Rival Referrals

Business-Stealing, Countercyclical Congestion, and the Limits of Search-Effort Policy in Job-Search Networks

Joseph Kopecky*

August 2026

Abstract

Job information transmitted through social networks is often rival. An employed worker may broadcast a vacancy widely, but is limited in the number of direct referrals she can make. I microfound a job-search network model in which referral rivalry, rather than a free elasticity, generates the convex network effects macroeconomic models assume, and show that a single statistic — the share of delivered referrals whose recipient was the referrer's only usable candidate — governs the economics of search effort. It is the fraction of the effort margin that creates matches rather than stealing them, its complement is the elasticity at which searchers crowd one another's referral chances, and it prices the queue-relief externality of cross-sector connections. In the model it falls sharply in recessions, so that under pure rivalry effort subsidies finance mostly business-stealing exactly when policymakers most wish to reach for them. The statistic and the rival/non-rival blend are in principle measurable from referral micro-data, and the headline results survive on fixed, non-mean-field networks.

JEL classification codes: D62, D85, E24, J64

Keywords: job-search networks, referrals, rivalry, business-stealing, congestion, mean-field limits

*Trinity College Dublin (jkopecky@tcd.ie). Replication code and the machine-verification ledgers behind the paper's quantitative claims are available from the author on request.

1. Introduction

Most jobs are found through people. Granovetter's Newton survey put numbers to that observation half a century ago — a majority of jobs in his sample were found through personal contacts (Granovetter, 1995) — and his companion finding was that the ties doing the work are disproportionately weak ones (Granovetter, 1973). The decades of evidence since put the informal share of hiring somewhere between thirty and sixty percent (Ioannides and Datcher Loury, 2004; Topa, 2011). Economists have built this fact into theories of how networks transmit job information, and the workhorse models make a quiet assumption about the good being transmitted: that information, once held, can be given to everyone. A vacancy announcement works that way. A referral does not. An employed worker who learns of an opening at her firm can tell every unemployed person she knows, but ultimately she can only vouch for a limited number. When she spends her recommendation on one friend, the others wait. And whether job information is repeated for free or spent once turns out to matter a great deal from the perspective of a policymaker.

This paper is about the rival case, and its central object is easy to state. When leads are scarce, a searcher's effort — keeping her network warm, checking in, being top of mind when a referrer has something to give — does one of two things. Sometimes she is the only candidate her referrer can actually reach, and her effort causes a match that would not otherwise have happened. The rest of the time the referrer had somebody else, and her effort moves her up a line that another searcher then stands further back in. I call the first fraction the *sole-candidate share*. It is a number between zero and one, and it is in principle countable in referral records. I will argue that it informs the policy economics of job-search networks. It is the fraction of the search-effort margin that creates matches rather than stealing them, and its complement is the elasticity at which searchers crowd one another's referral chances. It prices the externality that makes connections across sectors socially valuable. In the model it also moves with the business cycle: queues behind referrers lengthen in recessions, so search effort is at its most appropriative exactly when subsidizing it is most tempting.

The vehicle is a deliberately spare model. Workers hold a stock of social ties that behaves like capital. This stock is built while employed, decaying while unemployed, decaying fastest when the people at the other end are unemployed too. Job information travels along ties under one of two polar protocols: broadcast, or one-lead-one-referral in the manner of Calvó-Armengol and Jackson (2004). A share μ_{NR} of information travels the first way; the paper treats that share as a structural parameter to be measured rather than a modeling decision to be defended.

The paper makes four contributions. First, I show that rivalry *generates* the convex network effects that quantitative network models otherwise assume. Amplification through networks requires increasing returns somewhere in matching (Diamond, 1982), but the aggregate matching function is constant-returns in the data (Petrongolo and Pissarides, 2001); rivalry supplies the convexity inside the referral channel, which aggregate matching estimates do not separately identify. Calvó-Armengol and Zenou (2005) already derive a word-of-mouth matching function that fails constant returns because unemployed contacts compete for the same lead; what is new is the form the failure takes. Under one-lead-one-referral, the effective elasticity of referral matching with respect to network employment is two to seven rather than one, disciplined by an observable contact scale, while tie capital saturates this margin. The amplification that drives network models of scarring is a prediction of the referral protocol, not a parameter chosen to produce it. Second, I characterize the effort margin: the private return to search effort decomposes exactly into match creation and appropriation, the creation share is the sole-candidate share s_{piv} with closed form $xe^{-x}/(1-e^{-x})$ in the representative queue intensity x, and s_{piv} falls, at the calibration, from roughly 0.80 in normal times to 0.43 at the trough of a severe slump. The policy consequence is a sign threshold: search-effort subsidies are warranted only if μ_{NR} exceeds a cutoff that is about 0.18 in normal times and 0.71 in recessions. Third, I put the mean-field method that makes all of this tractable on firm ground: a propagation-of-chaos theorem for the tie-and-referral process, and an averaging theorem whose error term has an economic interpretation. It is the selection of the unemployed pool toward the tie-poor, which is how, in the model, scarring concentrates on the least connected. Simulations on fully fixed networks then bound the cost of annealing, at roughly five to twenty percent in the conservative direction. Fourth, I show μ_{NR} and s_{piv} are identified from referral micro-data by a transparent design (yield per employed contact, flat in searcher density under broadcast, falling under rivalry) and supply an estimator whose simulated precision (± 0.04) settles the policy sign question by more than an order of magnitude, along with a precise account of the data that design requires.

My work touches on several literatures. The theory of networks in labor markets has had rival information at its core since Calvó-Armengol and Jackson (2004) and Calvó-Armengol and Jackson (2007), whose token protocol I use; Calvó-Armengol and Zenou (2005) showed that word-of-mouth competition breaks constant returns, and Galenianos (2014) that referral matching deteriorates in slumps because employed referrers grow scarce — a supply mechanism, distinct from the queue crowding studied here. That work treats rivalry as a structural assumption rather than a measurable share, and none of it prices the pivotal margin of search effort. A second strand studies the efficiency of network search directly: search intensity can be over-provided when workers coordinate on the costly formal channel (Cahuc and Fontaine, 2009), and corrective referral instruments have been derived against bargaining and contact-spillover wedges (Zaharieva, 2013; Stupnytska and Zaharieva, 2017). Those instruments are all subsidies. Rivalry pushes the correction toward a tax. A third strand formalizes rivalry as a property of meeting technologies (Eeckhout and Kircher, 2010; Lester, Visschers, and Wolthoff, 2015; Cai, Gautier, and Wolthoff, 2017). The token protocol I study is, in their taxonomy, a bilateral, non-invariant technology. Their efficiency-restoring instrument, meeting fees posted by the meeting's owner, is not available to referrers, which is why an unpriced margin survives here. The business-stealing logic itself is old (Mankiw and Whinston, 1986), and the rule that pivotal contributions must be compensated for search to be efficient is Mortensen (1982)'s; the contribution here is to locate both in job-search networks, give the stealing share a closed form and a cyclical path, and tie all three of its policy prices to one statistic. On the empirical side, a large body of work measures what referrals do — for firms and match quality (Burks, Cowgill, Hoffman, and Housman, 2015; Brown, Setren, and Topa, 2016; Pallais and Sands, 2016; Dustmann, Glitz, Schönberg, and Brücker, 2016), through register-based links (Kramarz and Skans, 2014; Cingano and Rosolia, 2012; Glitz, 2017), and through network crowding in reduced form (Beaman, 2012; Wahba and Zenou, 2005; Galeotti and Merlino, 2014; Schmutte, 2016). The crowding studies, in particular, find the evidence for rivalry. Job-finding through contacts deteriorates where networks crowd, whether the crowding comes from an influx of similar arrivals (Beaman, 2012), from population density beyond a threshold (Wahba and Zenou, 2005), from separations filling the queue of searchers (Galeotti and Merlino, 2014), or from other workers' reliance on referrals itself (Schmutte, 2016). However, none of these estimate the composition of referral information or an elasticity with the structural interpretation the sole-candidate share supplies. Estimating either requires observing offers received, not only the hires that result, and the job-search supplement of the Survey of Consumer Expectations now records exactly that (Faberman, Mueller, Şahin, and Topa, 2022).

The corrective half of the paper belongs to the tradition on the efficiency of decentralized search. Since Hosios (1990), whether a search economy provides the right amount of matching effort has turned on how the congestion externalities in the matching function are split (Pissarides, 2000), with competitive and directed search giving conditions under which the split comes out right (Moen, 1997; Acemoglu and Shimer, 1999). The cyclical warning is not new either. In Michaillat (2012), jobs are rationed in slumps and marginal search mostly redistributes them; Landais, Michaillat, and Saez (2018b,a) derive the corresponding correction, a rat race in market tightness that strengthens in recessions; and the displacement experiments record the same pattern (Crépon, Duflo, Gurgand, Rathelot, and Zamora, 2013; Gautier, Muller, van der Klaauw, Rosholm, and Svarer, 2018; Cheung, Egebark, Forslund, Laun, Rödin, and Vikström, 2025). The externality here is a different animal. It runs not through vacancy creation, a bargained wage, or market tightness, but through the composition of a fixed flow of referrals, and the statistic it turns on is a within-channel business-stealing margin, countable in referral records rather than estimated from cross-market policy variation. The difference matters most when a shock is concentrated: after a mass layoff or in a declining region, aggregate tightness can be normal while referral queues crowd, and there the aggregate corrections read approximately zero while the network correction does not. Its classical analogue is the ranking externality of Blanchard and Diamond (1994) — a searcher who advances in a queue pushes another back — which is exactly the reallocation the appropriation share prices. The effort game of Section 5, for its part, is a mean-field game in the sense of Lasry and Lions (2007) and Carmona and Delarue (2018), played through the queue intensity; I use its fixed-point logic and not its analytical machinery.

A companion paper takes the laws of motion validated here into a full macroeconomic model of network-capital scarring, and derives there a dynamic counterpart of the match-value wedge that the policy section treats as a free parameter. This paper investigates the mechanism, while Kopecky (2026) pursues its aggregate consequences. Every policy threshold below is characterized over the full range of the wedge.

The paper proceeds as follows. Section 2 lays out the model. Section 3 establishes the mean-field foundations. Section 4 contains the core results on rivalry; Section 5 their policy content. Section 6 tests everything on fixed networks, Section 7 develops identification and the estimator, and Section 8 concludes. Proofs and constructions are in the appendices, and every quantitative claim in the paper is backed by a machine-checked ledger in the replication package.

2. A Model of Ties with Rival Information

Two ingredients drive everything in this paper. The first is a stock of social ties that behaves like capital. It builds during employment and decays while unemployed, decaying faster when a worker's contacts are unemployed too. The second is a distinction between whether job information traveling along those ties is a public good or a private one. A vacancy announcement costs a referrer nothing to repeat; a personal recommendation is spent once. The model makes that choice a parameter, and studies how the rival case affects the economics of search.

2.1. Workers and ties

Time is discrete (a period is a quarter). Worker i has an employment status $z_i \in \{0, 1\}$ and a tie stock $n_i \in [0, 1]$, normalized so that $c \cdot n_i$ is the expected number of active contacts, with c the number of contacts a worker maintains at full capacity. Employed workers separate with probability s each period; the unemployed find work through a direct channel, with probability f_{dir}, and through a referral channel described below.

Ties are built at work. While employed, a worker adds $\phi(1 - n_i)$ to her stock each period; while unemployed she adds nothing. The premise — that the ties worth having, ties to employed workers, accumulate during employment and deplete during unemployment — follows Bramoullé and Saint-Paul (2010), whose "economic inbreeding" assumption makes employed–unemployed links the rare ones; my formation rule is a starker special case, shutting down formation during unemployment entirely. Either way the tie stock is a form of capital rather than a fixed trait: a worker's network embodies her employment history. Tie dynamics of this kind have a precedent: Zenou (2015) models job information arriving through weak and strong ties whose use evolves with employment. The state variable here is different, a stock of ties rather than their strength.

Decay is where the model departs from a standard capital stock. A tie is maintained by two people, and what happens to it depends on the status of both:

Table 1: *Joint-status tie decay.*

tie between	decay rate	per quarter
two employed workers	d_{ee}	0
employed and unemployed	d_{eu}	0.11
two unemployed workers	d_{uu}	0.24

Calibration values (Appendix A, backed out of the reduced-form amplification $\kappa = 1$ in A.1). A coworker tie self-maintains, $d_{ee} = 0$; the other two rates order as the text describes.

The ordering in Table 1 is the critical mechanism: a tie between coworkers renews itself daily; a tie with one side out of work survives on the employed side's attachment; a tie between two unemployed workers has no workplace to live in and no employment news to carry. At the calibration used throughout (Appendix A), ties between two unemployed workers decay at roughly twice the rate of employed–unemployed ties. That gap is not a free parameter of convenience, but rather a statement about tie survival that any panel recording relationships and both parties' employment could test directly.

Matching between tie ends is *annealed*. This means that each period, the partner at the end of each unit of tie stock is a fresh draw from the population, weighted by tie stock. The operative aggregate is therefore the tie-weighted employment rate

$$\tilde{e}_w = \frac{\sum_i n_i z_i}{\sum_i n_i}, \tag{1}$$

the probability that the person at the other end of a randomly chosen tie is employed. Annealing is the tractability assumption of the paper, and I do not take it on faith. In Section 6, I re-run every headline result on fixed networks, the other extreme where a worker's contacts are the same people period after period, with all findings remaining intact.

Under annealing, an unemployed worker's per-tie decay rate is $d_{eu}\tilde{e}_w + d_{uu}(1 - \tilde{e}_w)$, so that her ties decay faster when the people at the other end are themselves out of work. This is the doom loop the model is built to study, stated at the level of a single relationship.

2.2. Job information: broadcast or token

Referral information arrives through employed contacts, and the paper studies two polar protocols for how it travels.

Under the *broadcast* protocol (non-rival, NR), an employed worker's job information — "my firm is hiring, apply" — reaches each of her unemployed contacts independently, at rate α per tie unit. An unemployed worker with tie stock n therefore has referral hazard $\alpha\, n\, \tilde{e}_w$, which is linear in her own ties and linear in the employment of the network around her. Nothing one searcher receives is taken from another.

Under the *token* protocol (rival, R), an employed worker receives one usable lead with probability α_R per period (one vacancy she can vouch for) and passes it to exactly one unemployed contact, chosen uniformly among those of her ties that are currently usable. If no usable unemployed contact exists, the lead expires. Throughout, a tie is *usable* if the worker at its other end is unemployed and reachable this period. Search effort will enter by scaling the probability that a tie is reachable when a referrer needs one (Section 4.2). This is the protocol of Calvó-Armengol and Jackson (2004), and it is deliberately polar. Referral programs in practice take many forms: Burks et al. (2015), studying nine large firms in three industries, find referral bonuses paid per hire and employees who make anywhere from zero to several referrals. What the token isolates is the feature these forms share — a recommendation, once spent on one candidate, is not available to another.

