

Welfare-Optimal Matching in Hiring Markets via Min-Cost Flow with Ensemble Slot Allocation

Jeremy McEntire

TalentSync

USA

jandrewmcentire@gmail.com

Abstract

We formulate the talent-to-job matching problem as a min-cost max-flow optimization with tiered job nodes and ensemble slot allocation, departing from the Gale–Shapley stable matching paradigm that dominates the matching market literature. Our formulation exploits a structural feature of brokered hiring platforms: because the platform computes compatibility scores centrally rather than eliciting preference rankings from participants, the strategy-proofness constraint that drives much of mechanism design becomes irrelevant. This frees the optimizer to pursue welfare-weighted global optima subject to fairness constraints encoded directly in the flow network.

We introduce a hiring-science-informed scoring function with component weights derived from Sackett et al.’s (2022) meta-analytic validity coefficients, and an ensemble slot allocation strategy that partitions each job’s candidate capacity across welfare-maximizing, equity-distributing, and exploratory passes. We validate across three independent datasets with different labeling methodologies. In simulation (1,000 candidates \times 200 jobs, 50 random seeds), the MCF-Tiered engine achieves 95.3% company satisfaction versus 24.5% for the greedy per-job strategy that approximates current industry practice, with all advantages significant at $p < 10^{-20}$. On 8,000 human-labeled resume–job description pairs, MCF achieves a 90.8% match quality rate versus 74.9% for greedy ($p = 0.039$). On 851 resume–JD pairs with GPT-4o-generated fit scores, MCF-Tiered achieves 87.3% job coverage versus 40.0% for greedy and 68.5% for plain welfare-max at a 2:1 ratio ($p < 0.001$), validating that tiered nodes contribute independently of the welfare optimization itself. The ensemble variant fills 94.2% of jobs over a 15-week lifecycle versus 90.9% for pure optimization strategies ($p < 0.001$). The algorithm’s advantage is largest precisely when it matters most: at 1:1 ratio on real data, MCF-Tiered achieves 62.2% job coverage versus greedy’s 20.0%—a threefold advantage at the cold-start condition every new platform faces.

CCS Concepts

• **Theory of computation** \rightarrow **Algorithmic mechanism design; Network flows.**

Keywords

matching markets, min-cost flow, hiring, welfare optimization, ensemble matching, scoring validity

1 Introduction

The labor market matching problem has attracted sustained attention from economists since Gale and Shapley’s [8] foundational work on stable matchings. The resulting theoretical edifice—deferred

acceptance, the rural hospitals theorem, the impossibility results of Roth [13] and Bogomolnaia and Moulin [4]—provides elegant characterizations of what is achievable when self-interested agents report preferences over potential partners. The National Resident Matching Program, school choice systems in New York and Boston, and kidney exchange networks represent celebrated practical applications of this theory.

Hiring markets, however, remain conspicuously under-served by this tradition. The dominant paradigm in commercial talent platforms is what we call *greedy per-job ranking*: each employer independently reviews candidates ranked by some similarity score, makes overtures to the top of its list, and hopes for the best. This approach is not merely theoretically suboptimal; it is catastrophically wasteful. When multiple employers target the same popular candidates, the “losing” employers receive no introductions of value, and the platform generates a pattern of concentrated attention on a few candidates and neglect of the rest—precisely the opposite of what a healthy market requires.

The root cause is a coordination failure. Greedy ranking treats each job as an independent optimization, ignoring the constraint that binds the entire market together: each candidate can accept at most one position. The resulting allocation is not welfare-maximizing, not fair, and not stable.

This paper presents an alternative. We formulate the hiring platform matching problem as a min-cost max-flow optimization on a bipartite graph with tiered job nodes, and solve it using the well-understood machinery of network flow algorithms [2]. We extend this core with an ensemble slot allocation strategy that partitions each job’s candidate capacity across multiple algorithmic passes, achieving both high single-cycle quality and robust multi-week coverage.

Contributions.

- (1) A **network flow formulation** of the many-to-one hiring match problem with native capacity constraints, quality floors, and equitable distribution—achieved through flow network structure rather than ad hoc post-processing.
- (2) A **hiring-science-informed scoring function** with component weights derived from meta-analytic validity coefficients [15], replacing the cosine-similarity approach common in industry.
- (3) An **ensemble slot allocation strategy** that partitions job capacity across welfare-maximizing, equity-distributing, and exploratory passes, with built-in A/B testing at the slot level.
- (4) **Comprehensive empirical validation** across three independent datasets—synthetic simulation (50 seeds), 8,000 human-labeled resume–JD pairs, and 851 GPT-4o-scored

real resume–JD pairs—with consistent directional findings under different labeling methodologies and market ratios.

2 Related Work

2.1 Stable Matching

The Gale–Shapley deferred acceptance algorithm [8] produces a stable matching—one in which no unmatched pair mutually prefers each other to their current partners. Roth [14] showed that the NRMP mechanism is equivalent to the hospital-proposing variant of deferred acceptance. Extensions to many-to-one matching [10], matching with contracts [9], and dynamic matching [3] have substantially generalized the framework.

