

Referral Networks and Inequality*

Manolis Galenianos[†]

June 2017

Abstract

I develop a theoretical model to study the interaction between labor markets, endogenous networks and worker heterogeneity and its implications for inequality and welfare. Consistent with empirical evidence, in the model referred workers are more likely to be hired, to receive a higher wage and to be more productive. The use of referrals exacerbates inequality. If heterogeneity is driven by productivity differences, then the use of referrals improves welfare. If workers face different probability of forming a match despite having the same productivity, as in the case of discrimination, then referrals worsen welfare.

*I would like to thank seminar and conference participants for useful comments.

[†]Royal Holloway, University of London. Email: manolis.galenianos@gmail.com

1 Introduction

It is well-known that personal contacts play an important role in labor markets: approximately half of all workers report finding their jobs through referrals (Granovetter, 1995).¹ Perhaps surprisingly, however, there is no consensus about the welfare consequences of using referrals in the labor market. On the one hand, some authors have argued that the use of referral networks may increase inequality by benefitting the networked at the expense of qualified but less well-connected workers.² On the other hand, several recent empirical studies find that referred workers have better labor market outcomes, suggesting that referrals might alleviate some of the informational frictions that pervade labor markets.³

These views are not necessarily incompatible with each other; rather, they suggest that determining the welfare consequences of referrals requires taking into account several factors which could individually point to different directions. The first such factor is the importance of worker heterogeneity and the frictions associated with identifying good workers. The second factor is that workers and firms use both formal and informal channels to search and these channels will, in general, have different informational content. The third, and most crucial, factor is that referral networks are the outcome of choices made by workers, rather than exogenous and immutable features of the environment. Taking into account the endogenous nature of referral networks is the key both for understanding the potential of referrals to alleviate informational frictions and exacerbate inequality and also for determining the welfare consequences of referral use.

In this paper I build the first theoretical model that studies the interaction between labor markets, endogenous referral networks and worker heterogeneity. The model provides an

¹The surveys of Ioannides and Loury (2004) and Topa (2011) provide a wealth of related evidence.

²Calvo-Armengol and Jackson (2004) make this point theoretically in a model where agents learn about jobs from their neighbors. Empirically, Topa (2001) finds strong local spillovers in unemployment rates across geographical locations in Chicago, Kugler (2003) finds that referrals are associated with higher wage inequality across equally productive workers and Beaman, Keleher and Magruder (2015) present evidence that the use of referrals reinforces unequal access to jobs between men and women.

³See Dustmann et al (2016), Brown, Setren and Topa (2015) and Hensvik and Skans (2015) among others.

accurate qualitative description of the informational advantage of referrals and predicts that the use of referrals exacerbates inequality in the labor market. Using referrals might either increase or decrease welfare and the outcome depends on the nature of worker heterogeneity. Specifically, when worker productivity is highly heterogeneous, the use of referrals improves welfare despite increasing inequality. When, however, workers have similar productivity but face consistently different probabilities of being hired, as in the case where discrimination is prevalent, the use of referrals exacerbates the effects of discrimination and is detrimental to welfare.

In the model there are two worker types, A and B . The network is formed at an initial stage and, subsequently, workers and firms interact in a frictional labor market. In the labor market vacancies are created in two ways: a new firm enters or a producing firm expands. Firms and workers meet either through a referral, which occurs when a producing firm expands and asks its current employee to refer a member of his network, or through search in the frictional market. At a meeting, the probability that a match is formed depends on the worker's type. Type- A workers weakly dominate type- B s in productivity when employed and probability of forming a match when meeting a firm.

The equilibrium admits a sharp characterization. The benefit of forming a link is that a referral may be provided at some future date. Type- A workers have higher employment rates and are, therefore, more likely to provide a referral, which makes them more desirable as links. In equilibrium, workers *of both types* have most of their links with type- A workers and type- A workers enjoy a higher arrival rate of referrals. As a result, the equilibrium network has a hierarchical structure.⁴ This is a novel result: the economics literature typically assumes homophily (Montgomery, 1991) motivated by sociological evidence that social networks are homophilous across observable characteristics. The connection between this paper's results and sociological evidence is discussed in Section 4.1.