Remark 1 (Why one token is not a constraint)*.* The single-token formulation may look like an assumption doing real work. It is not, for two reasons. First, the deep scarcity is the opening, not the message. Consider instead a referrer who tells *every* usable unemployed contact about her single vacancy, with the firm hiring uniformly among those who apply. The position is filled exactly when at least one usable contact exists, and each candidate wins it with the same lottery share as under the token. The mathematics of everything that follows would be unchanged, with the winnowing merely relocated from the referrer's discretion to the firm's hiring stage. What cannot be relaxed is that one vacancy yields one hire. Second, there are good reasons the winnowing often does happen at the referrer. An endorsement stakes the referrer's standing with her employer, which makes vouching a deliberately rationed resource — referrals carry information about candidates (Montgomery, 1991), referrers respond to incentives in choosing whom to put forward (Beaman and Magruder, 2012), and firms lean on the referrer's continued presence as a bond (Heath, 2018). The token is the reduced form of that rationing.

The object that organizes everything downstream is the referrer's *usable-contact intensity*,

$$x \;=\; c\,n\,(1-\tilde{e}_w), \tag{2}$$

the expected number of unemployed contacts an employed worker with tie stock n can actually reach. With Poisson contact counts, a token holder delivers her lead with probability $1-e^{-x}$, and x governs how crowded the queue behind each referrer is. In good times x is small; most contacts are employed, and a searcher is often the only candidate her referrer has. In bad times x is large, and every lead has a line behind it. Section 4 makes this observation quantitative; it is the engine of the paper's countercyclical results.

Reality is presumably neither pole. I let a share $\mu_{\mathrm{NR}} \in [0,1]$ of referral information travel by broadcast and the remainder by token, and treat μ_{NR} as a structural parameter that I will later identify in the data. The right reading of μ_{NR} is behavioral rather than technological: it indexes the share of network-transmitted opportunity whose marginal recipient does not crowd out another. Announcements that steer different contacts toward different openings sit at one pole; exclusive endorsements of a single slot at the other; and a channel that mixes the two (a mass announcement that concentrates applicants on one opening, say) loads onto the blend in proportion to how rival its *outcomes* are, which is exactly the object my later policy analysis needs.

2.3. Aggregation and the macro reduced form

Averaging over workers gives two aggregate states: the employment rate $\bar{e}$ and the mean tie stock K. Under broadcast, the model's aggregate laws of motion are

$$\bar{e}' = \bar{e}(1-s) + (1-\bar{e})\big[f_{\mathrm{dir}} + \alpha K \tilde{e}_w\big], \tag{3}$$

$$K' = K + \phi\,\bar{e}\,(1-K) - (1-\bar{e})\,K\big[d_{eu}\tilde{e}_w + d_{uu}(1-\tilde{e}_w)\big]. \tag{4}$$

Versions of these laws — a slow-moving network capital stock whose decay accelerates when network employment falls — appear as reduced forms in quantitative work on labor-market scarring, concretely in the companion paper (Kopecky, 2026), whose reduced form this model microfounds. The bracket in (4) is exactly the aggregation of the joint-status decay table above (the affine change of variables is in Appendix A), so the reduced form's "amplification" parameter is not a modeling choice: it is the u–u versus e–u tie survival gap, and it is observable.

Two questions now present themselves. Do these deterministic aggregate laws actually follow from the individual process, or are they an approximation with hidden slack? And what replaces the linear referral term $\alpha K \tilde{e}_w$ when information is rival? Section 3 answers the first question with a pair of limit theorems. Section 4 answers the second.

2.4. Numbers, for illustration

Every magnitude quoted in the paper comes from one quarterly calibration, held fixed across every experiment; I call it the baseline calibration. Its discipline is standard. The separation and direct-finding rates deliver a steady-state unemployment rate of six percent; referrals account for $\chi = 0.40$ of hires, the middle of the surveyed one-third-to-one-half range; the reduced-form amplification $\kappa = 1$ of Appendix A maps to the decay rates of Table 1; and workers maintain $c = 10$ contacts at capacity, swept from 5 to 20 wherever the contact scale matters. The full parameter table is in Appendix A.

The referral slopes are set so that both protocols deliver the identical steady state: total referral finding equals the referral share of hires at the same $(\bar{e}, K)$ whether information travels by broadcast, by token, or by any blend. Every protocol comparison in the paper is therefore a comparison of mechanisms, never of steady states. Finally, these numbers discipline illustrations; they are not estimates. Section 7 is about what estimating the objects that matter would take.

3. Mean-Field Foundations

The aggregate laws (3)–(4) treat the economy as two deterministic numbers. The individual process of Section 2 is neither deterministic nor two-dimensional: each worker carries her own history, and aggregates are averages over many correlated fates. This section establishes that the reduced form is not an approximation adopted for convenience but the exact limit of the individual process, and that in the one place where the reduction does lose something, the loss is itself informative.

Both results are stated here with the arguments needed to follow their logic; full proofs are provided in Appendices B and C.

3.1. The mean-field limit

Consider N workers evolving as in Section 2, each with an independent uniform draw each period and interacting through a short list of population aggregates: the tie-weighted employment rate $\tilde{e}_w$ of (1) under broadcast, and additionally the delivered yield of referrals under the token protocol, each a bounded functional of the current population. Let m_t^N collect these aggregates in the N-worker economy, and let m_t denote their deterministic counterpart — the flow generated by the McKean–Vlasov dynamics, in which a single representative worker responds to the deterministic aggregate that her own law generates.

Theorem 3.1 (Propagation of chaos)**.** *Fix a horizon T and suppose the initial tie stocks are bounded below by some $n_0 > 0$. Then there is a constant $C(T)$, explicit in the model's primitives, such that*

$$\sup_{t \le T} \mathbb{E}\left|m_t^N - m_t\right| \;\le\; C(T)\, N^{-1/2},$$

and for any fixed group of k workers, their joint law at each $t \le T$ converges to the product of k independent copies of the representative worker's law, at the same rate. The statement and rate are unchanged with S sectors and a fixed mixing matrix, and under the token protocol and rival/non-rival blends, where the interaction includes the delivered-yield functional (Appendix B).

The economic interpretation has two parts. First, aggregates are non-random. At the scale of a labor market, the employment rate and the mean tie stock evolve by laws of motion, rather than chance, and the error of treating them so vanishes at the standard $\sqrt{N}$ rate. Second, and less obviously, individual workers become independent. The correlation between any two workers' fates is a finite-population artifact of order $1/N$. Whatever co-movement survives in the limit travels entirely through the aggregate state. That is precisely the structure the reduced form assumes, and the theorem says the assumption is the $N \to \infty$ limit of the exact economy, not an approximation to it.

The proof follows the standard synchronous-coupling scheme of Sznitman (1991); the only model-specific ingredient is Lemma 1, which establishes an invariant domain on which the tie-weighted-employment functional is Lipschitz. Under the token protocol $\tilde{e}_w$ is joined by the delivered-yield functional, equally bounded and Lipschitz on that domain, so the same scheme carries the rival case and the blend; Appendix B records the one point of care, the searcher-pool denominator. The theorem is stated for this semi-analytic delivered-yield hazard (the object the aggregate laws and the closure use), with the discrete one-recipient-per-lead allocation reproducing it to within 0.2% in the ledger (Appendix F).

The difficulty the lemma resolves is easy to state. The interaction $\tilde{e}_w$ is a ratio, and ratios misbehave when their denominators approach zero: if aggregate tie stock could collapse, small perturbations in the population could move $\tilde{e}_w$ violently, and the coupling argument that drives the theorem would not close. The model itself rules this out:

Lemma 1 (Invariant domain)**.** *In every version of the process (finite N or the limit), tie stocks satisfy $n_i^t \geq n_0(1-d_{uu})^t$ almost surely. Hence on any horizon T every denominator entering $\tilde{e}_w$ is bounded below by a deterministic constant, and $\tilde{e}_w$ is Lipschitz along the entire evolution.*

The proof is three lines. Ties decay multiplicatively at a bounded rate, so they cannot reach zero in finite time. Thus the model's own structure supplies, pathwise and with no probabilistic argument, exactly the regularity the standard machinery needs. The remaining steps (Lipschitz estimates, the coupled construction, a law of large numbers for the auxiliary system, and a discrete Gronwall argument) are standard, with constants tracked explicitly throughout.

Two qualifications are needed. First, the constant $C(T)$ grows with the horizon, as finite-horizon coupling constants do. In my calibration the measured errors are far smaller than the theorem's bound and are flat over time, which a uniform-in-time refinement would capture but which I do not pursue. Second, the theorem concerns the annealed process. What happens when the network is fixed rather than redrawn is an empirical question, and Section 6 answers it by simulation rather than assumption.

3.2. The two-state reduction, and what its error is made of

Theorem 3.1 delivers a deterministic limit, but that limit is still an entire distribution of workers over employment states and tie stocks. The reduced form (3)–(4) compresses the distribution to two moments, and the compression involves one substitution: where the true referral outflow depends on the mean tie stock *of the unemployed*, $\mathbb{E}[n \mid u]$, the two-state map uses the population mean K. Are these close?

They are close due to the relatively fast rate of churn. Workers move between employment and unemployment about twenty-five times faster than tie stocks adjust, so a worker's ties average over many employment spells. As a result, the covariance between today's status and today's tie stock stays small. The same slowness that makes network capital an engine of persistence is what validates aggregation. One assumption does two jobs.

This timescale logic can be formalized. Scale the tie dynamics by λ (so $\lambda = 1$ is the calibration and $\lambda \to 0$ is the limit of perfect timescale separation), and consider any stationary state of the limit dynamics.

Theorem 3.2 (The averaging closure)**.** *Under conditions verified at the calibration, every stationary law of the McKean–Vlasov dynamics satisfies*

$$\mathrm{Cov}(n,z) \;=\; \lambda\, C_\infty + O(\lambda^{3/2}),$$

with C_∞ explicit in the model's primitives (Appendix C). Consequently the pool-selection gap and the tie-weighted employment tilt satisfy, to the same order,

$$\mathbb{E}[n \mid u] - \mathbb{E}[n] \;=\; -\frac{\lambda C_\infty}{1-\bar{e}}, \qquad \tilde{e}_w - \bar{e} \;=\; \frac{\lambda C_\infty}{K}.$$

At the calibration the constant delivers a pool-selection gap of -10.6% of the mean tie stock K (a raw $\mathbb{E}[n \mid u] - \mathbb{E}[n] = -0.074$ from the displayed formula) and a tie-weighted employment tilt of $+0.63$ percentage points; simulations of the full process return -10.3% and $+0.63$. The measured remainder behaves like λ^2, better than the theorem's exponent, and the appendix records which single term resists the sharper bound and why.

The closure error, $\mathrm{Cov}(n,z)$, supplies further economic insight. The two displayed quantities are its two faces. The first says the unemployed pool selects toward the tie-poor. That is, workers with thin networks exit unemployment more slowly, so at any moment the unemployed hold fewer ties than average. Thus the two-state map, which cannot see the composition of the pool, overstates referral outflow by about ten percent. That selection is, in the model, the mechanism by which network scarring concentrates on the least-connected workers. The second face is the friendship paradox: because partners are drawn tie-weighted, the person at the other end of a randomly chosen tie is better-employed than the average worker, by just over half a percentage point at the calibration, a tilt that emerges from employment histories. The closure error of the reduced form is not simply noise, but rather the model's distributional prediction, quantified.

One more feature of C_∞ deserves note. Its derivation is off by about a fifth of the total if one treats employment status as evolving independently of ties. Tie-rich searchers exit unemployment faster, status responds to ties, and this feedback enters the covariance at first order. The who-scars selection is not a passive consequence of the dynamics; it helps build the very covariance that measures it.

This paper's sharpest results are cyclical, but the theorem is about a stationary relationship. Ultimately, it governs the one critical compression that the reduced form makes (a distribution collapsed to two moments), not the path the economy takes through a recession. The countercyclical fall in s_{piv} of Section 4.3 does not rest on it: s_{piv} is a cross-sectional object (Proposition 2), read off the simulated distribution at each date, and that distribution along a recession path is delivered by Theorem 3.1, whose coupling covers a time-varying separation rate. What is left to simulation is the accuracy of the two-state map (3)–(4) in transition rather than at rest, and the finite-population and fixed-network runs of Section 6 carry shock paths and bound that error next to the stationary one.

4. Rivalry

Rivalry does two things to this economy, which this section takes up in turn. First, in Section 4.1, I show that it bends the aggregate matching function, as network employment starts to matter more than proportionally, and tie capital starts to matter less. Second, in Sections 4.2 and 4.3, I show that it changes what a unit of search effort buys: some of what a searcher gains, another searcher loses. The first result is why these networks matter for macroeconomics, as the convexity that amplification requires arises endogenously rather than by assumption. The second delivers the paper's central policy statistic: the sole-candidate share, which is the fraction of search effort's private return that creates matches rather than appropriating them from other searchers. Formally, the sole-candidate share s_{piv} is the fraction of delivered referrals whose recipient was the referrer's only usable candidate at the moment of delivery. In particular, Section 4.3 shows that this statistic falls by nearly half in slumps, weakening the public return to encouraging search precisely when a policymaker might be most interested in subsidizing it.

4.1. Emergent convexity and saturation

Under broadcast, the referral hazard is simply $\alpha\, n\, \tilde{e}_w$: the aggregate referral flow is linear in tie capital and linear in network employment, which is the form (3) assumes. Under the token protocol the same two aggregates enter through entirely different channels, and neither enters linearly.

Proposition 1 (Rivalry bends the matching function)**.** *Take any cross-section with employment rate $\bar{e} \in (0,1)$ and usable-contact intensities $x_j = c\, n_j(1-\tilde{e}_w)$ not all zero. Under broadcast the elasticity of the aggregate referral flow is one with respect to network employment and one with respect to tie capital. Under the token protocol both elasticities depart from one, in opposite directions, and both are governed by the sole-candidate share s_{piv}. The tie-capital elasticity is the sole-candidate share itself,*

$$\varepsilon_K \;=\; s_{\text{piv}} \;\in\; (0,1), \tag{5}$$

because a referrer needs only one usable contact to place her lead, so a uniform rise in tie stocks delivers a match only in the states where it makes an otherwise idle referrer pivotal, exactly the sole-candidate states. It is strictly below one whenever some referrer can reach more than one searcher *at once, and equals $s_{\text{piv}}(x) = xe^{-x}/(1-e^{-x})$ under Poisson usable counts of mean x. The effective employment elasticity exceeds one; when every referrer faces the same usable-contact intensity x (the representative case),*

$$\eta_{\text{eff}} \;=\; 1 \;+\; \frac{\bar{e}}{1-\bar{e}}\,\big(1 - s_{\text{piv}}(x)\big) \;>\; 1, \tag{6}$$

its excess over one being the competition-relief term $\frac{\bar{e}}{1-\bar{e}}(1-s_{\text{piv}})$, which is absent under broadcast and vanishes only as the appropriation share $1-s_{\text{piv}}$ does. Heterogeneity across referrers raises η_{eff} further; Appendix D carries the general expressions and the derivations.