However, the stable matching literature is fundamentally concerned with *elicited preferences* from strategic agents. When a platform computes scores centrally—as our system does—the strategy-proofness requirement that motivates much of this work becomes inapplicable. Candidates do not submit rankings; the platform observes their profiles and computes compatibility. This observation, discussed further in Section 8, is what permits our departure from the deferred acceptance paradigm.

2.2 Welfare-Optimal Assignment

The assignment problem—maximizing total weight in a bipartite matching—has been studied since Kuhn’s [11] formulation of the Hungarian algorithm. The min-cost max-flow generalization [2, 7] extends this to many-to-one matching with capacity constraints, minimum flow requirements, and cost structures that encode fairness objectives directly in the network topology.

The IJCAI 2017 work on diverse weighted bipartite b -matching is the closest prior art to our approach, though it focuses on diversity constraints rather than the tiered-node structure we propose for equitable quality distribution.

An alternative allocation framework is Optimal Transport (OT) with fairness constraints [23]. OT minimizes the total cost of mapping a candidate distribution to a job distribution, and *FairSinkhorn* introduces group fairness targets into the entropic regularized OT problem via block-scaling matrices. This approach is natural when demographic fairness constraints are the primary concern, as the group targets are specified explicitly. Our tiered node structure achieves equitable distribution through network topology without requiring explicit demographic constraints—fairness emerges from the structure rather than being imposed as an external target. The approaches are complementary: *FairSinkhorn* could be layered onto our flow network when demographic parity is a regulatory requirement.

2.3 Impossibility Results

Three impossibility results frame the design space:

- **Roth [13]:** No mechanism is simultaneously stable and strategy-proof for both sides.
- **Bogomolnaia and Moulin [4]:** No mechanism is simultaneously efficient, envy-free, and strategy-proof.
- **The welfare–stability tradeoff:** Welfare-maximizing assignments are generally not stable, and stable assignments are generally not welfare-maximizing.

These results constrain mechanism designers who must accommodate strategic agents. For a platform that computes scores centrally, the binding constraint is the welfare–fairness tradeoff, which we address through the tiered node structure (Section 5.2).

2.4 Neural Resume–JD Matching

A substantial body of work treats resume–JD matching as a per-pair scoring problem. Li et al. [17] used context-aware transformer models with chunk segmenting and section encoding to achieve 73.3% accuracy on competence-level prediction and 79.2% on resume–JD matching, validated with triple annotation ($\kappa = 0.61$). The dominant production architecture by 2025 is the two-stage dense retrieval pipeline: a bi-encoder (BERT, RoBERTa) maps resumes and JDs independently into a shared embedding space for fast retrieval via approximate nearest neighbors, followed by a cross-encoder that concatenates the pair for precision re-ranking [19, 20]. Resume2Vec [19] reports +15.9% nDCG over conventional ATS platforms; ConFit v2 [20] introduces hypothetical resume embeddings and runner-up hard-negative mining for +17.5% nDCG gains.

More recently, graph-based approaches using knowledge graphs and Graph Neural Networks represent skills, roles, and competencies as interconnected nodes, enabling explicit relational reasoning and auditability. MPNET-RGCN architectures operating over the ESCO skill ontology report 25% RMSE reduction over transformer baselines in high-relatedness regions [21]. LLM-driven active listwise tournaments with Plackett–Luce aggregation address the ranking stage, treating candidate ordering as a structured prediction problem rather than independent pointwise scoring [22].

We treat compatibility scoring as a black box and focus on the allocation problem that scoring-only approaches leave unsolved. A better scoring function deployed in a greedy per-job architecture still fails to coordinate across the market—the contribution of the present work is orthogonal to, and downstream of, the scoring literature.

2.5 Hiring Science

Schmidt and Hunter [16] established that structured interviews and cognitive ability tests are the strongest predictors of job performance. Sackett et al. [15] substantially revised these findings, demonstrating that structured interviews ($r = .58$) now surpass cognitive ability tests (corrected from $r = .65$ to $r = .31$) as the single most predictive assessment method. McKinsey’s 2024 workforce report [12] found that skills-based hiring is five times more predictive of performance than credential-based hiring. Our scoring function (Section 4) operationalizes these findings.

3 Model

3.1 Problem Definition

Let $C = \{c_1, \dots, c_N\}$ be a set of candidates and $J = \{j_1, \dots, j_M\}$ be a set of jobs. Each candidate c_i has capacity 1 (can accept at most one job per matching cycle). Each job j_k has capacity K_k (can receive up to K_k candidate introductions per cycle; $K_k = 5$ in our standard configuration).

A *matching* μ is an assignment of candidates to jobs such that:

$$|\mu(c_i)| \leq 1 \quad \forall i \quad (1)$$

$$|\mu^{-1}(j_k)| \leq K_k \quad \forall k \quad (2)$$

The platform computes a compatibility score $s(c_i, j_k) \in [0, 1]$ for each candidate–job pair (Section 4). The objective is to find a matching that maximizes total welfare:

$$W(\mu) = \sum_{(c_i, j_k) \in \mu} s(c_i, j_k) \quad (3)$$

subject to capacity constraints and, optionally, fairness constraints (quality floor, equitable distribution, minimum fill).