An implication of the network's hierarchical structure is that referrals are mostly received

⁴I use the term "hierarchical network" in its literal sense rather than the graph-theoretic terminology.

by type-*A* workers and, therefore, the use of referrals exacerbates inequality by disproportionately improving the employment opportunities of *A*-types. This referral-induced inequality is increasing in the strength of network hierarchy which, in turn, depends positively on the relative importance of informal search channels and on the cross-type differential in the probability of forming a match when meeting with a firm.

The positive predictions of the model about referred workers' labor market outcomes are consistent with a wealth of empirical findings: conditional on observable characteristics, referred candidates are more likely to be hired (Fernandez and Weinberg, 1997; Castilla, 2005; Brown, Setren and Topa, 2015; Burks et al, 2015), to receive higher wages (Simon and Warner, 1992; Bayer Ross and Topa, 2008; Brown, Setren and Topa, 2015; Hensvik and Skans, 2015; Dustmann et al, 2016) and to be more productive (Castilla, 2005; Pinkston, 2012; Burks et al, 2015; Barr, Bojilov and Munasinghe, 2016).

The optimal network is generically different from the equilibrium network. When productivity differentials are high the optimal network exacerbates inequality and, therefore, in that case the equilibrium network's hierarchical structure is welfare-improving. When productivity across types is similar but the probability of forming a match is greater for type-*A* workers the optimal network *reduces* inequality and, therefore, in this case the use of referrals is strictly detrimental to welfare. The latter case is consistent with a labor market where discrimination is prevalent. Policies that promote hiring through the market can help alleviate (though not resolve) the effects of discrimination.

The theoretical literature that studies the interaction between social networks and economic activity is vast and I will focus on the subset that relates to labor markets. The interaction between heterogeneous workers and referrals was first studied by Montgomery (1991) who assumes an exogenous and homophilous network. In that paper, as in the present one, using referrals helps firms hire high-productivity workers but there is no exploration of the welfare properties of equilibrium or the determinants of network structure. All of the following papers assume that workers are homogeneous in terms of their productivity and

the role of referrals is to facilitate the search process.⁵ The effect of network structure on sustaining inequality is examined in Calvo-Armengol and Jackson (2004) where the network structure is exogenous. Calvo-Armengol (2004) and Galeotti and Merlino (2014) endogenize the network and examine its equilibrium features such as network density and architecture. Cahuc and Fontaine (2009) studies the welfare consequences of using informal search methods and find that network use is inefficient. Mortensen and Vishwanath (1994) and Igarashi (2016) examine the effect of heterogeneous access to referrals on the wage distribution and on the welfare consequences of banning referrals, respectively. Galenianos (2014) studies the effect of referrals on matching efficiency.

Section 2 presents the model, Section 3 characterizes the equilibrium and derives its theoretical implications, Section 4 compares the model’s empirical predictions with the data and Section 5 examines the welfare properties of equilibrium. Section 6 concludes and proofs are collected in the Appendix.

2 The Model

Workers and firms populate the economy. There is measure 1 of type- A workers and measure 1 of type- B workers.⁶ A free entry condition determines the measure of firms.

There are two distinct stages. In the first stage workers form a referral network. In the second stage workers and firms interact in a frictional labor market. Network formation is costly but creates the opportunity of finding a job through a referral in the frictional labor

⁵In a different strand of the literature, referrals help firms screen for better match quality, e.g. Simon and Warner (1992), Galenianos (2013), Brown, Setren and Topa (2016) and Dustmann et al (2016). These papers focus on the firm’s screening problem and do not, typically, include an explicit model of the network among workers.

