only through its exact loss. Growing families outside the hub class (Figure 2): nested neighborhoods (arrival j present with probability $1 - 3/m$ and adjacent to $v_0, \dots, v_j$; $m = ck$, $c \in \{1.5, 2.3, 4\}$); random presence ($30k$ arrivals present with probability 0.8, each with a fixed

uniformly random neighborhood of size $4k$ among $18k$ offline vertices); random heterogeneous ($30k$ arrivals with three states of Dirichlet$(1,1,1)$ probabilities floored at 0.05, neighborhood sizes in $\{1,2,4\} \cup \{4k\,2^{-i} : 0 \le i \le 5\}$ weighted toward the largest, offline vertices drawn with Zipf(1) popularity among $21k$). The sensitivity check and the selector comparison also use three fixed instances: RH(300), a random heterogeneous instance with 300 arrivals and 210 offline vertices, three states per arrival with Dirichlet$(1,1,1)$ probabilities floored at 0.05, neighborhood sizes in $\{1,2,4,8,16,32\}$ with geometric weights favoring the smaller sizes, and Zipf(1) offline popularity; and SCAT$(300,d)$ for $d \in \{8,32\}$, 300 arrivals present with probability 0.8, each with a fixed uniformly random neighborhood of size d among 180 offline vertices.

Benchmark and marginals. The benchmark is Hopcroft–Karp on the realized graph with the arrivals in a uniformly random order and the offline vertices relabeled by a uniformly random permutation; this is a random maximum matching but not a uniform one. Conditional marginals z_{itv} are empirical frequencies over N_z unconditional arrival profiles: for each state (i,t) we divide the number of profiles in which (i,v) is matched by the number in which $T_i = t$. We use $N_z = 3000$, reduced to 2000 for $k \ge 48$ and for MH$_k$ with $k \le 32$, and to 1200 for MH$_k$ with $k \ge 48$. Every state with a nonempty neighborhood has probability at least 0.05 in all families, so each receives at least 60 profiles in expectation. The estimates can assign zero weight to an edge with positive true marginal, so the implementation need not satisfy (D1) relative to the true benchmark marginals. The finite menu banks described below introduce an additional approximation.

Menus. VarOpt$_k$ inclusion probabilities are $\min\{1, z_{itv}/\tau\}$ with $\sum_v \min\{1, z_{itv}/\tau\} = k$, after adding 10^{-12} to every estimate so that edges of zero estimated weight share the remaining budget uniformly; a neighborhood with at most k edges is retained entirely. The probabilities are rounded by the pivotal method with an independent uniformly random processing order for each sample, which realizes them exactly, has pairwise nonpositive covariances, and, when the probabilities are equal, yields a uniformly random k-subset by exchangeability. For each state we draw 64 menus once and, at evaluation, select one uniformly at random at every occurrence of the state. The exact inclusion probabilities and covariance guarantees above describe the underlying sampler; they need not hold exactly after conditioning on this finite menu bank. The comparison rules are *sim*, which draws k partners with replacement from the empirical partner pool of the state and retains the distinct ones (possibly fewer than k); *top-k*, the k largest estimated marginals with random tie-breaking; and *uniform*, k uniformly random neighbors.

Estimation of the loss. Marginal estimation and evaluation use independent profiles. For each of N_{ev} evaluation profiles (1500; 1200 for $k \ge 48$ and for MH$_k$ with $k \le 32$; 800 for MH$_k$ with $k \ge 48$) we compute $\nu(\mathcal{G})$ and $\nu(H)$ for every rule on the same profile; the loss is the mean difference divided by the mean of $\nu(\mathcal{G})$, and the reported standard error is that of the mean difference, divided by the mean of $\nu(\mathcal{G})$. This is an approximate evaluation standard error that treats the denominator as fixed and conditions on the estimated marginals and menu banks. It omits their estimation uncertainty and is not adjusted for selecting the largest estimated loss over the scale grid. Thus the plotted bars are not simultaneous confidence bands for the maxima. On NC, where every menu with k entries has loss $(a^k - a^m)/(m(1-\rho) + 1 - a^m)$ with $a = 1-\rho$, $\rho = 1/m$, the estimator reproduces the exact value within two standard errors at all 49 tested (k,c) pairs. Re-estimating the marginals with three independent seeds and with $N_z = 12000$ on RH(300), TS$_4(64)$, MH$_4(48)$ and NCH(18) changes the VarOpt loss by at most 6% (relative), within about two evaluation standard errors; the deterministic top-k rule on NCH varies by up to 25% across seeds, since its menus change discontinuously with the estimates. Losses below two standard errors are omitted from the log-scale plot.