3.2 Weekly Cycle Model

The platform operates on a weekly matching cycle. Each week:

- (1) The score matrix S is computed for all active candidate–job pairs.
- (2) A sigmoid decay function $d(t) = (1 + e^{0.5(t-10)})^{-1}$ is applied based on listing and resume age.
- (3) The matching algorithm produces introductions.
- (4) Some introductions result in hires (candidate and job exit the market).
- (5) Previously introduced pairs are blocked from re-introduction.

3.3 Evaluation Metrics

We evaluate algorithms on metrics that reflect market health:

- **Company satisfaction:** Fraction of active jobs that receive at least one candidate above the hireability threshold (score ≥ 0.70).
- **Candidate satisfaction:** Fraction of matched candidates whose best match exceeds the acceptability threshold (score ≥ 0.65).
- **Market health:** Harmonic mean of company and candidate satisfaction.
- **Quality floor:** The worst job’s best candidate score—a measure of equitable distribution.
- **Job fill rate** (multi-week): Fraction of jobs resulting in a hire over the simulation period.

4 Scoring Function

4.1 Design Principles

Two design principles govern the scoring function’s construction.

First, *validity-weighted components*: each scoring dimension receives a weight proportional to its meta-analytic predictive validity for job performance, as reported by Sackett et al. [15]. This grounds the weighting in empirical evidence rather than intuition.

Second, *culture add over culture fit*: we score for the diversity of perspective a candidate brings to an existing team—what unique contribution they would make—rather than their similarity to the current team composition. Culture-add teams produce 35% more innovative output and are 70% more likely to capture new markets.

4.2 Component Weights

Table 1 lists the component weights. These weights produce 72% top-1 disagreement with a standard cosine-similarity scoring function

Table 1: Scoring function component weights, derived from meta-analytic validity coefficients.

Component	Weight	Validity Source
Structured assessment	0.28	$r = .58$ (Sackett et al., 2022)
Skills overlap	0.25	$r = .53/.48$ combined
Culture add	0.17	35% innovation boost
Psychological match	0.15	7-axis trait alignment
Seniority alignment	0.10	Experience level match
Role alignment	0.05	Archetype match (weakest)

on our test data: the revised scoring function picks a fundamentally different best candidate for nearly three-quarters of all jobs.

4.3 Component Definitions

Skills overlap. is not cosine similarity of skill vectors. We decompose it into three sub-dimensions: *coverage* (fraction of required skills the candidate possesses), *depth* (average proficiency in required skills weighted by job importance), and *surplus* (additional skills beyond requirements—the “T-shaped engineer” bonus). The composite is $0.50 \cdot \text{coverage} + 0.35 \cdot \text{depth} + 0.15 \cdot \text{surplus}$.

Culture add. measures alignment on core values (conscientiousness, emotional stability) combined with diversity on approach dimensions (adaptability, communication, proactive initiative). The score peaks when a candidate shares the team’s fundamental reliability but brings a distinct perspective on execution. A candidate who diverges on core values receives a low score regardless of perspective diversity.

Structured assessment. simulates the signal that a well-designed AI-mediated structured interview would extract. In production, this comes from the actual AI conversation; in simulation, it is approximated from psychological profile components with noise. The weight of 0.28 reflects Sackett et al.’s finding that structured interviews are the single most predictive assessment method.

5 Algorithm

5.1 Min-Cost Max-Flow Formulation

We construct a flow network $G = (V, E)$ as follows.

Nodes. A source node s ; one node per active candidate $\{v_{c_i}\}$; one node per active job $\{v_{j_k}\}$ (or tiered variants; see Section 5.2); and a sink node t .

Arcs.

- $s \rightarrow v_{c_i}$: capacity 1, cost 0.
- $v_{c_i} \rightarrow v_{j_k}$: capacity 1, cost = $-[s(c_i, j_k) \cdot \text{SCALE}]$ where $\text{SCALE} = 10,000$.
- $v_{j_k} \rightarrow t$: capacity K_k , cost 0.

Arcs with $s(c_i, j_k)$ below the quality floor are omitted, encoding the minimum quality constraint in the network structure. The supply at s equals $\min(N, \sum_k K_k)$; the demand at t matches.

Solving for minimum-cost maximum flow yields the welfare-maximizing matching subject to capacity constraints and quality floors. We use Google OR-Tools’ SimpleMinCostFlow solver, which

implements a cost-scaling push-relabel algorithm with $O(|V| \cdot |E| \cdot \log(|V| \cdot C_{\max}))$ complexity.

Scalability. At 10,000 candidates and 1,000 jobs, the network contains approximately 11,000 nodes and up to 10 million edges. The solver completes in seconds. At our standard configuration (1,000 \times 200), solve time is under 100 ms.

5.2 Tiered Job Nodes

Plain min-cost max-flow maximizes total welfare but does not guarantee equitable distribution. To address this, we introduce *tiered job nodes*.