The two distortions come from one scarcity. When network employment rises, a searcher gains twice: more of her contacts hold leads, and fewer rival searchers stand behind each one. The second, competition-relief term is what pushes η_{eff} above one, and it has no counterpart under broadcast, where rivals are irrelevant by construction. Tie capital, by contrast, saturates: a referrer needs only one searching contact to place her lead, so once most referrers can reach somebody, additional ties add reach the token cannot use. Hence $\varepsilon_K < 1$, falling as contacts multiply. Both statements are the sole-candidate share seen from two sides: s_{piv} is the fraction of a marginal activation that creates a match rather than reordering the queue, and so is what an extra unit of tie capital saturates to ($\varepsilon_K = s_{\text{piv}}$); its complement $1-s_{\text{piv}}$, the appropriation share, is what drives the competition relief that lifts η_{eff} above one.

At the calibration the elasticities take the values in Table 2, evaluated on the exact discrete queue rather than the representative approximation.

Table 2: *Emergent convexity and saturation under the token protocol.*

contacts at capacity c	ε_K	η_{eff} (steady state)	η_{eff} (slump)
5	0.90	2.7	2.2
10	0.80	4.3	3.1
20	0.64	7.1	4.2

Elasticities of Proposition 1 at the calibration of Appendix A, computed on the exact cross-section. The ε_K column is the exact sole-candidate share (5) (equivalently, s_{piv} at each contact scale); the representative form (6) understates η_{eff} by the referrer-heterogeneity term, so the table reports the larger exact values. Both broadcast elasticities equal one. The slump column evaluates the trough of the slump experiment of Section 4.3 (separations quadrupled for two years), holding tie stocks at their pre-shock level; that section lets them erode.

This is the section's message for quantitative modeling. Amplification through networks requires a matching elasticity above one somewhere, and the aggregate evidence says it is not in the aggregate matching function (Petrongolo and Pissarides, 2001). In reduced-form work the convexity is therefore a free parameter, set to whatever the target moments require. Under rivalry it is not free: convexity above one and saturation below it are what one-lead-one-referral produces, and the two are a joint prediction, since the same sole-candidate share sets both. The qualitative possibility is the word-of-mouth congestion result of Calvó-Armengol and Zenou (2005). This section adds to their result a closed-form solution for both elasticities, in which every term is either directly observable or the sole-candidate share. Convexity stays above one in a deep slump, though lower than at the steady state (the competition-relief term weakens when searchers are plentiful), so rivalry delivers an amplification *region*, not a knife-edge. The convexity is an allocation effect, not an aggregation one: η_{eff} and ε_K are elasticities of the token flow placed with searchers (the leads themselves), not of the finding rate that turns tokens into hires, so they measure how the queue reallocates a fixed lead flow, whatever the form that aggregates a searcher's tokens into a job.

4.2. The sole-candidate share

Now fix the aggregate environment and ask the question a policymaker would. If one searcher raises her search effort — checks in with her contacts more often, keeps her ties warm — what does the effort purchase? Under broadcast the answer is uninteresting. Her hazard rises, nobody else's moves. Under the token protocol her effort acts on a fixed flow of leads, and the accounting is where rivalry gains teeth.

Effort acts as activation, as a searcher's effort scales the probability that each of her ties is usable when a referrer needs one. Her gain then decomposes into two pieces. Sometimes her newly activated tie is the referrer's *only* usable option, and her effort causes a lead to be delivered that would otherwise have expired. In this case her effort is pivotal. The rest of the time the referrer already had somebody, and her effort merely wins a lottery another searcher would have won.

Proposition 2 (Creation versus appropriation)**.** *The private return to search effort under the token protocol is the sum of a pivotal component, which raises aggregate matches one for one, and an appropriation component, which is a pure transfer among searchers. Because delivered referrals equal received referrals identically, appropriation nets to zero in the aggregate. The share of the effort margin that creates matches is*

$$s_{\text{piv}} = \mathbb{P}\big(\textit{the recipient was the referrer's only usable candidate} \mid \textit{lead delivered}\big),$$

and this identity holds for any distribution of ties and any market state. With Poisson usable-contact *counts of mean x,*

$$s_{\text{piv}}(x) \;=\; \frac{x\,e^{-x}}{1-e^{-x}},$$

which falls monotonically from one (empty queues: every activation is pivotal) toward zero (crowded queues: effort only reorders the line).

I will refer to s_{piv} as the *sole-candidate share*, and it is the central object of the paper. Three features make it more useful than the decomposition alone. It is a share of *delivered referrals* — a conditional frequency, in principle countable in referral records, not a construct requiring the model to compute. It depends on the environment through the queue — in the representative case through the single intensity x of equation (2), in general through the distribution of usable-contact intensities across referrers — so everything that moves it is observable. And, as Section 5 shows, the same number prices every correction that rivalry forces on policy.

One measurement remark is worth noting, because three numbers are easily confused. The identity of Proposition 2 defines s_{piv} as an exact ratio on the referrers' usable sets; on the calibrated network it is 0.80, and that is the value the policy section uses. Its homogeneous Poisson evaluation at the mean queue intensity $x = cK(1-\tilde{e}_w) \approx 0.42$ (the closed form $s_{\text{piv}}(x) = xe^{-x}/(1-e^{-x})$) also comes to 0.80, because the lumpiness and the heterogeneity of an actual network pull in opposite directions and here roughly cancel. What does *not* track is averaging the Poisson formula over the queue distribution, $\mathbb{E}_j[x_j e^{-x_j}]/\mathbb{E}_j[1-e^{-x_j}] = 0.58$, which understates the exact share: a referrer with a few strong ties is the sole candidate more often than a Poisson queue of the same mean would be. The exact ratio is what carries over to data, and it is what keeps the steady-state economy near the policy boundary of Section 5, where the Poisson average would place it well inside the taxing region. The closed form is a convenience, accurate at this calibration; Appendix D treats the discrete case.

4.3. Job stealing is countercyclical

Everything in s_{piv} moves with the cycle through the queue intensity $x = cK(1-\tilde{e}_w)$: more unemployment means more searchers standing behind each referrer. The comparative static is immediate and its size is not small, as can be seen in Figure 1. At the calibrated steady state, with unemployment at six percent, $s_{\text{piv}} \approx 0.80$: four fifths of the effort margin creates matches. At the trough of a severe recession, with separations quadrupled for two years and unemployment above twenty percent, the sole-candidate share falls to 0.43. More than half of what search effort buys at the bottom of a slump, it buys by taking from others.

Two refinements discipline the headline. First, the trough figure uses ties as they stand

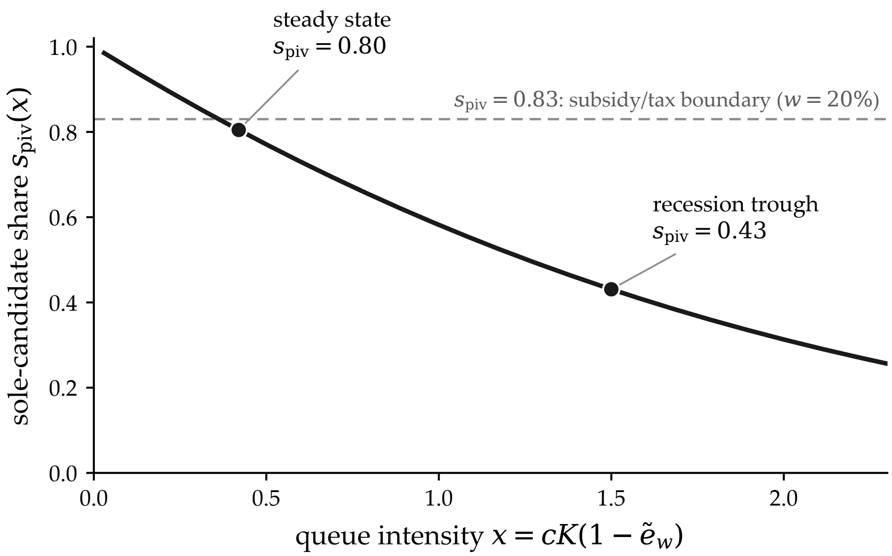

Figure 1: *The sole-candidate share $s_{\text{piv}}(x) = xe^{-x}/(1-e^{-x})$ falls with the queue intensity $x = cK(1-\tilde{e}_w)$. At the calibrated steady state the economy sits just below the corrective-policy boundary of Section 5.1 ($s_{\text{piv}} = 0.83$ at the illustrative match-value wedge $w = 20\%$, w being the proportional excess of a match's social over private value), while a severe recession drives it deep into the region where effort is taxed.*

when the shock hits; a long slump erodes tie capital, which thins the queues and pulls s_{piv} partway back (a market that has been slack for years, with unemployment stabilized near nineteen percent, shows $s_{\text{piv}} \approx 0.71$). The stealing problem is at its worst *early* in a downturn, while networks are still dense and queues are longest. Second, the numbers above are steady-state and trough snapshots of the same calibrated economy, not a sweep over parameters; Appendix D reports the sensitivity, and the monotonicity in x. The economics of the result survives everywhere.

The picture entering the policy section is this. In good times the line behind a referral is short, and helping a worker search harder mostly helps her into a job. In bad times the line is long, and the same help mostly moves her up a queue at her neighbor's expense — exactly when the political demand for search-assistance policy is greatest.

5. Policy: Three Prices, One Statistic

Every way rivalry distorts policy runs through the sole-candidate share. This section makes that claim precise three times: for subsidies to search effort, for the congestion externality among searchers, and for subsidies to cross-sector connections. In each case the correction that rivalry forces is a closed-form function of s_{piv}.

5.1. The sign of the effort instrument

Consider the planner's problem in its most transparent form: a continuum of searchers choose costly effort; effort activates ties; tokens are delivered as in Section 4.2; each match carries a social value that exceeds its private value by a wedge w, reflecting the network externalities of employment.[1] The searcher weighs her private return against her cost; the planner weighs the social return. Write σ^* for the corrective effort instrument this comparison implies, with subscripts R and NR marking the token and broadcast protocols.

Proposition 3 (The sign threshold)**.** *In the static effort economy, the corrective instrument on search effort is a subsidy if $s_{\text{piv}}(1+w) > 1$ and a tax if the inequality reverses. Equivalently, relative to the subsidy a broadcast economy would warrant, the fraction that survives rivalry is*

$$\frac{\sigma^*_R}{\sigma^*_{NR}} = \frac{s_{\text{piv}}(1+w) - 1}{w},$$

which is negative — a net tax — whenever $s_{\text{piv}} < 1/(1+w)$.

The logic is straightforward. A searcher's private calculus counts every referral she wins; the planner counts only the pivotal ones, valued at $(1+w)$, because the appropriated ones are transfers. When queues are crowded enough that the pivotal share falls below the private share of a match's social value, effort is privately profitable and socially wasteful at the margin. This is the free-entry logic of Mankiw and Whinston (1986) and the pivotal-contribution accounting of Mortensen (1982), relocated from firms entering a market to searchers entering a queue.

Consider some numbers at the illustrative wedge of 20%: the threshold is $s_{\text{piv}} = 0.83$, and the calibrated steady-state value is 0.80. Here the calibrated economy sits just inside the taxing region. In normal times the optimal effort instrument is approximately zero, and the intuition that says helping people search harder must help is, under pure rivalry, a rounding error away from the wrong sign. That near-coincidence pairs a calibrated s_{piv} with an illustrative wedge, and the boundary moves one-for-one with the wedge. In a slump the verdict is not close. At the trough value $s_{\text{piv}} = 0.43$ the surviving fraction is -2.4, a substantial tax, precisely when subsidies to search are most politically attractive.

Because the wedge is ad hoc, it is worth saying what the sign result needs from it and what it does not. Set $w = 0$ first: with no social premium on a match, the instrument is a tax whenever $s_{\text{piv}} < 1$, that is, whenever referrals are rival at all. Rivalry alone, correcting

[1] The wedge is an input here, not a result. For illustration I use $w = 20\%$; a dynamic derivation of such a wedge from network-capital scarring is in the companion paper (Kopecky, 2026). The analysis below characterizes the instrument's sign over the whole range $w \geq 0$.

no externality, already makes search effort privately over-provided, because the private return counts the appropriated referrals that the planner nets out. This direction is the opposite of the existing corrective results in the referral channel, which price bargaining and contact-spillover wedges and prescribe subsidies (Zaharieva, 2013; Stupnytska and Zaharieva, 2017). A positive w pushes back, and a subsidy is warranted only once the wedge clears $w^*(s_{\text{piv}}) = (1 - s_{\text{piv}})/s_{\text{piv}}$: at the steady-state $s_{\text{piv}} = 0.80$ that cutoff is 25%, and at the trough $s_{\text{piv}} = 0.43$ it is 133%. So the recession verdict survives any plausible wedge (a slump makes the instrument a tax unless the social value of a match exceeds its private value by more than 130%), while the normal-times sign is the wedge-sensitive one, flipping from tax to subsidy as w passes a quarter. The illustrative 20% falls below the steady-state cutoff, which is what places the normal-times instrument near zero; but the qualitative map — a tax under rivalry at low wedges, a subsidy only at high ones, and a cutoff that climbs steeply as queues crowd — is a property of s_{piv} and holds for every $w \geq 0$.

Neither polar case describes actual referral flows; the operative question is the mix. With a share μ_{NR} of referral information traveling by broadcast, a net subsidy is warranted if and only if

$$\mu_{\text{NR}} \;>\; \bar{\mu} \;=\; \frac{1 - s_{\text{piv}}(1+w)}{1 - s_{\text{piv}}(1+w) + w},$$

and at the illustrative wedge $\bar{\mu} \approx 0.18$ in normal times but ≈ 0.71 in a slump. The threshold falls monotonically in w, from $\bar{\mu} = 1$ at $w = 0$ toward zero as the wedge grows (a larger externality lowers the non-rival share a subsidy requires), so the countercyclical jump in $\bar{\mu}$, and with it the empirical weight the policy question puts on μ_{NR}, is there at every wedge. The sign of search-effort policy is an empirical question, and the parameter that answers it is measurable in principle. That is the paper's central policy claim.

5.2. Congestion, measured by the same statistic

The appropriation component of Proposition 2 is, seen from the other side, an externality. One searcher's activated ties dilute every rival's chances. Its size requires no new analysis.

Corollary 1 (Congestion elasticity). *The elasticity of an incumbent searcher's referral hazard with respect to the usable tie mass of competing searchers is $s_{\text{piv}} - 1$, for any size of the perturbing group. It is zero under broadcast.*

The polar cases discipline this number. A global lottery over all searchers, which is the assumption reduced-form treatments of referral congestion reach for, corresponds to $x \to \infty$ and an elasticity of -1. In other words, every referral won is a referral taken. The local queues of the calibrated economy deliver $s_{\text{piv}} - 1 \approx -0.22$ at the steady state (the fixed-network simulation of Section 6, with the semi-analytic value at -0.20). Congestion among searchers is real and is countercyclical for the reasons of Section 4.3: at the steady state it is roughly a fifth of the global-lottery polar case a mean-field-of-everything intuition would reach for, but that gap narrows as queues crowd. At the trough $s_{\text{piv}} - 1 \approx -0.57$, roughly half the polar -1. Both the sign and its state-dependent attenuation matter for how one reads reduced-form estimates of network crowd-out.

5.3. Bridges and the queue behind workers

A third price concerns a different instrument: subsidies to connections that cross sectors. If employment shocks hit specific sectors, policymakers may wish to subsidize "bridges": connections into healthy sectors, where the rivalry externality is weaker. These bridges have externalities of their own in any network model; rivalry adds one that is invisible under broadcast. When a bridged searcher redirects her search toward the healthy sector, she exits her home queue, and every searcher standing behind her benefits. Write τ for a searcher's expected token receipts per period, with subscripts h and b marking the home and bridged-to sectors and τ_b' the receipts the bridged searcher draws from the destination queue.