⁶Having arbitrary measures of the two types yields qualitatively identical results in the labor market (proof available upon request). The population share of the two types affects network formation but in a somewhat uninteresting way: if there are very few type- B workers then the network of a type- A worker will mechanically include very few B -types. For this reason, the focus of this paper is on the case of equal measures for each worker type which, in one of the model’s applications, can be thought of as above-median and below-median according to some attribute such as employability or productivity. Modeling race or ethnic minorities introduces some complications which are discussed in the Conclusions.

market. A worker’s referral network represents the set of workers that he might refer, and be referred by, if a job opportunity appears, rather than the full set of people with whom he interacts socially (see Section 4.1 for a detailed discussion).

2.1 Network Formation

The formation of referral networks is modeled as a large, non-cooperative game with non-transferable utility and the focus is on symmetric networks. Denote the network of worker j of type i by n_i^j , the cost of forming the network by $C_i(n_i^j)$ and the steady state utility in the labor market by $\Lambda_i(n_i^j)$.⁷ Worker j chooses n_i^j to maximize:⁸

$$\mathcal{L}_i(n_i^j) = \Lambda_i(n_i^j) - C_i(n_i^j). \quad (1)$$

The network of a worker consists of the measure of links that he has with type- A and type- B workers. Denoting the measure of links that worker j of type i has with workers of type k by n_{ik}^j , his network is given by $n_i^j = (n_{iA}^j, n_{iB}^j) \in [0, 1] \times [0, 1]$.

Modeling a worker’s network as a continuum is consistent with the (spirit of the) sociology literature’s finding that the most helpful links for finding a job are the more numerous weak links (Granovetter, 1973; 1995) and is crucial for the model’s tractability (Galenianos, 2014). A worker’s employment opportunities will in general depend on how many of his links are employed which necessitates keeping track of each link’s time-varying employment status. Having a continuum of links means that the aggregate (un)employment rate of a worker’s contacts is deterministic (and constant, in steady state) due to the law of large numbers, thereby greatly simplifying the analysis. Although some of the richness of network architecture that can be found, for instance, in graph-theoretic models of networks cannot

⁷In the labor market, the worker transits between employment and unemployment. Steady state utility is calculated using the proportion of time that he spends at each labor market state which, in general, depends on his network.

⁸To be more precise, the worker chooses effort which maps uniquely into the network outcome in a way that depends on other workers’ effort. This is detailed below.

be replicated in a continuum model, this is less important for this paper's focus.

The measure of links that a worker acquires depends on the effort he exerts in network formation and on the aggregate effort of all other workers. Workers simultaneously choose their effort levels. Denote the effort of worker j of type i by (e_{iA}^j, e_{iB}^j) where $e_{ik}^j \geq 0$ is the effort he exerts in linking with type- k workers. The aggregate effort of all workers is denoted by $(E_{AA}, E_{AB}, E_{BA}, E_{BB})$ where E_{ik} denotes the effort of type- i workers towards linking with type- k workers and will be referred to as the “demand” for such links. I will focus attention on equilibria where $e_{ik}^j = E_{ik}$ for all i, j, k , which will deliver symmetric networks conditional on type.

Effort is mapped to links as follows. If there is positive demand for within- i links ($E_{ii} > 0$), then the measure of links that worker j of type i forms is equal to his effort: $n_{ii}^j = e_{ii}^j$; if there is no demand ($E_{ii} = 0$), then he forms no such links ($n_{ii}^j = 0$) regardless of his effort level. Therefore:

$$n_{ii}^j(e_{ii}^j) = e_{ii}^j I[E_{ii} > 0]$$

where the indicator function takes a value of 1 if $E_{ii} > 0$ and 0 if $E_{ii} = 0$.