Each job j_k is split into two nodes: a *premium* node $v_{j_k}^P$ with capacity P (default $P = 1$) and a *standard* node $v_{j_k}^S$ with capacity $K_k - P$. Arcs from candidates to the premium node carry a bonus:

$$\text{cost}_P = -[(s(c_i, j_k) + \beta) \cdot \text{SCALE}] \quad (4)$$

where β is an adaptive premium bonus set to $\beta = \alpha \cdot \sigma(S)$, with $\alpha = 0.7$ and $\sigma(S)$ the standard deviation of active scores. This makes the bonus scale-invariant: on any scoring function, β is a fixed fraction of the score spread, preventing the tiered structure from dominating routing when score differences are small. Arcs to the standard node carry the unmodified cost.

The effect is that the optimizer preferentially fills premium slots across all jobs before filling standard slots, because the premium bonus makes it “cheaper” (in the minimization sense) to fill a premium slot at an unfilled job than a standard slot at a popular job. This forces top candidates to be distributed more equitably across the market without requiring an explicit fairness constraint—the fairness emerges from the network topology.

At 1.5:1 candidate-to-job ratio, tiered nodes increase company satisfaction from 59.8% (plain MCF) to 95.3%, a dramatic improvement achieved solely through network structure.

5.3 Ensemble Slot Allocation

The core MCF-Tiered engine optimizes for a single objective. Over a multi-week lifecycle, however, we observe that pure optimization strategies fill approximately 91% of jobs, while an *ensemble* strategy fills 94.2%—a statistically significant improvement ($p < 0.001$).

The ensemble partitions each job’s 5 candidate slots across three algorithmic passes:

- (1) **Pass 1—Welfare** (2 slots): Plain MCF maximizing total welfare.
- (2) **Pass 2—Equity** (2 slots): MCF-Tiered with premium bonus, distributing quality candidates across jobs.
- (3) **Pass 3—Exploration** (1 slot): Stochastic selection from the top-30% viable candidate pool, weighted by $\sqrt{1/\text{rank}}$ rather than by score.

Each pass operates on the residual market after previous passes have claimed their candidates. The exploration slot’s $\sqrt{1/\text{rank}}$ weighting is a deterministic approximation of Upper Confidence Bound (UCB) exploration from the contextual bandit literature [24], which balances exploitation of known-good candidates with exploration of uncertain ones. Research on UCB-based hiring shows that exploration more than doubles interview rates for under-represented candidates without reducing long-term match quality [24]. The

exploration slot also provides a natural A/B testing framework—the strategy that produced each match is tracked, enabling empirical comparison over time.

6 Experimental Design

6.1 Synthetic Data

We generate candidates and jobs using an O*NET-inspired skill taxonomy comprising 79 dimensions across seven categories (languages, frontend, backend, data, infrastructure, ML, practices). Each candidate has a skill proficiency vector in $[0, 1]^{79}$, a 7-axis psychological profile, a role archetype (one of seven), and a seniority level (one of five). Jobs have corresponding requirement vectors, desired psychological profiles, and team need types.

Skill vectors are generated from role archetype distributions with seniority scaling and Gaussian noise. Job requirements are inflated by a factor drawn from Uniform(0.9, 1.2), modeling the well-documented tendency of employers to over-specify requirements.

6.2 Real Data

We validate on the `cnamuangtoun/resume-job-description-fit` dataset [6], hosted on Hugging Face: 8,000 resume–job description pairs constructed from 642 unique resumes and 280 unique job descriptions, with human-assigned labels (*No Fit*, *Potential Fit*, *Good Fit*). The dataset is split 78/22 into 6,240 training and 1,760 test pairs.

Provenance note. The dataset does not publish its labeling methodology or inter-annotator agreement metrics. Provenance limitations are discussed in Section 8.6.

TF-IDF features (5,000 dimensions, bigrams, `min_df = 3`) are used to compute cosine similarity, which serves as the base compatibility score. We construct synthetic markets of 200 candidates and 50 jobs, ensuring that known-labeled pairs are present, and measure whether each algorithm discovers the expert-labeled good matches.

6.3 Vanetik & Kogan Dataset

We additionally validate on the Vanetik and Kogan [18] vacancy-resume matching dataset: 65 anonymized CVs and 5 job vacancies from data.world, with ground-truth rankings produced by two independent annotators following instructions from a senior HR specialist. Rankings cover the first 30 CVs against all 5 vacancies. This dataset has documented annotation methodology and a published paper, addressing the provenance gap in the `cnamuangtoun` dataset. Its small size (30 \times 5) limits statistical power but provides confirmatory evidence under a different labeling regime.

6.4 Netsol Dataset

We validate on the `netsol/resume-score-details` dataset: 851 valid resume–JD pairs with GPT-4o-generated compatibility scores on a 0–10 scale across macro (leadership, technical expertise) and micro (skill-specific) dimensions, plus boolean requirement satisfaction. We construct a composite score normalized to $[0, 1]$ from macro, micro, and requirement components. Markets of varying sizes are assembled from unique JDs and resumes, with GPT-4o

Table 2: Single-cycle metrics at 1.5:1 ratio, averaged over 30 seeds.