Number of hubs. At $k = 8$, k times the loss of VarOpt menus on MH$_h$ at the best tested scale ($c \in \{0.5, 0.75, 1, 1.25, 1.5, 2, 2.3, 3\}$ for $h \le 32$, $\{0.75, 1, 1.25\}$ for $h \in \{64, 128\}$) is 0.12, 0.20, 0.27, 0.32, 0.36, 0.39, 0.41, 0.42 for $h = 1, 2, 4, \ldots, 128$. The tested procedure samples each hub's menu independently and does not preassign disjoint blocks of columns, which the model would allow.

Rules and benchmark selector. Table 1 compares the four rules at $k \in \{4, 8\}$ (maximum over tested scales for the scaled and growing families). Uniform menus have substantially larger measured loss on the heterogeneous families and can omit the private resource in TS$_4$. On symmetric NC, every menu with k entries has the same exact loss; the differences between full-menu rules in the table are

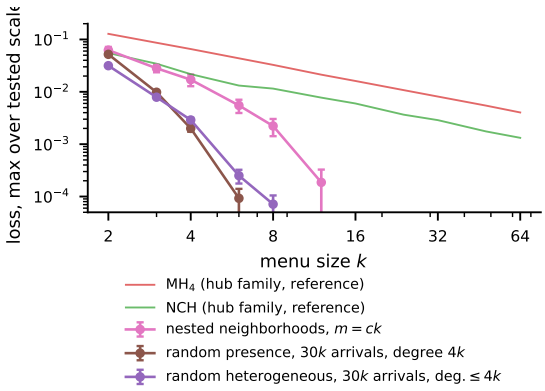

Figure 2: Loss of VarOpt$_k$ menus on growing families outside the hub class, maximum over the tested scales, with two hub families of Figure 1 for reference; error bars show twice the estimated evaluation standard error, and points below that level are omitted.

Table 1: Loss $\times 10^3$ (estimated evaluation standard error) by rule

family	k	VarOpt	sim	top-k	uniform
NC	4	33.6 (1.4)	46.5 (2.0)	32.5 (1.5)	32.2 (1.3)
NC	8	15.0 (0.7)	24.4 (1.0)	15.2 (0.7)	15.8 (0.7)
NCH	4	21.7 (1.0)	45.1 (2.0)	18.0 (0.7)	35.5 (1.4)
NCH	8	11.5 (0.4)	20.4 (0.9)	10.3 (0.4)	16.4 (0.7)
MH_4	4	65.6 (1.1)	68.7 (1.1)	114.3 (1.1)	63.4 (1.0)
MH_4	8	32.7 (0.5)	33.9 (0.5)	51.3 (0.6)	30.6 (0.5)
TS_4	4	58.6 (0.9)	64.7 (1.0)	65.8 (1.0)	67.7 (1.0)
TS_4	8	25.9 (0.5)	29.6 (0.5)	33.4 (0.5)	35.1 (0.5)
nested	4	17.1 (2.1)	24.4 (1.8)	43.9 (3.3)	92.9 (2.2)
nested	8	2.2 (0.4)	7.7 (0.8)	3.3 (0.5)	65.6 (1.2)
random presence	4	2.0 (0.1)	19.3 (0.4)	0.9 (0.1)	3.8 (0.2)
random presence	8	0.0 (0.0)	0.5 (0.1)	0.3 (0.0)	0.0 (0.0)
random heterogeneous	4	2.9 (0.2)	24.1 (0.4)	1.1 (0.1)	80.9 (0.7)
random heterogeneous	8	0.1 (0.0)	1.2 (0.1)	0.0 (0.0)	12.2 (0.2)

sampling noise. Top-k has the smallest measured loss on NCH and the random heterogeneous family, but the largest on MH_4, consistent with collisions between hubs' preferred choices. The simulation menu can retain fewer than k distinct edges and has larger measured loss than VarOpt in these comparisons. Replacing the Hopcroft–Karp selector by a maximum matching that, among maximum matchings, prefers low-degree arrivals (dense assignment with edge weights $1+1/(2n \deg i)$) roughly halves the VarOpt loss on the fixed random heterogeneous instance RH(300) at every k at which it is measurable (31.5 to 19.9, 3.6 to 2.0 and 0.69 to 0.30, all $\times 10^{-3}$, at $k = 2, 4, 6$), increases it by 10–20% on the random presence graphs, and changes it by at most 7% on the hub instances NC(64) and $TS_4(64)$. These comparisons show that the maximum-matching selector can materially affect the measured loss; they do not determine an asymptotic decay rate.