Corollary 2 (Queue relief)**.** *A searcher's marginal exit from the home queue raises remaining home searchers' referral flow by fraction $(1 - s_{\text{piv},h})$ of her former receipts: conditional on having received a lead, $s_{\text{piv},h}$ is the probability it would otherwise have expired, and the complement is the probability it passes to a rival. The externality is zero under broadcast. Net of the congestion she imposes on the destination queue, the marginal queue externality of a bridge is $(1 - s_{\text{piv},h})\tau_h - (1 - s_{\text{piv},b})\tau_b'$ (derivation in Appendix D).*

The netting is where the instrument design lives. Bridging a searcher from a slack sector into a healthy one relieves a crowded queue and congests an empty one; at the calibration the net externality is worth about six percent of the searcher's home referral hazard, every quarter. Bridging her into an *equally slack* sector relieves one crowded queue and congests another, and the net is approximately zero. The queue component of a bridge subsidy should therefore target *de-correlating* connections (links to sectors that do not crash when the home sector does) and, echoing Section 4.3, it is worth the most early in a downturn, before tie erosion thins the queues it relieves. Zaharieva (2018) also gives a planner a reason to favor cross-sector connections, as insurance against occupational mismatch. The queue externality is a different reason, and one the sole-candidate share prices.

Instruments that expand the supply of leads (hiring credits above all) carry no appropriation component under either protocol. Tie-formation subsidies are less clean: a searcher's extra ties raise her share of a fixed token flow, so part of their private return is the same reallocation margin as effort. The appropriation-free case is for instruments that add leads or capacity rather than redirect access to a fixed flow. Rivalry is a problem for the allocation of leads specifically; where an instrument creates leads it is left alone, and if anything the relative case for it strengthens.

6. Beyond Mean Field: Fixed Networks

Everything so far rests on annealed matching: the person at the end of each tie is redrawn every period. Real people have the same friends tomorrow, and fixed relationships are exactly where mean-field reasoning could fail — neighbors' fortunes correlate, local clusters of unemployment form, and the averaging that justified Section 3 has no obvious warrant. This section reports what actually breaks. The tools are simulations of the full model on fixed graphs: ties live on the edges of a configuration-model network with realistic sector structure, decay according to the actual neighbor's status, and tokens pass to actual unemployed neighbors. The design, which requires rebuilding the tie layer from scratch rather than patching the annealed simulator, is in Appendix F.

I show the comparison as a three-way race in Table 3. There I report the mean field versus a finite annealed population versus the fixed graph, identically calibrated. These are studied in the steady state, in a transient-shock scarring experiment, and in the experiment that should stress the mean-field assumption hardest: a shock that hits a sector *together with* its main partner sector, so that a worker's own bad luck coincides with her contacts'. Correlated shocks are precisely where neighbors' states stop being independent — and independence across neighbors is what the mean field assumes.

Table 3: *Mean field versus finite annealed and fixed-graph economies, identically calibrated.*

	mean field	annealed (N=48k)	fixed graph (N=48k)
steady state ($\bar{e}$, broadcast)	0.940	0.938	0.937
transient-shock scar (broadcast)	+0.0030	+0.0030	+0.0029
correlated-shock deepening, broadcast ($\bar{e}$)	3.8%	2.5%	3.0%
correlated-shock deepening, token ($\bar{e}$)	24.7%	32.4%	21.9%
standard error (pp)	—	(12.5)	(9.3)
correlated-shock deepening, token (K)	6.9%	6.6%	5.7%
standard error (pp)	—	(0.4)	(0.2)
η_{eff} at $c = 10$ (token)	—	4.31	4.97
congestion elasticity (token)	−0.20	—	−0.22

Deepening is the extra scar from a sector-plus-partner (correlated) shock relative to a sector-plus-unrelated (spread) shock; standard errors in percentage points, from eight seeds with three coupled replicates (Appendix F). The employment ($\bar{e}$) deepening sits near the demographic noise floor, so its standard error is large and the three columns are statistically indistinguishable; the tie-capital (K) channel is an order of magnitude more precise and the variable the mechanism moves, and there the ordering mean field $\geq$ annealed $\geq$ fixed graph is monotone. Elasticities are evaluated on the respective cross-sections; broadcast congestion is exactly zero.

There are three findings, which I discuss in order of importance. First, the mean-field quantification survives, in both information regimes, with corrections of roughly five to twenty percent that run in the conservative direction. On employment the token deepening is noisy: the annealed value (32.4%) sits above mean field (24.7%) and the fixed graph (21.9%) below, but the standard errors are near ten percentage points, so the three are statistically indistinguishable; the apparent annealed overshoot is noise, not evidence that mean field understates the deepening. The verdict rests on the tie-capital channel, both because its standard errors are an order of magnitude smaller and because tie capital is the variable the mechanism actually moves. There the ordering is monotone: mean field 6.9%, annealed 6.6%, fixed graph 5.7%. So annealing adds almost nothing, and fixing the graph shrinks the deepening, to 0.83 of the mean-field figure under the token protocol and to 0.94 under broadcast, whose tie-capital values the table omits and Appendix F reports. Both corrections are conservative. And the deepening is the queue, not the concave aggregator that turns tokens into hires: recomputing the closure with a linear aggregator at the same steady state leaves the token deepening essentially unchanged — 24.6 against 24.7 percent on employment, and slightly larger on tie capital (7.1 against 6.9) — while broadcast, that same linear aggregator without the queue, deepens only 3.8 percent; Appendix F reports the decomposition.

Second, the amplification of Section 4 is not an artifact of annealing: the rivalry-generated convexity is slightly *stronger* on the fixed graph, 4.97 against 4.31 at ten contacts, with the full pattern across contact scales reproduced. Locality concentrates competition relief on the searchers who remain, and the fixed graph, if anything, sharpens the mechanism.

Third, the agreement is not because fixed networks fail to generate local structure. At the trough of the correlated-shock experiment, the correlation of unemployment across actual network edges reaches $+0.08$, twenty-five times its steady-state level. The information annealing discards is really there. It fails to move sector-level aggregates because the between-sector co-movement that drives the correlated-collapse result is carried explicitly by the sector mixing structure, which all three models share, while within-sector clustering washes out of sector averages when hazards are close to linear in neighborhood composition. Where hazards are convex (the token protocol) the wash-out is imperfect, and that is exactly where the corrections are largest and where the local-queue value of the congestion elasticity (-0.22, against the global-lottery polar case -1) parts company with mean-field intuition.

The robustness sweep behind the table also varied the graph itself. Heavy triangle clustering (ring-lattice sectors) and sparser networks (half the contacts) leave the correlated-shock deepening statistically unchanged. The claims are correspondingly scoped: one calibration, one shock design, and configuration-model graphs with no dominant hubs. The hub case is the known danger, because a shock aimed at the most-connected workers, the brokers through whom many ties run, is precisely where a sector-level description would fail. I leave this experiment for future work.

7. Identification and Measurement

Section 5 left two numbers to be measured: the sole-candidate share s_{piv} and the non-rival blend μ_{NR}. This section shows both are identified from referral micro-data, describes the estimator, and reports its performance on data simulated from the model, including from the fixed-network economies of Section 6, which the estimator's assumptions do not strictly cover. It closes with the data required to take this to the field.

Neither number has been estimated, to my knowledge. Structural work with referral networks fixes the transmission technology by assumption: Arbex, O'Dea, and Wiczer (2019) impose a fully rival protocol and state that their object is the structure of the network rather than the technology of transmission, and Lester, Rivers, and Topa (2021) estimate referral effects on offer arrivals and wages with no rivalry margin. The composition of referral information is the estimand these papers set aside.

7.1. The separating moment

Define the *yield*: referral offers received per unit of a searcher's employed-contact mass. The two protocols make opposite predictions about how the yield varies with the number of competing searchers. Under broadcast the yield is a constant. An employed contact passes what she knows regardless of who else is looking. Under the token protocol the yield falls with searcher density, with elasticity $s_{\text{piv}} - 1$ by Corollary 1, because every additional searcher lengthens the queue behind each referrer. In a blended economy the yield across markets m traces

$$y_m \;=\; A_{NR} \;+\; A_R \cdot R_m,$$

a flat broadcast component plus a rival component whose shape R_m declines in market slackness. Identification is then transparent: the *slope* of yield in searcher density tests for rivalry (any decline rejects pure broadcast); the *floor* the yield approaches as queues crowd is the broadcast component, since the rival component is competed away; and the level split at a reference state delivers $\mu_{\text{NR}} = A_{NR}/(A_{NR} + A_R)$.

Two assumptions make this cross-market variation informative. First, the channel levels A_{NR} and A_R are common across markets, so that what varies with m is the queue shape R_m and not the composition or productivity of referral information itself. This is an exclusion restriction, which states that searcher density must vary across markets or over time for reasons orthogonal to μ_{NR}, α, and the contact scale c. A tight labor market and a slack one must differ in how crowded their referral queues are, not in what share of their referrals is rival or how much each contact is worth. Second, the network is taken as given. If c or the tie stock responds to local conditions, slack markets differ in the very primitives the estimator holds fixed, and the cross-market comparison confounds queue crowding with endogenous network formation. Neither assumption is innocuous. Density variation isolated by a plausibly exogenous shock (a plant closing, a migration wave) is what would make them credible, and $\hat{\mu_{\text{NR}}}$ is best read as a yield-share estimand whose mapping to the structural information share the model imposes rather than the data establish. Selection sits on the left-hand side too: offers received are self-reported, and the propensity to report an offer may itself move with search intensity. The design assumes it does not.

7.2. The estimator

The practical question is what data make the split feasible, and the answer contains the section's one methodological finding. The rival shape R_m is close to linear over realistic ranges of market slackness, so with yield data alone the floor is an extrapolation: formally identified, but hopelessly imprecise. In Monte Carlo, a joint nonlinear-least-squares fit of (A_{NR}, A_R) and the queue scale wanders across the unit interval (standard deviation 0.24 on $\hat{\mu_{\text{NR}}}$, against 0.03 for the estimator below) and collapses to the corners. Yield data alone support the rivalry *test*, not the composition estimate.

What restores precision is measuring the queue directly. If the data record employed workers' contacts by employment status (how many unemployed contacts each potential referrer has), then the rival shape R_m is computable from observables (the distribution matters, not just the mean: deliverability is concave in contacts, and using the mean alone leaves a bias of -0.13 on $\hat{\mu_{\text{NR}}}$ at the design's hardest point). The regression becomes linear in the two channel levels, and the two-step estimator (queue shape from the contact module, channel split from the yield curve) performs as in Table 4, on 20 replications of 24 markets of 6,000 workers observed for 50 quarters; the construction is in Appendix E.

Table 4: *The two-step estimator on data simulated from the blended model.*

true μ_{NR}	0	0.25	0.50	0.75	1
mean $\hat{\mu_{\text{NR}}}$	0.011	0.253	0.505	0.744	0.978
RMSE	0.022	0.039	0.039	0.044	0.033
reject broadcast (%)	100	100	100	100	5

Twenty replications of 24 markets of 6,000 workers observed 50 quarters after burn-in, spanning unemployment 3.5 to 23% (Appendix E); RMSE is taken over replications. The last row is the rejection rate of the pure-broadcast null at the 5% level; at $\mu_{\text{NR}} = 1$, where broadcast is the truth, the 5% is the nominal size, not low power.

Bias never exceeds 0.022, and reaches that only at the $\mu_{\text{NR}} = 1$ endpoint. Across the interior of the grid it stays under 0.006. The test of pure broadcast has full power against blends of one half or less at exact size. Two robustness results matter for practice. Fitted to data simulated from the fully fixed networks of Section 6 — discrete contacts, lumpy tie strengths, none of the estimator's Poisson scaffolding — the estimator returns 0.47 against a truth of 0.50: measuring the deliverability functional from the contact distribution, rather than assuming a functional form, absorbs the misspecification. And the fitted yield-density elasticity recovers $s_{\text{piv}} - 1$ directly, so the same regression that splits the blend delivers the statistic all three policy corrections need.

Recall the thresholds these estimates would be held against: $\bar{\mu} \approx 0.18$ in normal times, 0.71 in a slump. An estimator with root-mean-squared error near 0.04 settles which side of either threshold the non-rival share falls on, unless the truth lies within roughly a tenth of the boundary itself.

7.3. Data requirements

Four ingredients are needed, in descending order of how often real data supply them: (i) market or time variation in searcher density spanning tight to slack markets; (ii) referral offers or leads received, not only realized hires (with hires only, a composition correction mapping multi-offer distributions to at-most-one hire is required, and I flag it as the first modeling task of any empirical implementation); (iii) searchers' contact counts by the contacts' employment status; and (iv) the same composition for the *employed* (the queue side). Ingredients (i)–(ii) exist in job-search surveys; (iii)–(iv) are present in network modules and population registers; no readily available source has all four. The practical sequencing this implies is an ambitious data project in its own right: first a rivalry test on survey data, then the full split where two-sided contact composition exists. What this section establishes is that the object is identified, the estimator exists, and the precision on offer would suffice to settle the policy question if the data could be assembled.

8. Conclusion

What should a policymaker take from a model of rival referrals? Before subsidizing search, measure the queue. A single statistic — the share of delivered referrals whose recipient was the referrer's only usable candidate — says how much of the search-effort margin creates matches rather than reallocating them, falls short of one by the elasticity at which searchers crowd one another's referral chances, and prices the queue-relief externality that makes connections across sectors socially valuable. At the baseline calibration it sits near 0.80 in normal times and 0.43 at the trough of a severe recession, and the sign of the optimal effort instrument turns on it. The instrument is approximately zero in normal times under pure rivalry, a substantial tax in slumps, a subsidy only if enough referral information is non-rival. That last condition is measurable. The identification design of Section 7 recovers the non-rival share to within ± 0.04, an order of magnitude finer than the thresholds the policy question needs, given referral data with two-sided contact composition. Assembling such data is likely the more binding constraint and is a natural application of this theory.

The mechanism has the most bite where a shock lands on many people who search the same networks at once — concentrated layoffs, a region's long decline, an occupation thinned by automation — because that is when the queue behind each referrer is longest and de-correlating bridges are scarcest. Those are the episodes in which knowing the sole-candidate share, rather than assuming it, changes what a government should do.

For quantitative models the takeaways are two. Convex network effects in matching need not be assumed: one-lead-one-referral produces them, with the convexity pinned by an observable contact scale and disciplined by a companion saturation prediction. And the mean-field laws that make such models tractable are trustworthy. They are exact in the large-population limit by Theorem 3.1, accurate on fixed networks to within roughly five to twenty percent in the conservative direction, with a closure error that is not slack but the model's own "who-scars" prediction.