Strategy	Company Sat.	Market Health	Floor
MCF-Tiered	95.3%	97.6%	0.600
Ensemble (MCF)	80.7%	87.4%	0.520
Gale-Shapley	60.1%	74.2%	0.325
Welfare-Max	59.8%	74.8%	0.423
Greedy	24.5%	37.6%	0.180

scores used for known pairs and TF-IDF cosine similarity for unknown pairs. This dataset provides a third labeling methodology (LLM-generated rather than human) and enables evaluation at realistic market ratios.

6.5 Experimental Conditions

- **Single-cycle benchmark:** 1,000 candidates \times 200 jobs (5:1 ratio), 30 random seeds.
- **Noise analysis:** 50 seeds \times 1,000 \times 200, with threshold sweep (0.40–0.95) and ratio sweep (1:1 through 10:1).
- **Multi-week lifecycle:** 15 weekly cycles, $P(\text{hire}) = 0.15$ score, 20 seeds per strategy.
- **Real data backtest (cnamuangtoun):** 30 trials \times 200 candidates \times 50 jobs, sampled from the 8K-pair dataset.
- **Real data backtest (Netsol):** 20 seeds at three market ratios (5:1, 2:1, 1:1) with GPT-4o composite scores.
- **Confirmatory (Vanetik):** 30 CVs \times 5 vacancies against dual-annotator rankings.

6.6 Baselines

- **Greedy per-job:** Each job independently selects its top- K candidates by score. Candidates claimed first-come. Approximates current industry practice.
- **Gale-Shapley** (candidate-proposing): Converges to a stable matching.
- **Welfare-Max** (Hungarian): Maximum-weight bipartite matching via the Hungarian algorithm with job node expansion for many-to-one capacity.

7 Results

7.1 Single-Cycle Performance

Table 2 reports single-cycle metrics. The tiered node structure is the decisive innovation: plain welfare-max and Gale-Shapley perform comparably, while greedy is catastrophically bad—nearly four out of five employers receive no hireable candidate.

7.2 Statistical Validation

Across 50 random seeds at 1,000 \times 200, all pairwise advantages of MCF-Tiered over greedy are significant at $p < 10^{-20}$ by paired t -test. The advantage is robust across the full threshold sweep (0.40 to 0.95) and grows monotonically as the market tightens:

The algorithm’s advantage is largest at cold start and in competitive markets—precisely the conditions under which a new platform must demonstrate value (Table 3).

Table 3: Market health by candidate-to-job ratio.

Ratio	MCF-Tiered	Greedy	Δ
1:1 (tight)	62%	27%	+35%
1.5:1	97.6%	37.6%	+60%
5:1 (standard)	99%+	91%	+8%
10:1 (surplus)	100%	100%	0%

Table 4: 15-week lifecycle at 1.5:1 ratio, averaged over 20 seeds.

Strategy	Total Hires	Jobs Filled
MCF (welfare)	103.8	90.7%
Gale-Shapley	103.4	90.9%
Ensemble (2W+2T+1E)	100.8	94.2%
Greedy	100.7	82.9%

Table 5: Backtest on 8,000 human-labeled resume-JD pairs.

Strategy	Success Rate	Good Fit	p vs. Greedy
MCF	90.8%	44.2%	—
Welfare-Max	90.8%	44.2%	—
Greedy	74.9%	40.1%	0.039
Gale-Shapley	72.6%	30.2%	0.020

7.3 Multi-Week Lifecycle

The ensemble fills 94.2% of jobs versus 90.9% for the best pure strategy ($p < 0.001$; Table 4). Total hires are modestly lower (100.8 vs 103.8), reflecting the exploration slot’s lower conversion rate. The tradeoff is favorable: 3.5 more jobs filled for 3 fewer total hires.

7.4 Real Data Validation

MCF matches agree with human expert labels 9 times out of 10 (Table 5). The good-fit rate (44.2% vs. 30.2% for Gale-Shapley) indicates that MCF not only avoids bad matches but actively discovers the best ones. The provenance limitations noted in Section 6.2 apply.

7.5 Netsol: Real Data at Multiple Ratios

Table 6 presents results on the Netsol dataset across three market ratios. The MCF-Tiered advantage is most striking at 2:1: 87.3% job coverage versus 40.0% for greedy and 68.5% for plain MCF—an 18.8 percentage-point advantage of tiered nodes over welfare maximization alone ($p < 0.001$). This validates the tiered node structure as an independent architectural contribution, not merely a side effect of global optimization.

At 1:1 ratio—the cold-start condition—MCF-Tiered achieves 62.2% job coverage versus greedy’s 20.0%, a threefold advantage on real data. Gale-Shapley’s stability constraint distributes matches somewhat more broadly than pure welfare maximization (71.7% vs. 68.5% at 2:1), but the tiered node structure accomplishes this more effectively by design rather than as a side effect of the stability property.

Table 6: Netsol dataset (851 GPT-4o-scored resume-JD pairs), 20 seeds. MCF-Tiered vs. baselines at three market ratios. $k = 5/\text{job}$.

Ratio	Strategy	Job Covg	Co. Sat	p vs. Greedy
5:1	MCF-Tiered	100.0%	81.5%	< 0.001
	Welfare-Max	100.0%	81.8%	< 0.001
	Gale-Shapley	100.0%	74.0%	< 0.001
	Greedy	100.0%	56.5%	—
2:1	MCF-Tiered	87.3%	58.8%	< 0.001
	Gale-Shapley	71.7%	55.0%	< 0.001
	MCF	68.5%	55.8%	< 0.001
	Greedy	40.0%	16.8%	—
1:1	MCF-Tiered	62.2%	43.5%	< 0.001
	GS / WM / MCF	~50%	~39%	< 0.001
	Greedy	20.0%	11.8%	—

Table 7: 15-week lifecycle on cnamuangtoun (30 seeds, TF-IDF scores).

Ratio	Strategy	Hires	Filled
1.5:1	MCF-Tiered	29.4	28.7
	Gale-Shapley	29.8	29.0
	Welfare-Max	29.6	28.7
	Greedy	20.5	19.7
1:1	MCF-Tiered	18.7	18.6
	Welfare-Max	20.4	20.1
	Gale-Shapley	19.9	19.6
	Greedy	12.2	11.8

At 5:1, all algorithms converge on job coverage, consistent with the synthetic ratio sweep—the algorithm matters less when candidates are abundant.

7.6 Vanetik: Confirmatory Null at Surplus Ratio

On the Vanetik and Kogan [18] dataset (30 CVs, 5 vacancies, $k = 6/\text{job}$), all algorithms converge to 100% job coverage and statistically indistinguishable performance. Inter-annotator agreement is low (Kendall’s $\tau = 0.129$, top-1 agreement 16.7%), limiting the achievable agreement ceiling. This result is consistent with the ratio sweep finding: at a 6:1 effective ratio, the market is not tight enough for the matching algorithm to differentiate. The confirmatory null strengthens the claim that the MCF-Tiered advantage is specifically a tight-market phenomenon.

7.7 Multi-Week Lifecycle on Real Data

Table 7 reports multi-week lifecycle results on real resume-JD data. MCF-Tiered retains a decisive advantage over greedy at both ratios (+8.9 and +6.8 hires, $p < 0.001$). However, MCF-Tiered is statistically tied with Gale-Shapley and Welfare-Max at 1.5:1, and *slightly worse* at 1:1 (−1.5 filled vs. Welfare-Max, $p = 0.002$).

Table 8: Large-scale synthetic (2,000c × 400j), 15-week lifecycle, 30 seeds. Jobs filled (%).

Ratio	Strategy	Fill Rate
1.5:1	MCF-Tiered	96.9%
	GS / WM	94.9%
	Greedy	88.5%
1:1	MCF-Tiered	74.1%
	GS / WM	~68.5%
	Greedy	58.2%

This divergence from the synthetic results—where MCF-Tiered dominates all alternatives—traces to a specific mechanism: the premium bonus β . On the hiring-science scorer (score range 0.30–0.84), $\beta = 0.10$ is approximately 18% of the score spread, producing a meaningful but non-dominating equity incentive. On TF-IDF cosine scores (range ~0.00–0.20 on real data), $\beta = 0.10$ exceeds most score *differences* between candidates, causing the tiered structure to dominate routing decisions—effectively randomizing which job receives the “premium” pick rather than equitably distributing genuinely strong candidates.

On large-scale synthetic data with the hiring-science scorer (Table 8), the MCF-Tiered advantage holds strongly: 96.9% fill rate versus 94.9% for GS/WM ($p < 0.001$) and 88.5% for greedy at 1.5:1. At 1:1: 74.1% versus 68.5% ($p < 0.001$). This confirms that tiered nodes contribute genuine value when the scoring function has sufficient discriminative power.

We address this by making β adaptive: $\beta = \alpha \cdot \sigma(S)$, where $\sigma(S)$ is the standard deviation of active scores. With $\alpha = 0.7$, the ratio $\beta/\sigma(S)$ is constant at 0.7 regardless of scorer. On TF-IDF ($\sigma = 0.043$), adaptive $\beta = 0.030$ versus fixed $\beta = 0.10$ —a threefold reduction that eliminates the regression (−0.2 filled vs. Welfare-Max, $p = 0.56$, versus −1.3 filled with fixed β , $p = 0.002$). On the hiring-science scorer ($\sigma = 0.138$), adaptive $\beta = 0.097$, preserving the full tiered advantage (+21.9 filled over Welfare-Max at 1:1, $p < 0.001$). Adaptive β thus strictly dominates fixed β across all tested conditions.

7.8 Scoring Function Impact

The hiring-science-informed scoring function disagrees with a standard cosine-similarity scorer on the top-1 candidate for 72% of jobs. The revised function has a more discriminating score range (0.30–0.84 vs. 0.26–0.94) and produces a culture-add component averaging 0.859, confirming that it correctly identifies candidates who bring diverse perspectives.

8 Discussion

8.1 Why Strategy-Proofness Does Not Apply

The most common objection from the mechanism design community will be: where is the strategy-proofness guarantee? The answer is that it is unnecessary in this context, and its absence is the key that unlocks welfare optimization.