The paper has important boundaries. The welfare analysis is static, built to isolate the creation-versus-appropriation margin; embedding the sign threshold in a full dynamic setting with endogenous effort is left to future work. It is also partial equilibrium in the vacancy dimension: referral flows do not feed back into vacancy creation or wages, and the match-value wedge is a free parameter rather than solved for, so how firms respond to a change in referral efficiency sits outside the model. The lead flow α_R is also held fixed over the cycle. If referrers solicit or ration leads cyclically, the magnitudes above move with them; the sign threshold does not, because s_{piv} is a property of the queue rather than of the lead flow. The stark verdicts (a tax in slumps, a knife-edge in normal times) are the pure-rivalry ones; the blended rule of Section 5 governs once a share of referral information is non-rival. The referral here is pure allocation, carrying no signal about match quality, so any interaction of rivalry with the endorsement content of referrals is left open. The endorsement logic points toward reinforcement, since vouching is rationed hardest exactly when applicant pools are noisiest. The policy magnitudes are illustrations at one calibration, not estimates. The fixed-network results cover sector-structured configuration graphs, not hub-dominated degree distributions. And μ_{NR} remains, for now, unmeasured — the paper supplies the estimand, the estimator, and the data specification, not the number.

The literature's starting point is that a worker's network is her asset. This paper adds that in bad times the same network is also a queue, and that much of labor-market policy comes down to whether a subsidy creates matches or merely reorders the queue.

References

Acemoglu, D. and R. Shimer (1999). Holdups and efficiency with search frictions. *International Economic Review 40*(4), 827–849.

Arbex, M., D. O'Dea, and D. Wiczer (2019). Network search: Climbing the job ladder faster. *International Economic Review 60*(2), 693–720.

Beaman, L. and J. Magruder (2012). Who gets the job referral? evidence from a social networks experiment. *American Economic Review 102*(7), 3574–3593.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the U.S. *Review of Economic Studies 79*(1), 128–161.

Benaïm, M. and J.-Y. Le Boudec (2008). A class of mean field interaction models for computer and communication systems. *Performance Evaluation 65*(11–12), 823–838.

Blanchard, O. J. and P. Diamond (1994). Ranking, unemployment duration, and wages. *Review of Economic Studies 61*(3), 417–434.

Bramoullé, Y. and G. Saint-Paul (2010). Social networks and labor market transitions. *Labour Economics 17*(1), 188–195.

Brown, M., E. Setren, and G. Topa (2016). Do informal referrals lead to better matches? evidence from a firm's employee referral system. *Journal of Labor Economics 34*(1), 161–209.

Burks, S. V., B. Cowgill, M. Hoffman, and M. Housman (2015). The value of hiring through employee referrals. *Quarterly Journal of Economics 130*(2), 805–839.

Cahuc, P. and F. Fontaine (2009). On the efficiency of job search with social networks. *Journal of Public Economic Theory 11*(3), 411–439.

Cai, X., P. Gautier, and R. Wolthoff (2017). Search frictions, competing mechanisms and optimal market segmentation. *Journal of Economic Theory 169*, 453–473.

Calvó-Armengol, A. and M. O. Jackson (2004). The effects of social networks on employment and inequality. *American Economic Review 94*(3), 426–454.

Calvó-Armengol, A. and M. O. Jackson (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory 132*(1), 27–46.

Calvó-Armengol, A. and Y. Zenou (2005). Job matching, social network and word-of-mouth communication. *Journal of Urban Economics 57*(3), 500–522.

Carmona, R. and F. Delarue (2018). *Probabilistic Theory of Mean Field Games with Applications I–II*. Probability Theory and Stochastic Modelling. Springer.

Chaintron, L.-P. and A. Diez (2022). Propagation of chaos: A review of models, methods and applications. I. Models and methods. *Kinetic and Related Models 15*(6), 895–1015.

Cheung, M., J. Egebark, A. Forslund, L. Laun, M. Rödin, and J. Vikström (2025). Does job search assistance reduce unemployment? Evidence on displacement effects and mechanisms. *Journal of Labor Economics 43*(1), 47–81.

Cingano, F. and A. Rosolia (2012). People i know: Job search and social networks. *Journal of Labor Economics 30*(2), 291–332.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Quarterly Journal of Economics 128*(2), 531–580.

Diamond, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy 90*(5), 881–894.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016). Referral-based job search networks. *Review of Economic Studies 83*(2), 514–546.

Eeckhout, J. and P. Kircher (2010). Sorting versus screening: Search frictions and competing mechanisms. *Journal of Economic Theory 145*(4), 1354–1385.

Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2022). Job search behavior among the employed and non-employed. *Econometrica 90*(4), 1743–1779.

Galenianos, M. (2014). Hiring through referrals. *Journal of Economic Theory 152*, 304–323.

Galeotti, A. and L. P. Merlino (2014). Endogenous job contact networks. *International Economic Review 55*(4), 1201–1226.

Gautier, P. A., P. Muller, B. van der Klaauw, M. Rosholm, and M. Svarer (2018). Estimating equilibrium effects of job search assistance. *Journal of Labor Economics 36*(4), 1073–1125.

Glitz, A. (2017). Coworker networks in the labour market. *Labour Economics 44*, 218–230.

Granovetter, M. (1995). *Getting a Job: A Study of Contacts and Careers* (2nd ed.). Chicago: University of Chicago Press. First edition: Harvard University Press, 1974.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Heath, R. (2018). Why do firms hire using referrals? evidence from Bangladeshi garment factories. *Journal of Political Economy 126*(4), 1691–1746.

Hosios, A. J. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies 57*(2), 279–298.

Ioannides, Y. M. and L. Datcher Loury (2004). Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature 42*(4), 1056–1093.

Kopecky, J. (2026). Friends in low places: Social networks, the friendship paradox, and structural scarring. Working paper, in preparation.

Kramarz, F. and O. N. Skans (2014). When strong ties are strong: Networks and youth labour market entry. *Review of Economic Studies 81*(3), 1164–1200.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability 7*(1), 49–58.

Landais, C., P. Michaillat, and E. Saez (2018a). A macroeconomic approach to optimal unemployment insurance: Applications. *American Economic Journal: Economic Policy 10*(2), 182–216.

Landais, C., P. Michaillat, and E. Saez (2018b). A macroeconomic approach to optimal unemployment insurance: Theory. *American Economic Journal: Economic Policy 10*(2), 152–181.

Lasry, J.-M. and P.-L. Lions (2007). Mean field games. *Japanese Journal of Mathematics 2*(1), 229–260.

Lester, B., D. A. Rivers, and G. Topa (2021). The heterogeneous impact of referrals on labor market outcomes. Staff Report 987, Federal Reserve Bank of New York. Revised November 2023. Forthcoming, Review of Economic Studies.

Lester, B., L. Visschers, and R. Wolthoff (2015). Meeting technologies and optimal trading mechanisms in competitive search markets. *Journal of Economic Theory 155*, 1–15.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. *RAND Journal of Economics 17*(1), 48–58.

Michaillat, P. (2012). Do matching frictions explain unemployment? Not in bad times. *American Economic Review 102*(4), 1721–1750.

Moen, E. R. (1997). Competitive search equilibrium. *Journal of Political Economy 105*(2), 385–411.

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review 81*(5), 1408–1418.

Mortensen, D. T. (1982). Property rights and efficiency in mating, racing, and related games. *American Economic Review 72*(5), 968–979.

Pallais, A. and E. G. Sands (2016). Why the referential treatment? evidence from field experiments on referrals. *Journal of Political Economy 124*(6), 1793–1828.

Petrongolo, B. and C. A. Pissarides (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature 39*(2), 390–431.

Pissarides, C. A. (2000). *Equilibrium Unemployment Theory* (2nd ed.). Cambridge, MA: MIT Press.

Schmutte, I. M. (2016). Labor markets with endogenous job referral networks: Theory and empirical evidence. *Labour Economics 42*, 30–42.

Stupnytska, Y. and A. Zaharieva (2017). Optimal policy and the role of social contacts in a search model with heterogeneous workers. *Journal of Public Economic Theory 19*(5), 957–985.

Sznitman, A.-S. (1991). Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*, Volume 1464 of *Lecture Notes in Mathematics*, pp. 165–251. Springer.

Topa, G. (2011). Labor markets and referrals. In *Handbook of Social Economics*, Volume 1B, pp. 1193–1221. North-Holland.

Wahba, J. and Y. Zenou (2005). Density, social networks and job search methods: Theory and application to Egypt. *Journal of Development Economics 78*(2), 443–473.

Zaharieva, A. (2013). Social welfare and wage inequality in search equilibrium with personal contacts. *Labour Economics 23*, 107–121.

Zaharieva, A. (2018). On the optimal diversification of social networks in frictional labour markets with occupational mismatch. *Labour Economics 50*, 112–127.

Zenou, Y. (2015). A dynamic model of weak and strong ties in the labor market. *Journal of Labor Economics 33*(4), 891–932.

A. Model Algebra and Calibration

A.1 The decay bracket as a change of variables

Reduced-form treatments of network capital — the companion paper (Kopecky, 2026) among them — write the unemployed's tie decay as $\delta[1+\kappa(1-\tilde{e}/\bar{e}^{\circ})]$: a baseline rate amplified when network employment falls below its steady state, with κ governing the amplification. Under annealed matching, the model of Section 2 delivers exactly this form. An unemployed worker's per-tie decay rate is

$$d_{eu}\,\tilde{e}_w + d_{uu}(1-\tilde{e}_w) \;=\; d_{eu}\big[1+\tilde{\kappa}\,(1-\tilde{e}_w)\big], \qquad \tilde{\kappa} = \frac{d_{uu}-d_{eu}}{d_{eu}},$$

and matching level and slope against the reduced form at the reference employment rate $\bar{e}^{\circ}$ gives the affine system

$$\delta(1+\kappa) = d_{eu}(1+\tilde{\kappa}), \qquad \delta\kappa/\bar{e}^{\circ} = d_{eu}\tilde{\kappa}$$

— two equations in (d_{eu}, d_{uu}), exact for every value of $\tilde{e}_w$ (the two forms agree to machine precision over the relevant range; this is verified in the replication ledger). Inverting at the calibration below yields $d_{eu} = 0.112$ and $d_{uu} = 0.240$ per quarter. The reduced form's amplification parameter is therefore not internal to the model: κ *is* the u–u versus e–u tie survival gap, an object estimable from any panel that records relationships and both parties' employment status.

A.2 Calibration

All magnitudes in the text use one quarterly calibration (Table 5), disciplined to standard values and held fixed across every experiment.

Table 5: *Calibration.*

parameter	value	source of discipline
separation rate s	0.035	steady-state $u^* = 6\%$ with $f^* = 0.548$
referral share of hires χ	0.40	survey range one third to one half
direct finding f_{dir}	$(1-\chi)f^* = 0.329$	residual
steady-state tie stock K°	0.70	normalization
decay scale δ	0.12	unemployed-tie half-life ≈ 5.8 quarters
amplification κ	1	maps to $(d_{eu}, d_{uu}) = (0.112, 0.240)$
formation ϕ	0.0179	backed out of the K steady state
contacts at capacity c	10	baseline; swept 5–20 in the text
referral slopes α, α_R	0.333, 0.074	referral flow $= \chi f^*$ at steady state

Rates are per quarter. The amplification κ is the reduced-form parameter of A.1, distinct from the coupling floor κ_0 of Appendix B.

The referral slopes are calibrated so that both protocols deliver the same steady state: total referral finding equals χf^* at $(\bar{e}, K) = (0.94, 0.70)$ whether information travels by broadcast, by token, or by any blend. The two protocols convert a searcher's leads into a hire as their micro-structure dictates: broadcast enters the finding rate linearly, as in (3), while a token searcher is hired if her direct search succeeds or she receives at least one usable token, giving $f = 1 - (1 - f_{\text{dir}})\, e^{-\tau}$ in her expected token count τ — the Poisson counterpart of the linear term. Both are pinned to $f^* = 0.548$ at the steady state, which fixes $\alpha_R = 0.074$. Protocol comparisons in the text are therefore comparisons of mechanisms, never of steady states.

B. Proof of Theorem 3.1

This appendix proves Theorem 3.1. Both this appendix and Appendix C proceed through explicit one-step inequalities, and every inequality is checked numerically on simulated paths via a literal implementation of the coupled construction; the measured convergence exponents, -0.49 and -0.51, sit on the theorem's $-1/2$. The scheme is the synchronous coupling of Sznitman (1991), in discrete time, where it is elementary; see Chaintron and Diez (2022) for the modern toolkit, Kurtz (1970) for the density-dependent ancestry, and Benaïm and Le Boudec (2008) for discrete-time mean-field interaction models of this flavor. The one model-specific step is Lemma 1, already stated in the text. Throughout, $h(n, m) = \min(f_{\text{dir}} + \alpha n m,\ \bar{h})$ with cap $\bar{h} < 1$, $d(m) = d_{eu} m + d_{uu}(1 - m)$, and one period of the process at flow m maps (z, n) with a uniform U to

$$z' = \begin{cases} \mathbf{1}\{U > s_t\} & z = 1 \\ \mathbf{1}\{U < h(n, m)\} & z = 0 \end{cases} \qquad n' = \begin{cases} n + \phi(1-n) & z = 1 \\ n\,(1 - d(m)) & z = 0, \end{cases}$$

where $(s_t)_{t \le T}$ is any deterministic separation path (the shock experiments of the text are covered). The N-system uses $m_t^N = \tilde{e}_w(\mu_t^N)$, the empirical tie-weighted employment; the McKean–Vlasov (MV) limit uses the deterministic flow $m_t = \tilde{e}_w(\mu_t)$, $\mu_{t+1} = \Phi_t(\mu_t)$, which in discrete time exists and is unique by construction.

Standing assumption. The initial law is supported on $\{n \ge n_0\}$ for some $n_0 > 0$, and the N-system starts i.i.d. from it.

B.1 The invariant domain (novel step)

Proof of Lemma 1. In each period either $n' = n + \phi(1-n) \ge n$, or $n' = n(1 - d(m)) \ge n(1 - d_{uu})$, since $d(m) \le d_{uu}$ for $m \in [0, 1]$. Induction gives $n_i^t \ge n_0(1 - d_{uu})^t$ pathwise, in the N-system, the MV system, and any coupled copy; both branches also preserve $n \le 1$. Averaging, every denominator entering $\tilde{e}_w$ is bounded below by $\kappa_0 := n_0(1 - d_{uu})^T$ on the horizon, almost surely. □

The force of the lemma is that the bound is pathwise and holds with no concentration argument, which is what lets the ratio functional be treated as Lipschitz everywhere the coupling ever evaluates it.