In classical matching markets, agents submit preference rankings, and the mechanism must be robust to strategic misreporting.

In a brokered hiring platform, neither side submits preferences. The platform observes candidate profiles (resumes, assessment results, behavioral data) and employer requirements (job descriptions, team composition, hiring history), and computes compatibility scores from these observations. A candidate cannot “misreport” their compatibility with a job; they can only present their profile as favorably as possible, which is a different problem—profile optimization, not preference manipulation.

Roth’s [13] impossibility result binds only when agents have private preferences that the mechanism must elicit. When the platform possesses the relevant information directly, the impossibility result does not apply, and the designer is free to optimize for whatever objective the market requires.

8.2 The Fairness–Welfare Interface

Our tiered node structure resolves the fairness–welfare tension through network topology rather than constrained optimization. Fairness constraints expressed as linear programming constraints often lead to infeasible problems or corner solutions; the tiered approach guarantees feasibility by construction and produces smooth quality distributions.

The intuition: in a plain flow network, the optimizer may route all flow through a single high-capacity job node. The tiered structure creates a bottleneck at the premium tier—once one premium slot is filled, additional candidates must compete for standard slots, which have lower priority than premium slots at other jobs.

8.3 Exploration as Insurance

The exploration slot is not a concession to randomness; it is insurance against scoring function errors. Any scoring function encodes assumptions about what constitutes a good match. The exploration slot deliberately relaxes these assumptions for one out of five slots per job, drawing from the viable-but-non-obvious portion of the candidate pool. This serves three purposes: it surfaces candidates whose value the scoring function underestimates, it provides empirical data for scoring function recalibration, and it introduces productive surprise that static optimization cannot produce.

8.4 Scorer Independence of the Allocation Advantage

A natural concern is that the MCF-Tiered advantage may depend on properties of the scoring function rather than the allocation algorithm itself. To test this, we ran the full ratio sweep with three different scorers feeding the same MCF-Tiered engine: cosine similarity (the industry baseline), our hiring-science-informed scorer, and a graph-based scorer that incorporates skill prerequisite chains, category co-occurrence, and IDF-weighted specificity.

At 1:1 ratio, MCF-Tiered achieves 83.5%–91.8% coverage depending on scorer, while greedy achieves ~20% regardless of scorer. The MCF-Tiered advantage over greedy ranges from +63.5 to +71.8 percentage points across all three scorers (all $p < 0.001$). Greedy per-job allocation is catastrophically bad *independent of scoring quality*. This confirms that the allocation algorithm, not the scoring function, is the binding constraint in hiring market efficiency—and that our contribution is orthogonal to the scoring literature reviewed in Section 2.4.

8.5 Regulatory Auditability

The EU AI Act classifies AI recruitment systems as high-risk, requiring documented risk management, bias testing, and explainable decision logic. NYC Local Law 144 mandates annual independent bias audits for automated employment decision tools. Our approach is structurally compatible with these requirements: the scoring function has explicit, citable component weights derived from peer-reviewed meta-analyses, and every allocation decision traces to the flow network structure—which arcs carried flow, which tiered slots were filled, which exploration pass produced a given match. This is auditable in a way that opaque dense-vector embedding spaces are not.

8.6 Limitations

Simulation fidelity. Our synthetic data generator, while based on O*NET skill taxonomies, does not capture the full complexity of real labor markets. The structured assessment component (0.28 weight) is approximated in simulation from psychological profiles—in production, this signal comes from actual AI-mediated conversations with different distributional properties.

Hire probability model. We model hire probability as proportional to match score. Real hiring involves multiple stages with nonlinear attrition.

Dataset provenance. The cnamungtoun dataset lacks published labeling methodology. The Netsol dataset uses LLM-generated labels, which may embed systematic biases different from human annotation. The Vanetik dataset has documented methodology but is small. The consistency of directional findings across all three datasets, each with a different labeling regime, mitigates but does not eliminate concerns about any single dataset’s validity.

Scale. While the MCF solver handles $10K \times 1K$ markets in seconds, the scoring function itself is $O(NMD)$. Pre-clustering may be necessary at very large scale.

Single market. Our experiments model a single labor market. Behavior in highly segmented markets remains untested.

9 Conclusion

Hiring markets are currently served by tools that ignore the fundamental constraint of the problem: that candidates are a shared, rivalrous resource. The greedy per-job approach that dominates industry practice leaves 75% of employers without a hireable candidate at moderate market ratios.

We have shown that formulating the hiring match as a min-cost max-flow problem with tiered job nodes and ensemble slot allocation achieves 95.3% company satisfaction in simulation and 87.3% job coverage on real resume–JD pairs, validated across three independent datasets with $p < 10^{-20}$ on the synthetic benchmark and $p < 0.001$ on both real-data evaluations. The approach is grounded in hiring science [15] rather than ad hoc similarity metrics, and it provides a natural mechanism for balancing welfare, fairness, and exploration without requiring the strategy-proofness guarantees that constrain classical mechanism design—because the platform, not the participants, computes the scores.