B.2 Lipschitz estimates (standard)

Lemma 2. *For $n, n', m, m' \in [0, 1]$: (i) $|h(n, m) - h(n', m')| \le \alpha|n - n'| + \alpha|m - m'|$; (ii) $|d(m) - d(m')| = (d_{uu} - d_{eu})\,|m - m'|$; (iii) if μ, ν are empirical measures of two configurations on the same index set with $\langle n \rangle_\mu, \langle n \rangle_\nu \ge \kappa_0$, then*

$$\left|\tilde{e}_w(\mu) - \tilde{e}_w(\nu)\right| \;\le\; \frac{1}{\kappa_0}\frac{1}{N}\sum_i \Big(2\,|n_i - n_i'| + \mathbf{1}\{z_i \ne z_i'\}\Big).$$

Proof. (i) $f_{\mathrm{dir}} + \alpha n m$ is α-Lipschitz in each argument on $[0, 1]^2$ and $\min(\cdot, \bar{h})$ is 1-Lipschitz. (ii) is affine. (iii) Write $A = \langle nz \rangle$, $B = \langle n \rangle$. Then $|A/B - A'/B'| \le |A - A'|/B + (A'/B')|B - B'|/B \le \kappa_0^{-1}(|A - A'| + |B - B'|)$ using $A' \le B'$ and $B \ge \kappa_0$; and $|A - A'| \le N^{-1}\sum_i(|n_i - n_i'| + \mathbf{1}\{z_i \ne z_i'\})$ (add and subtract $n_i' z_i$; $n \le 1$), while $|B - B'| \le N^{-1}\sum_i |n_i - n_i'|$. □

B.3 The synchronous coupling (standard construction, two-channel bookkeeping)

On one probability space carry the N-system (z_i^t, n_i^t) and the *auxiliary system* $(\bar{z}_i^t, \bar{n}_i^t)$: N independent copies of the MV worker, driven by the deterministic flow m_t and — the coupling — by the *same* initial draws and the *same* uniforms U_i^t as their twins. The auxiliary particles are i.i.d. with law μ_t by construction. Define

$$\delta_t = \mathbb{E}\,|n_1^t - \bar{n}_1^t|, \qquad \varepsilon_t = \mathbb{P}\,(z_1^t \neq \bar{z}_1^t), \qquad \eta_t = \mathbb{E}\,|m_t^N - m_t|,$$

the same for every i by exchangeability; $\delta_0 = \varepsilon_0 = 0$. Set $c_n = \phi + d_{uu}$, $c_d = d_{uu} - d_{eu}$, $c_h = \alpha$, $L = 1/\kappa_0$.

Lemma 3 (One-step recursions)**.** *For all $t < T$:*

$$\begin{aligned}
\delta_{t+1} &\le \delta_t + c_n\,\varepsilon_t + c_d\,\eta_t, \\
\varepsilon_{t+1} &\le \varepsilon_t + c_h\,(\delta_t + \eta_t), \\
\eta_t &\le L\,(2\delta_t + \varepsilon_t) + r_t^N, \qquad r_t^N := \mathbb{E}\,\big|\tilde{e}_w(\bar{\mu}_t^N) - m_t\big|,
\end{aligned}$$

where $\bar{\mu}_t^N$ is the auxiliary empirical measure.

Proof. Ties. On $\{z = \bar{z} = 1\}$: $|n' - \bar{n}'| = (1-\phi)|n - \bar{n}|$. On $\{z = \bar{z} = 0\}$: $|n' - \bar{n}'| \le (1 - d(m_t))|n - \bar{n}| + n\,|d(m_t^N) - d(m_t)| \le |n - \bar{n}| + c_d\,|m_t^N - m_t|$. On $\{z \neq \bar{z}\}$ each branch moves n by at most $\max(\phi, d_{uu})$, so $|n' - \bar{n}'| \le |n - \bar{n}| + (\phi + d_{uu})$. Take expectations.

Status. $\mathbb{P}(z' \neq \bar{z}') \le \mathbb{P}(z \neq \bar{z}) + \mathbb{P}(z' \neq \bar{z}', z = \bar{z})$. On $\{z = \bar{z} = 1\}$ the same uniform and the same s_t force $z' = \bar{z}'$. On $\{z = \bar{z} = 0\}$, $z' \neq \bar{z}'$ exactly when U_1^t lands between $h(n, m_t^N)$ and $h(\bar{n}, m_t)$; since U_1^t is independent of time-t states (including m_t^N), the conditional probability is $|h(n, m_t^N) - h(\bar{n}, m_t)| \le \alpha(|n - \bar{n}| + |m_t^N - m_t|)$ by Lemma 2(i). Take expectations.

Flow. $m_t^N - m_t = [\tilde{e}_w(\mu_t^N) - \tilde{e}_w(\bar{\mu}_t^N)] + [\tilde{e}_w(\bar{\mu}_t^N) - m_t]$. Both empirical measures have denominators $\ge \kappa_0$ almost surely (Lemma 1), so Lemma 2(iii) bounds the first term's expectation by $L(2\delta_t + \varepsilon_t)$; the second is r_t^N. □

B.4 The auxiliary law of large numbers (standard)

Lemma 4. $r_t^N \le 1/(\kappa_0\sqrt{N})$ *for every $t \le T$.*

Proof. $A_N = N^{-1}\sum_i \bar{n}_i\bar{z}_i$ and $B_N = N^{-1}\sum_i \bar{n}_i$ are means of i.i.d. $[0,1]$ variables with means $a_t \le b_t$ and $b_t \ge \kappa_0$; also $B_N \ge \kappa_0$ almost surely (Lemma 1). As in Lemma 2(iii),

$|A_N/B_N - a_t/b_t| \le \kappa_0^{-1}(|A_N - a_t| + |B_N - b_t|)$, and each L^1-deviation is at most its standard deviation, at most $1/(2\sqrt{N})$ for $[0,1]$ variables. □

B.5 Gronwall and the proof of Theorem 3.1

Let $e_t = \delta_t + \varepsilon_t$. Substituting the flow bound into the first two recursions of Lemma 3 and using $2\delta + \varepsilon \le 2e$,

$$e_{t+1} \;\le\; (1+\gamma)\, e_t + \frac{\beta}{\kappa_0\sqrt{N}}, \qquad \gamma = \max(c_n, c_h) + 2L(c_d + c_h), \quad \beta = c_d + c_h,$$

and discrete Gronwall from $e_0 = 0$ gives $e_t \le \big[\beta/(\gamma\kappa_0)\big]\big[(1+\gamma)^t - 1\big] N^{-1/2}$ for $t \le T$. The flow bound then delivers part (i) of the theorem with

$$C_1(T) \;=\; \frac{2\beta}{\gamma\kappa_0^2}\,(1+\gamma)^T + \frac{1}{\kappa_0}.$$

For part (ii), any k coupled workers sit within $\sum_{i\le k}\big(|n_i - \bar{n}_i| + \mathbf{1}\{z_i \ne \bar{z}_i\}\big)$ of the i.i.d. vector $(\bar{X}_1, \ldots, \bar{X}_k)$ whose law is exactly $\mu_t^{\otimes k}$; for any F that is 1-Lipschitz in the summed metric, $|\mathbb{E}F(X) - \mathbb{E}F(\bar{X})| \le k\, e_t \le k\, C_2(T)\, N^{-1/2}$ with $C_2(T) = [\beta/(\gamma\kappa_0)](1+\gamma)^T$. □

B.6 Sectors, and two qualifications

With S sectors of fixed shares $\rho_j > 0$ and a row-stochastic mixing matrix M, the flow is the vector $m_j = \sum_l M_{jl}\tilde{e}_w^{(l)}$. Lemma 1 holds per sector with the same κ_0; row-stochasticity gives $|m_j - m_j'| \le \max_l |\tilde{e}_w^{(l)} - \tilde{e}_w^{(l)\prime}|$, so Lemma 3 goes through with the same constants, and Lemma 4 with $\kappa_0\sqrt{\rho_{\min} N}$. Theorem 3.1 holds verbatim with an extra factor $\rho_{\min}^{-1/2}$ in the constants.

Two remarks. First, $\kappa_0 = n_0(1 - d_{uu})^T$ is rigorous but pessimistic in T — the worst case is a worker unemployed for the entire horizon. A sharper, horizon-free floor is available by a standard good-event argument (employment concentrates above $f_{\mathrm{dir}}/(s + f_{\mathrm{dir}})$, which feeds the formation term); I use the simple version because the theorem's content is the rate in N, not the constant. Second, the constants at the calibration: $c_n = 0.258$, $c_d = 0.128$, $c_h = 1/3$; with the *measured* floor ($\kappa_0 \approx 0.59$ across every simulated path), $\gamma \approx 1.9$. The replication package implements the coupling of Section B.3 literally — same uniforms, reference flow from an independent two-million-worker simulation — and verifies each inequality of Lemmas 1–4 period by period at $N \in \{2{,}000,\ 8{,}000,\ 32{,}000\}$, on constant and shock separation paths, together with the endpoint rate (measured exponents -0.49 for the coupled distance and -0.51 for the flow error).

B.7 The token protocol and blends

Sections B.1–B.6 couple the broadcast process, where workers interact only through $\tilde{e}_w$. Under the token protocol the searcher's referral hazard is

$$\tau_i \;=\; \alpha_R\, n_i\, G(\mu), \qquad G(\mu) \;=\; \frac{\left\langle z\left(1-e^{-cn(1-\tilde{e}_w(\mu))}\right)\right\rangle_\mu}{\big\langle (1-z)\, n\big\rangle_\mu},$$

structurally the broadcast hazard $\alpha n \tilde{e}_w$ with the interaction scalar $\tilde{e}_w$ replaced by the delivered-yield functional G. The theorem extends once G, like $\tilde{e}_w$, is a bounded Lipschitz functional of the empirical measure on the invariant domain, and the argument splits cleanly at the two parts of the statement.

Part (i): the aggregates. The reduced form is driven by two functionals only — the decay flow through $\tilde{e}_w$, and the delivered referral flow $T = \alpha_R \langle z(1-e^{-x})\rangle$, which is the numerator of G scaled by α_R. Both are denominator-free and bounded in $[0,1]$: $1-e^{-x}$ is 1-Lipschitz in $x = cn(1-\tilde{e}_w)$, which is Lipschitz in n and in $\tilde{e}_w$ by Lemma 2, and the average is over a bounded integrand. So the interaction vector $(\tilde{e}_w, T)$ satisfies the estimates of Lemmas 2–3 with the same κ_0 and enlarged constants γ, β, and the convergence of the aggregates at rate $N^{-1/2}$ (part (i)) holds verbatim. The blend μ_{NR} mixes the two hazards linearly and is covered by the same estimate; the rate is unchanged and only γ grows.

Part (ii): the individual allocation. The one point of care is the denominator of G, the unemployed tie mass $\langle (1-z)n\rangle$, which enters the individual searcher's hazard τ_i. Unlike $\langle n\rangle$ it has no pathwise floor — a population could momentarily hold no searchers — but it concentrates around its deterministic value $(1-\bar{e})\,\mathbb{E}[n \mid u] > 0$ and drops below a floor $\frac{1}{2}(1-\bar{e})\kappa_0$ only on an event of probability exponentially small in N, by the good-event argument of the first remark of §B.6 (unemployment is a bounded-increment functional with a strictly interior mean). On that event G is Lipschitz with constant $O\big(1/((1-\bar{e})\kappa_0)\big)$ and the coupling of §B.5 closes; the complementary event contributes below the $N^{-1/2}$ rate. Equivalently, capping the referral finding probability at $\bar{h}$, as the direct hazard is capped, removes the denominator from the individual hazard outright. The replication ledger runs its coupling checks in the token and blended economies as well as the broadcast one. The theorem thus covers the token protocol through its semi-analytic delivered-yield hazard; the discrete one-recipient-per-lead allocation of Section 2.2 reproduces that limit to within 0.2% in the ledger (Appendix F), the sense in which the rival mean-field limit holds.

C. Proof of Theorem 3.2

This appendix proves Theorem 3.2 and derives the constant C_∞. At a stationary law of the MV dynamics the flow m is a constant, so the single worker follows an ordinary time-homogeneous Markov chain on $\{0,1\} \times [0,1]$; the proof is stationary-chain analysis plus a self-consistency step for m itself. Throughout, the tie dynamics carry the scaling λ: employed $n' = n + \lambda\phi(1-n)$, unemployed $n' = n(1-\lambda d(m))$.

C.1 Standing conditions

Assumption 1. (i) No clip: $f_{\mathrm{dir}} + \alpha \le \bar{h}$, and $\bar{\rho} := \max(1-s-f_{\mathrm{dir}},\ |1-s-\bar{h}|) < 1$. (ii) The $\lambda \to 0$ mean system (the aggregate map (3)–(4)) has a nonsingular Jacobian at the calibrated fixed point. (iii) Loop-gain smallness: $\beta(1-\bar{e})G/\psi < s+f$ and $\pi' G < 4\psi$, where $\beta = \alpha\bar{e}$, $G = \phi(1-K^\circ) + \delta K^\circ$, $\psi = \phi\bar{e} + \delta(1-\bar{e})$, and $\pi' = \alpha\bar{e}\, s/(s+f)^2$, all at the calibrated point.

All three are verified numerically at the calibration with wide margins: $\det(I-J) = 0.012$; the two loop gains are 0.12 and 0.03 against bounds of one. Existence of a stationary law follows from Schauder–Tychonoff on the compact convex set of laws (the relevant set is invariant by Lemma 5 below); uniqueness is not needed — the theorem holds at *every* stationary law.

C.2 Stationary floors are automatic

Lemma 5. *Any stationary law satisfies $\langle z\rangle \ge e_\star := f_{\mathrm{dir}}/(s+f_{\mathrm{dir}})$ and $\langle n\rangle \ge \phi\, e_\star/d_{uu}$, for every $\lambda > 0$.*

Proof. Stationarity of $\langle z\rangle$: $s\langle z\rangle = \langle(1-z)h(n,m)\rangle \ge (1-\langle z\rangle)f_{\mathrm{dir}}$. Stationarity of $\langle n\rangle$ (the λ's cancel): $\phi\langle z(1-n)\rangle = d(m)\langle(1-z)n\rangle$, i.e. $\langle n\rangle = [\phi\langle z\rangle + (d(m)-\phi)\langle zn\rangle]/d(m) \ge \phi\langle z\rangle/d_{uu}$ when $d(m) \ge \phi$ (true at the calibration: $d_{eu} = 0.112 > \phi = 0.018$). □

Note the contrast with Appendix B: the floors here are λ-free and horizon-free, so no pessimistic constant enters this part anywhere.

C.3 The binary-chain decorrelation lemma

Write $\pi(n) = h(n,m)/(s+h(n,m))$ and $\rho(n) = 1-s-h(n,m)$, so that $|\rho(n)| \le \bar{\rho} < 1$ and the exact one-step identity

$$\mathbb{E}\left[z_{t+1} \mid \mathcal{F}_t\right] \;=\; \pi(n_t) + \rho(n_t)\,\tilde{\zeta}_t, \qquad \tilde{\zeta}_t := z_t - \pi(n_t),$$

holds. The following lemma is the rigorous replacement for every "z is fast, so its covariance with slow variables is small" step.

Lemma 6. *Let $S_t = f(n_t)$ be any stationary sequence with $|S| \le M_S$. Then, at any stationary law,*

$$\big|\mathbb{E}[\tilde{\zeta}_t S_t]\big| \;\le\; \frac{\alpha\,\mathbb{E}|\xi S| + \mathbb{E}|S_t - S_{t-1}| + L_\pi\,\mathbb{E}|\Delta n|\,M_S}{1-\bar{\rho}},$$

where $\xi = n - n^\circ$ (deviation from the calibrated point), $L_\pi = \sup|\pi'|$, and $|\Delta n| \le \lambda(\phi \vee d_{uu})$ pathwise.

Proof. The one observation doing the work: ties update deterministically from time-t information, so n_{t+1} — and hence S_{t+1} — is $\mathcal{F}_t$-measurable, and the martingale part of z_{t+1} is orthogonal to it. Therefore

$$\mathbb{E}[\tilde{\zeta}_{t+1}S_{t+1}] = \mathbb{E}\big[\rho(n_t)\tilde{\zeta}_t S_t\big] + \mathbb{E}\big[\rho(n_t)\tilde{\zeta}_t(S_{t+1}-S_t)\big] + \mathbb{E}\big[(\pi(n_t)-\pi(n_{t+1}))S_{t+1}\big].$$

Write $\rho(n_t) = \rho_\star + (\rho(n_t) - \rho_\star)$ with $\rho_\star = \rho(n^\circ)$, $|\rho_\star| \le \bar{\rho}$ and $|\rho(n) - \rho_\star| \le \alpha|\xi|$ (h is αm-Lipschitz in n, $m \le 1$). By stationarity the left side equals $\mathbb{E}[\tilde{\zeta}_t S_t]$; bound $|\tilde{\zeta}| \le 1$, collect, and solve for $|\mathbb{E}[\tilde{\zeta} S]|$. □

The hazard feedback — tie-rich searchers exit faster — is carried by $\pi(n)$ varying with n and appears through the $\alpha\mathbb{E}|\xi S|$ term; it is a first-order piece of the constant below, not a refinement.

C.4 Moment scalings

Lemma 7. *Under Assumption 1, at any stationary law, with ξ_c the central deviation of n: (i) the means $(\bar{e}, \langle n\rangle, m)$ lie within $O(\lambda)$ of the calibrated point; (ii) $\mathrm{Var}(n) = O(\lambda)$; (iii) $\mathbb{E}[\xi_c^4] = O(\lambda^2)$ and hence $\mathbb{E}|\xi_c|^3 = O(\lambda^{3/2})$; (iv) $|\mathrm{Cov}(z, \xi_c^2)| = O(\lambda^{3/2})$.*

Proof sketch, with the two critical steps. (i) The three exact mean-stationarity identities form a perturbation of the $\lambda \to 0$ mean system with right side $O(|C| + V + \lambda)$; Assumption 1(ii) and the implicit function theorem give the claim once (ii) is in hand. (ii) Lemma 6 with $S = \xi$ bounds $|\mathbb{E}[\tilde{\zeta}\xi]|$ by $O(\mathbb{E}[\xi^2] + \lambda)$; the exact stationarity identity for $\mathbb{E}[\xi^2]$ reads $2\lambda\psi(1+O(\lambda))V = 2\lambda GC + \lambda^2\mathbb{E}[\tilde{g}^2] + \lambda\,O(|C| + V + |\mathbb{E}\xi|)$, where $\tilde{g}$ is the centered tie drive with gap G; substituting each bound into the other closes by Assumption 1(iii) (measured loop gain 0.12). (iii) is the step where naive bounds fail. The exact identity for $\mathbb{E}[\xi_c^4]$ has cross term $4\lambda\mathbb{E}[\xi_c^3\,\tilde{g}(z)]$; decompose $\tilde{g}(z) = \bar{g}(n) + G\tilde{\zeta}$ with $\bar{g}(n) = \mathbb{E}[\tilde{g} \mid n,\ \text{frozen}]$, note $\bar{g}(n) = \pi' G\,\xi + O(\xi^2 + \lambda)$ by the calibration, and bound $\mathbb{E}[\xi_c^3\tilde{\zeta}]$ by Lemma 6 with

$S = \xi_c^3$. Collecting, $(4\psi - \pi' G - O(\lambda))\mathbb{E}[\xi_c^4] \le c(\lambda \mathbb{E}|\xi_c|^3 + \lambda V + \lambda^2)$ — positive coefficient by Assumption 1(iii) — and with $\mathbb{E}|\xi_c|^3 \le (V\,\mathbb{E}[\xi_c^4])^{1/2}$ this is a quadratic inequality in $(\mathbb{E}[\xi_c^4])^{1/2}$ whose solution is $O(\lambda^2)$. Freezing the tie-drive gap at the calibrated G is harmless at this order: its state-dependent residual multiplies the martingale increment $\tilde{\zeta}$, so by Lemma 6 it decorrelates from the slow variables into the $\lambda\, O(V)$ and higher-order buckets of C.5 and does not enter the coercive coefficient $4\psi - \pi' G$; the measured λ^2 scaling confirms it. An absolute-value bound at the cross term would deliver only $O(\lambda)$ here; the signed decomposition and the smallness of $\pi' G$ rescue the correct order. (iv) Lemma 6 with $S = \xi_c^2$: every term on the right is $O(\lambda^{3/2})$ by (iii) and $\mathbb{E}|\Delta(\xi^2)| \le 2\lambda(\phi \vee d_{uu})\mathbb{E}|\xi| + O(\lambda^2)$; add $|\mathrm{Cov}(\pi(n), \xi_c^2)| \le L_\pi \mathbb{E}|\xi_c|^3$. □

C.5 The remainder ledger and the proof of Theorem 3.2

The closed form is obtained from two exact stationarity identities. Stationarity of $\mathbb{E}[\xi\zeta]$ and of $\mathbb{E}[\xi^2]$ (with $\zeta = z - \bar{e}$ and $\gamma = 1 - s - f$) give, exactly,

$$\begin{aligned} C\,(1-\gamma) &= \lambda\gamma G\,\bar{e}(1-\bar{e}) + \beta(1-\bar{e})\,V \;+\; R_C, \\ 2\lambda\psi\,V &= 2\lambda G\,C + \lambda^2\mathbb{E}[\tilde{g}^2] \;+\; R_V, \end{aligned}$$

where every term collected in R_C, R_V is one of: $\lambda \cdot$ (linear in $C, \mathbb{E}\xi$) [$O(\lambda^2)$ by Lemma 7(i)–(ii), using that z is binary so $\zeta^2 = (1-2\bar{e})\zeta + \bar{e}(1-\bar{e})$ closes the hierarchy]; $\beta \cdot \mathrm{Cov}(z, \xi_c^2)$ [$O(\lambda^{3/2})$ by (iv) — the binding term]; mean-shift corrections to the coefficients [$O(\lambda^2)$ by (i)]; and $\lambda \cdot O(V)$ cross terms [$O(\lambda^2)$]. Solving the two-by-two linear system — invertible with margin by Assumption 1(iii) — yields

$$C \;=\; \lambda\, C_\infty + O(\lambda^{3/2}), \qquad C_\infty \;=\; \frac{\gamma G\bar{e}(1-\bar{e}) + \beta(1-\bar{e})\,\mathbb{E}[\tilde{g}^2]/(2\psi)}{(s+f) - \beta(1-\bar{e})G/\psi},$$

and the two displayed conclusions of the theorem follow from the exact conditional-moment identities $\mathbb{E}[n \mid u] - \mathbb{E}[n] = -C/(1-\bar{e})$ and $\tilde{e}_w - \bar{e} = C/\langle n\rangle$, with the $O(\lambda)$ mean shifts absorbed at $O(\lambda^2)$. □

C.6 Remarks

Where $\lambda^{3/2}$ comes from. The single term resisting $O(\lambda^2)$ is $\beta\,\mathrm{Cov}(z, \xi_c^2)$: Lemma 6 bounds it by its scale, while its true size is its first-order corrector value — and the simulations measure its λ-exponent at 1.9, i.e. λ^2 in practice. One further iteration of the Lemma 6 expansion applied to $S = \xi_c^2$ would sharpen the theorem to $O(\lambda^2)$; it is mechanical and I have not done it.

Numbers. At the calibration $C_\infty = 0.00443$; the measured remainder $C - \lambda C_\infty$ has fitted exponent 1.6 over $\lambda \in \{1, \frac{1}{2}, \frac{1}{4}\}$ with C/λ within 2.4%/1.9%/0.8% of C_∞, and the implied pool-selection gap and tilt match full-process simulations as reported in the text. Every scaling asserted in Lemma 7 is verified in the replication ledger (measured exponents 0.94, 1.41, 1.89, 1.92 for (ii)–(iv)), as are the floors, the frozen-chain identities, and the orthogonality step of Lemma 6.

Scope. Stationary laws only. The theorem governs the two-state closure of the reduced form, not the cyclical path of s_{piv}, which is cross-sectional (Proposition 2) and rides on Theorem 3.1 instead. The shocked-path version of the closure statement itself is numerically supported — Appendix B's machine checks run the coupling along shock paths — and holds analytically in the slowly-varying limit, but is not proven here for sharp transitions.

D. Rivalry Propositions: Proofs

D.1 The token hazards and Proposition 1

Under the token protocol at a cross-section $\{(z_i, n_i)\}$, an employed worker j holds a lead with probability α_R; with Poisson contact counts her number of usable unemployed contacts has mean $x_j = c\, n_j(1 - \tilde{e}_w)$, so she delivers with probability $P_j = 1 - e^{-x_j}$, and the aggregate delivered flow per capita is $T = \alpha_R\, N^{-1} \sum_{j:z_j=1} P_j$. Delivered leads are allocated across searchers in proportion to their usable tie mass, so searcher i's hazard is $Tn_i/(N^{-1} \sum_{k:z_k=0} n_k)$. (This semi-analytic system is validated against an explicit discrete token simulation in the replication ledger.) The elasticities of Proposition 1 are defined and computed on the calibrated stationary cross-section: ε_K as the elasticity of the mean searcher hazard to a uniform scaling of all tie stocks, and η_{eff} as the elasticity of incumbent searchers' hazards to the employment margin (a small random set of searchers employed), with respect to the induced change in $\tilde{e}_w$. The broadcast benchmark $\alpha n \tilde{e}_w$ has both elasticities equal to one identically. Under the token protocol neither equals one, and both are pinned to the sole-candidate share.

Saturation. A uniform scaling of every tie stock by θ leaves $\tilde{e}_w = \langle nz \rangle / \langle n \rangle$ unchanged and scales every usability $a_{jk} \mapsto \theta a_{jk}$ (equivalently $x_j \mapsto \theta x_j$). The mean searcher hazard equals $T/(1 - \bar{e})$, so it moves with the delivered flow $D = \sum_{j:z_j=1} P_j$, $P_j = 1 - \prod_k (1 - a_{jk})$, alone (the searcher tie mass scales with θ in both the numerator and the denominator of the allocation and cancels). By the product-form derivative established in D.2, $dD/d\log\theta =$

$\sum_j P_j(\text{exactly one usable})$, so

$$\varepsilon_K \;=\; \frac{d\log D}{d\log\theta}\bigg|_{\theta=1} \;=\; \frac{\sum_j P_j(\text{exactly one usable})}{\sum_j P_j(\text{at least one usable})} \;=\; s_{\text{piv}},$$

the sole-candidate share exactly: the tie-capital elasticity and the creation share of Proposition 2 are the same object. It is strictly below one whenever some referrer can have two usable contacts at once (so $P_j(\geq 2) > 0$), and positivity is immediate. With Poisson usable counts it is $\mathbb{E}_j[x_j e^{-x_j}]/\mathbb{E}_j[1-e^{-x_j}]$, and $s_{\text{piv}}(x) = xe^{-x}/(1-e^{-x})$ in the homogeneous case; the Poisson average understates the exact value on a lumpy graph, which is why the text quotes the exact discrete ε_K. This is (5).

Convexity. Take the representative cross-section $x_j \equiv x = cK(1-\tilde{e}_w)$ with $\tilde{e}_w = \bar{e}$, so the incumbent searcher hazard is $T/(1-\bar{e}) = \alpha_R\,\bar{e}\,(1-e^{-x})/(1-\bar{e})$. Employing a vanishing random share of searchers raises $\bar{e}$, and since $x = cK(1-\bar{e})$ gives $d\log(1-e^{-x})/d\log\bar{e} = -\frac{\bar{e}}{1-\bar{e}}s_{\text{piv}}(x)$,

$$\eta_{\text{eff}} \;=\; \frac{d\log\!\big[\bar{e}(1-e^{-x})/(1-\bar{e})\big]}{d\log\bar{e}} \;=\; \underbrace{1}_{\text{more holders}} \;-\; \underbrace{\tfrac{\bar{e}}{1-\bar{e}}s_{\text{piv}}(x)}_{\text{queues thin}} \;+\; \underbrace{\tfrac{\bar{e}}{1-\bar{e}}}_{\text{fewer rivals}} \;=\; 1+\frac{\bar{e}}{1-\bar{e}}\big(1-s_{\text{piv}}(x)\big).$$

The deliverability loss and the competition relief combine into the appropriation share $1-s_{\text{piv}}$ scaled by the employment odds $\bar{e}/(1-\bar{e})$, so $\eta_{\text{eff}} > 1$ strictly; this is (6). Referrer heterogeneity acts through the same competition-relief channel with $1-s_{\text{piv}}$ replaced by its delivered-weighted average, which is why the exact cross-sectional η_{eff} of Table 2 exceeds the representative value; the replication ledger reports those values and confirms $\varepsilon_K = s_{\text{piv}}$ to three decimals at $c \in \{5, 10, 20\}$.

D.2 Proof of Proposition 2

Adding up. Every delivered lead is received by exactly one searcher, so aggregate receipts equal aggregate deliveries identically, at every realization; any redistribution of receipts across searchers at fixed deliveries is a transfer. It follows that the aggregate matching flow responds to a searcher's effort only through the *delivered total*.

The pivotal margin. Model effort as activation: searcher i's effort a_i scales the probability that each of her ties is usable. Consider the delivered probability of one referrer with usable-set composition given by independent activations: $1-\prod_k(1-q_k)$, where q_k is the usability probability of contact k. Its derivative in q_i is $\prod_{k\neq i}(1-q_k)$ — the probability that *no other* contact is usable. So the aggregate-flow effect of i's effort counts exactly the states in which she is the sole usable candidate.

The identity, for any distribution of ties. Let referrer j have contacts usable independently with probabilities $\{a_{jk}\}$, so $P_j = 1 - \prod_k (1 - a_{jk})$ is her delivery probability and the aggregate delivered flow is $D = \sum_j P_j$. Two routes both evaluate the creation share, exactly and with no distributional assumption.

Route 1 (delivered-flow elasticity). Scale every usability by a common activation, $a_{jk} = a\, b_{jk}$. The events "contact k is the *unique* usable contact of j" are disjoint over k and their union is "exactly one usable at j," so $\sum_k a_{jk} \prod_{l \neq k} (1 - a_{jl}) = P_j(\text{exactly one usable})$, and therefore

$$\frac{dD}{d \log a} = \sum_j \sum_k a_{jk} \prod_{l \neq k} (1 - a_{jl}) = \sum_j P_j(\text{exactly one usable}).$$

The elasticity of the delivered flow with respect to symmetric effort is thus $\big[\sum_j P_j(\text{exactly one})\big] / \big[\sum_j P_j(\text{at}$ Because aggregate matches equal the delivered total, this is precisely the share of the effort margin that creates rather than reallocates.