The mathematics is seventy years old. The hiring science is three years old. The combination is new.

The algorithm’s advantage is most pronounced in tight markets at cold start—at 1:1 ratio on real data, MCF-Tiered achieves 62.2% job coverage versus greedy’s 20.0%. These are exactly the conditions a new platform faces. The ensemble slot allocation strategy achieves higher job coverage than any pure optimization strategy over multi-week lifecycles, while providing built-in A/B testing and scoring function error correction. It is, in effect, a mechanism that improves itself over time by design.

References

- [1] A. Abdulkadiroglu and T. Sonmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.
- [2] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.
- [3] M. Akbarpour, S. Li, and S. O. Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815, 2020.
- [4] A. Bogomolnaia and H. Moulin. A new solution to the random assignment problem. *Journal of Economic Theory*, 100(2):295–328, 2001.
- [5] E. Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.
- [6] C. Na Muangtoun. resume-job-description-fit dataset. Hugging Face Datasets, 2024. <https://huggingface.co/datasets/cnamuangtoun/resume-job-description-fit>.
- [7] L. R. Ford and D. R. Fulkerson. Maximal flow through a network. *Canadian Journal of Mathematics*, 8:399–404, 1956.
- [8] D. Gale and L. S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.
- [9] J. W. Hatfield and P. Milgrom. Matching with contracts. *American Economic Review*, 95(4):913–935, 2005.
- [10] A. S. Kelso and V. P. Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica*, 50(6):1483–1504, 1982.
- [11] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955.
- [12] McKinsey & Company. Taking a skills-based approach to building the future workforce. Technical report, 2024.
- [13] A. E. Roth. The economics of matching: Stability and incentives. *Mathematics of Operations Research*, 7(4):617–628, 1982.
- [14] A. E. Roth. The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.
- [15] P. R. Sackett, C. Zhang, C. M. Berry, and F. Lievens. Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11):2040–2068, 2022.
- [16] F. L. Schmidt and J. E. Hunter. The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2):262–274, 1998.
- [17] C. Li, E. Fisher, R. Thomas, S. Pittard, V. Hertzberg, and J. D. Choi. Competence-level prediction and resume & job description matching using context-aware transformer models. In *Proceedings of EMNLP*, 2020.
- [18] N. Vanetik and G. Kogan. Job vacancy ranking with sentence embeddings, keywords, and named entities. *Information*, 14(8):468, 2023.
- [19] Resume2Vec. Transformer encoder-LLM decoder architectures for resume retrieval. Technical report, 2025.
- [20] ConFit v2. Hypothetical resume embeddings and runner-up hard-negative mining for resume-job matching. Technical report, 2025.
- [21] MPNET-RGCN. Relational graph convolutional networks over the ESCO skill ontology for job-candidate matching. Technical report, 2026.
- [22] Anonymous. LLM-driven active listwise tournaments with Plackett-Luce aggregation for candidate ranking. In *Proceedings of EACL*, 2026.
- [23] FairSinkhorn. Optimal transport with group fairness constraints for equitable resource allocation. Technical report, 2024.
- [24] UCB contextual bandits for hiring exploration. Balancing exploitation and exploration in automated candidate selection. Technical report, 2024.

A Network Construction Detail

A.1 Node and Arc Enumeration

For N active candidates and M active jobs with capacity K :

- **Nodes:** 1 (source) + N (candidates) + $2M$ (tiered job nodes) + 1 (sink) = $N + 2M + 2$.
- **Arcs:** N (source-to-candidate) + $2NM$ (candidate-to-tiered, minus pruned) + $2M$ (job-to-sink) = $O(NM)$.
- **Supply:** $\min(N, \sum_k K_k)$.

A.2 Cost Scaling

OR-Tools’ SimpleMinCostFlow requires integer costs. We multiply float scores in $[0, 1]$ by SCALE = 10,000 and negate. The premium bonus β is added before scaling: cost = $-\lfloor (s + \beta) \cdot 10,000 \rfloor$ for premium arcs.

A.3 Infeasibility Handling

When the quality floor prunes most edges, the flow problem may become infeasible at the requested supply level. We handle this by iteratively reducing supply until a feasible solution is found, equivalent to allowing some candidates to go unmatched rather than forcing low-quality matches.

B Simulation Parameters

Table 9: Simulation parameter settings.

Parameter	Value	Rationale
Skill dimensions	79	O*NET-derived taxonomy
Psychological axes	7	See Table 10
Role archetypes	7	Tech role categories
Seniority levels	5	Junior to principal
Team need types	5	Behavioral archetypes
Decay center	10 wk	Sigmoid midpoint
Hire probability	0.15s	Score-weighted
Quality floor	0.0	Default; 0.3 in experiments
Premium bonus β	$0.7\sigma(S)$	Adaptive; scale-invariant
Ensemble allocation	2W+2T+1E	Per job per cycle

Table 10: Psychological profile axes.

Axis	Construct
1	Cognitive ability
2	Conscientiousness
3	Adaptability
4	Collaboration
5	Communication
6	Emotional stability
7	Proactive initiative