Route 2 (recipient-side sole share). Condition on a delivered lead. It issues from referrer j with probability $P_j / \sum_l P_l$, and given j the uniformly chosen recipient was j's only usable candidate with probability $P_j(\text{exactly one})/P_j$. Averaging,

$$\mathbb{P}(\text{recipient sole} \mid \text{delivered}) = \sum_j \frac{P_j}{\sum_l P_l} \cdot \frac{P_j(\text{exactly one})}{P_j} = \frac{\sum_j P_j(\text{exactly one usable})}{\sum_j P_j(\text{at least one usable})}.$$

The two routes coincide identically, so

$$s_{\text{piv}} \;=\; \frac{\sum_j P_j(\text{exactly one usable})}{\sum_j P_j(\text{at least one usable})} \;\in\; (0, 1],$$

for any distribution of ties and any market state, at a symmetric activation margin. Given a realized configuration of usable sets the ratio is a property of that configuration; but under *asymmetric* effort an individual searcher's own creation share is weighted by her exposure across her referrers and need not equal this unconditional delivered-referral share. The policy analysis uses the symmetric aggregate margin throughout. In the homogeneous Poisson case, with the competing usable count $U \sim \text{Poisson}(x)$, both routes reduce to $\mathbb{P}(U = 0)/\mathbb{E}[1/(1 + U)] = e^{-x} / \frac{1 - e^{-x}}{x} = s_{\text{piv}}(x)$, the closed form. Monotonicity of $s_{\text{piv}}(x)$ is elementary; $s_{\text{piv}} < 1$ exactly when $P_j(\geq 2) > 0$ for a positive delivered mass. The replication ledger confirms the two routes agree to 10^{-3} on simulated fixed networks, and that the exact discrete value (0.80 at the calibration) exceeds the heterogeneity-averaged Poisson $\mathbb{E}_j[x_j e^{-x_j}] / \mathbb{E}_j[1 - e^{-x_j}]$ (0.58), because lumpy strong edges make two-usable states rarer than a Poisson queue of equal mean. □

Attention-weighted allocation. If leads are allocated within the usable set in proportion to attention rather than uniformly, effort gains a second, purely appropriative margin: with activation held fixed, the delivered total $1 - \prod_k (1 - q_k)$ does not depend on the allocation weights at all, so any effort-induced re-weighting is a pure transfer — the replication ledger checks this as an exact invariance. The pivotal share under the blended margin is correspondingly lower (0.73 against 0.80 at the calibrated steady state); the text's results use the conservative activation-only case.

D.3 Proof of Proposition 3

Searchers of mass one choose effort a at cost $\psi(a) = \frac{\kappa_a}{2} a^2$; activation scales usable intensity, so the delivered flow per searcher is $D(\bar{a}) = \varrho\,(1 - e^{-x_0 \bar{a}})$ at symmetric effort $\bar{a}$, and an individual deviating to a receives $\tau(a; \bar{a}) = D(\bar{a})\, a / \bar{a}$ (shares are proportional to activation under uniform allocation).

Lemma 8 (The effort game is well posed)**.** *For any instrument σ, the individual objective* $\mathrm{PV}\, \tau(a; \bar{a}) + \sigma a - \psi(a)$ *is strictly concave in a — second derivative $-\kappa_a < 0$, since τ is linear in the deviation a — so each searcher's best response $a(\bar{a}) = [\mathrm{PV}\, D(\bar{a})/\bar{a} + \sigma]/\kappa_a$ is unique and satisfies its second-order condition. Delivered flow per unit effort $D(\bar{a})/\bar{a}$ is strictly decreasing, because $u e^{-u} < 1 - e^{-u}$; hence $a(\bar{a})$ is strictly decreasing and the symmetric equilibrium $\bar{a} = a(\bar{a})$ is unique, and interior whenever $\sigma > -\mathrm{PV}\, \varrho x_0$. The planner's objective* $\mathrm{SV}\, D(\bar{a}) - \psi(\bar{a})$ *has second derivative* $\mathrm{SV}\, D''(\bar{a}) - \kappa_a < 0$ *(concave D), so its optimum is unique. The sign of the corrective instrument is therefore well defined relative to the unique equilibrium.*

The private first-order condition with a per-effort instrument σ is $\kappa_a \bar{a} = \mathrm{PV} \cdot D(\bar{a})/\bar{a} + \sigma$; the planner, valuing a match at $\mathrm{SV} = (1 + w)\mathrm{PV}$, maximizes $\mathrm{SV}\, D(\bar{a}) - \frac{\kappa_a}{2} \bar{a}^2$, giving $\kappa_a \bar{a}^* = \mathrm{SV}\, D'(\bar{a}^*)$. The decentralizing instrument is

$$\sigma^* \;=\; \mathrm{SV}\, D'(\bar{a}^*) - \mathrm{PV}\, \frac{D(\bar{a}^*)}{\bar{a}^*} \;=\; \frac{D(\bar{a}^*)}{\bar{a}^*}\, \mathrm{PV} \left[s_{\mathrm{piv}}(x_0 \bar{a}^*)(1 + w) - 1 \right],$$

using $D' \bar{a}/D = s_{\mathrm{piv}}$ (the aggregate pivotal share). The sign claim follows; dividing by the broadcast benchmark ($s_{\mathrm{piv}} \equiv 1$) gives the surviving fraction in the proposition, and setting the blended expression $\mu_{\mathrm{NR}} w + (1 - \mu_{\mathrm{NR}})[s_{\mathrm{piv}}(1 + w) - 1]$ to zero gives $\bar{\mu}$. The replication ledger verifies the sign prediction against the numerically solved planner and equilibrium efforts across a grid of wedges, along with the over-search ranking (equilibrium effort exceeds the planner's exactly when $\sigma^* < 0$). □

D.4 Corollary 1 and finite perturbations

For a group holding usability share s_g of the searcher mass, with effort factor $(1+d)$: the queue intensity becomes $x(1+s_g d)$, so delivered flow scales by $(1-e^{-x(1+s_g d)})/(1-e^{-x})$; the group's receipts scale by that times $(1+d)/(1+s_g d)$ and incumbents' by that times $1/(1+s_g d)$. Differentiating the incumbent expression at $d=0$ gives $d\log(\text{incumbent hazard})/dd = s_g(s_{\text{piv}}-1)$. The perturbation moves the *aggregate* competing usable mass at the same rate, $d\log(1+s_g d)/dd = s_g$, so the elasticity of the incumbent hazard with respect to that mass is $s_{\text{piv}}-1$, independent of the perturbed group's size s_g — the statement of Corollary 1. (With respect to the group's *own* mass, which moves at unit rate, it would instead be $s_g(s_{\text{piv}}-1)$.) The finite-change formulas above are exact and are the ones the ledger verifies. Under broadcast no searcher's hazard references any other searcher, and the elasticity is exactly zero.

D.5 Corollary 2

Marginal relief. By the posterior identity of D.2, conditional on a searcher's received lead, the lead passes to a rival upon her exit with probability $1-s_{\text{piv},h}$ and expires otherwise; hence the marginal relief $(1-s_{\text{piv},h})\tau_h$. A finite exit of usability share σ recovers the smaller fraction

$$\text{recov}(\sigma, x) \;=\; 1 - \frac{e^{-x(1-\sigma)} - e^{-x}}{\sigma\,(1-e^{-x})},$$

(each exit leaves the remainder more often pivotal), obtained directly from the delivered-flow expressions of D.4 with $d=-1$ on the exiting group. *Destination congestion* applies D.4 in reverse to the entered queue, with the bridged cohort's usability mass added to x_b; the private gain and the two external terms then satisfy the exact accounting $\Delta(\text{total delivered}) = \text{private} + \text{net externality}$, which the ledger verifies to numerical precision, along with the closed forms against exact cross-section expectations (within 7%) and the fixed-network replication of the relief (within 5%). The de-correlation premium quoted in the text is the difference between the healthy-destination and slack-destination nets computed on the same home cross-section and bridged cohort.

D.6 Sensitivity

The values of s_{piv} quoted in the text — 0.80 at the calibrated steady state, 0.43 at the transient trough, 0.71 under chronic slackness — are point evaluations of the exact discrete form on the respective simulated cross-sections. Monotonicity in the queue intensity x is a property of the formula, not the calibration, and is what the policy statements rely on; the level at any particular unemployment rate moves with the contact scale c and the state of tie capital as $x = cK(1 - \tilde{e}_w)$ dictates.

E. Estimator Construction and Monte Carlo

E.1 The yield decomposition

In market m, let y_m be referral offers received per unit of searchers' employed-contact mass. Under the blend, exactly,

$$y_m \;=\; A_{NR} \;+\; A_R\, R_m, \qquad R_m \;=\; \frac{\bar{e}_m\, F_m}{\tilde{e}_m\,(1-\bar{e}_m)\,\mathbb{E}_m[n \mid u]}, \qquad F_m \;=\; \mathbb{E}_{j\in\text{employed}}\big[1 - e^{-x_j}\big],$$

where F_m is the deliverability functional — computed from the *distribution* of employed workers' unemployed-contact intensities, not its mean. The distributional form matters: deliverability is concave in contacts, market heterogeneity makes $\mathbb{E}[1-e^{-x_j}]$ fall increasingly short of $1 - e^{-\bar{x}}$ as markets slacken, and a mean-based regressor loads that drift onto the intercept that carries μ_{NR} (a bias of -0.13 at $\mu_{\text{NR}} = 0.25$ in the Monte Carlo below). Every component of R_m is observable in data with two-sided contact composition.

E.2 The two-step estimator

Step one measures R_m per market from the contact module. Step two runs nonnegative least squares of y_m on $(1, R_m/R_{\text{ref}})$, with R_{ref} interpolated at the reference density ($u = 6\%$), and reports $\hat{\mu_{\text{NR}}} = \hat{A}_{NR}/(\hat{A}_{NR} + \hat{A}_R)$. The contact scale follows from the through-origin regression of measured queue intensity on (mean contacts $\times$ unemployment). The test of pure broadcast is the fit-improvement statistic of the unrestricted curve over a flat line, with critical values from the $\mu_{\text{NR}} = 1$ null distribution — exact size by construction.

E.3 Monte Carlo design and results

Data are simulated from the blended model: 24 markets spanning separation rates that generate unemployment from 3.5 to 23 percent, $N = 6{,}000$ workers per market, 50 quarters observed after burn-in, 20 replications per point of the μ_{NR} grid; the econometrician sees offers, contacts by status on both sides, and unemployment — never channel labels. The DGP is calibrated so every blend shares the same steady state (Appendix A), so markets differ by density, not by construction. Results are in the text (Section 7.2); beyond the table in the text, the ledger records: the label-free shape check (the rival component fits the lottery-share curve with $R^2 = 0.98$ on pure-rival data); the weak-identification exhibit (free-scale nonlinear least squares on yield data alone: $\hat{\mu_{NR}}$ standard deviation 0.24, corner collapse); the queue-scale measurement (median error 0.4%); and the misspecification run — the estimator fitted to data from the fixed-network model of Section 6, returning $\hat{\mu_{NR}} = 0.47$ against 0.50, because F_m is measured rather than assumed Poisson.

E.4 The offers-to-hires correction (open)

The design consumes offer or lead counts. Most field sources record realized referral *hires*; mapping the model's multi-offer distribution to an at-most-one-hire observable is a composition correction that is tractable — the offer process is fully specified — but not implemented here, and any hires-based application should build it first.

F. Fixed-Network Benchmark Designs

F.1 The fixed-network model

The annealed simulator cannot be patched into a fixed-network one; the tie layer is rebuilt at the edge level. Workers occupy the nodes of a block configuration-model graph (Poisson degrees with mean c, mutual edges, sector mixing from a symmetric matrix with the same strong-partner structure fed to all three models). Each directed edge slot $(i \to j)$ carries a strength $a_{ij} \in [0, 1]$ — worker i's side of the relationship — which grows by $\phi(1 - a_{ij})$ while i is employed and decays at d_{eu} or d_{uu} according to the *actual* status of neighbor j. Under broadcast, i's referral hazard collects $(\alpha/c) \sum_j a_{ij} z_j$ over her actual employed neighbors. Under the token protocol, edge strength is read as usability: a lead-holding worker's unemployed neighbors are each usable with probability a_{jk}, the lead goes to one uniformly among the usable, and expires if none — which reproduces the semi-analytic deliverability form with local quantities. Annealed matching is exactly this model with neighbors redrawn tie-weighted each period, so the three-way comparison isolates finite-population error (annealed versus mean field) from network-fixedness error (fixed graph versus annealed).

F.2 Experimental designs

All benchmarks share: identical calibration across the three models; populations of 20,000 (single-sector) and 48,000 (six-sector) workers; burn-in to stationarity with drift checks; and common-random-number coupling — shocked and unshocked runs share the burn-in state and the identical uniform sequence, so scars are measured as coupled differences (with per-period seeded substreams where token draws make counts data-dependent).

The correlated-shock experiment mirrors the structure that should stress mean field most: six sectors, two of them strong partners; separations quintupled for eight quarters in the focal sector and *either* its partner (correlated) or an unrelated sector (spread); the object is the difference in the focal sector's scar, measured eighteen to twenty-two quarters after the shock ends, averaged over the symmetric pair where design symmetry permits, with eight seeds and three coupled replicates per seed. Because employment-based scars sit near the demographic noise floor at feasible populations, the tie-capital scar — roughly ten times more precise, and the variable the mechanism acts through — accompanies every employment comparison. Success criteria for every benchmark were fixed before the full runs, and amendments made after machinery smoke tests (never after full-scale results) are documented in the scripts' headers.

F.3 Results behind the text's table

Broadcast regime: steady state within 0.3 percentage points of mean field; transient-shock scar within one percent; correlated-shock deepening 3.0% (employment, s.e. 1.7pp) and 3.1% (tie capital, s.e. 0.1pp) against mean field's 3.8%/3.3% — the tie-capital channel excludes deviations beyond $\sim$30% and puts the point estimate at a 0.8–0.95 conservative correction. Token regime: deepening 21.9% (s.e. 9.3pp) and 5.7% (s.e. 0.2pp) against mean field's 24.7%/6.9%; elasticities on the fixed graph $\varepsilon_K = 0.79$, $\eta_{\text{eff}} = 4.97$ against 0.81 and 4.31 for the annealed semi-analytic form evaluated on the same cross-sections, with the full pattern across $c \in \{5, 10, 20\}$ reproduced; congestion elasticity -0.22 ± 0.06 against the local closed form's -0.20. *Aggregator decomposition:* recomputing the mean-field-R closure with a linear finding rate $f = f_{\text{dir}} + \tau$ in place of $1 - (1 - f_{\text{dir}})e^{-\tau}$, with α_R recalibrated to restore the same steady state, gives a correlated-shock deepening of 24.6% (employment) and 7.1% (tie capital), against 24.7%/6.9% under the Poisson aggregator and 3.8%/3.3% for broadcast (the same linear aggregator with no queue): the token queue accounts for the entire token–broadcast gap, and the concave aggregator, if anything, dampens it. Power checks confirm the comparisons are not vacuous: edge-level unemployment correlation at the correlated-shock trough reaches $+0.08$, twenty-five times its steady-state level. Robustness: ring-lattice sectors with local clustering near 0.6 and sparse graphs ($c = 5$) leave the deepening statistically unchanged.

F.4 Scope

One calibration; one shock size and duration; configuration-model degree distributions. Hub-targeted shocks — removal of high-degree brokers, where sector-scalar aggregates are known to fail — are outside the designs and outside the paper's claims.