The measure of links between workers of different types needs to respect aggregate consistency, i.e. the A -types' measure of links with B -type workers is equal to the measure of links that B -types have with A -type workers: $n_{AB} = n_{BA}$ (recall that the economy is populated by equal measures of A - and B -workers). The measure of links as a function of effort is given by

$$n_{ik}^j(e_{ik}^j) = e_{ik}^j \frac{E_{ki}}{E_{ik} + E_{ki}}$$

if $E_{ik} > 0$ or $E_{ki} > 0$ and zero otherwise. Under this formulation, the probability of forming across-type links depends in a natural way on the relative demand for links across types.

It is worth reiterating that the mapping from a worker's effort to his actual network depends on the effort of other workers. This interaction will be crucial for the network structure that arises in equilibrium.

I assume that cost is quadratic in effort: $\tilde{C}_i(e_{ii}^j, e_{ik}^j) = \frac{c}{2} \left((e_{ii}^j)^2 + (e_{ik}^j)^2 \right)$. The cost of network formation satisfies:

$$C_i(n_{ii}^j, n_{ik}^j) = \frac{c}{2} \left((n_{ii}^j (e_{ii}^j))^2 + \left(\frac{n_{ik}^j (e_{ik}^j) (E_{ki} + E_{ik})}{E_{ki}} \right)^2 \right)$$

for $E_{ii} > 0$ and $E_{ki} > 0$. The cost is infinite in the case where $E_{ii} = 0$ and $n_{ii}^j > 0$ and in the case where $E_{ki} = 0$ and $n_{ik}^j > 0$. I assume throughout that c is high enough to have an interior solution. Notice that the cost of effort is symmetric across the two types and, therefore, any differences in their networks are due to behavior rather than exogenous factors.

The effort cost is separable in each of the two types which means that the marginal cost of forming links starts at zero for both types. The motivation behind this assumption is that some links might be easier to create for non-pecuniary and unmodeled reasons, such as friendship. To the extent that the distribution of non-pecuniary benefits is not perfectly correlated with one's type (and, for example, an A -type worker derives positive non-pecuniary benefits from B -type workers and vice versa) this effect is captured in a reduced-form way by the separability in the cost of network formation: the marginal cost, net of non-pecuniary benefits, of linking with workers of a type starts at zero because the initial link is formed with the worker who provides the highest level of non-pecuniary benefits.

Denote the proportion of links that a worker type i has with his own type by ϕ_i :

$$\phi_i = \frac{n_{ii}}{n_{ii} + n_{ik}}. \tag{2}$$

A referral network is called homophilous when $\phi_i \geq \frac{1}{2}$ for $i \in \{A, B\}$ and hierarchical when

$$\phi \geq \frac{1}{2} \geq \phi_k.^9$$

2.2 The Labor Market

I first describe a labor market where all workers have the same network, conditional on type. I then derive the value functions for a worker with an off-equilibrium network, which will be used to solve for the equilibrium network.

2.2.1 Production

Time runs continuously, the horizon is infinite, the discount rate is $r > 0$ and the labor market is in steady state. Firms are homogeneous, risk-neutral and maximize expected discounted profits. Each firm hires one worker and a firm is either filled and producing or vacant and searching, where k denotes the flow cost of a vacancy.

Workers are heterogeneous, risk-neutral and maximize expected discounted utility. A worker is either employed or unemployed and the flow utility of unemployment is $b > 0$. There are two dimensions of heterogeneity: the probability of forming a match when meeting with a firm (p) and the productivity when employed (y). The probability of forming a match and productivity when employed are weakly higher for A -types ($p_A \geq p_B$ and $y_A \geq y_B$) and B -types generate positive surplus when employed ($y_B > b$). All payoff-relevant variables (e.g. worker's type and network) are common knowledge when the match is formed. Matches are destroyed at rate δ .

Worker heterogeneity captures attributes that the firm cannot readily advertise for. Specifically, worker heterogeneity captures characteristics that are not contractible (e.g. productivity beyond observables) or which cannot legally be used for recruiting purposes (e.g. gender). The information embedded in a worker's network is valuable precisely in settings where information that is harder to advertise for is more relevant.

⁹For completeness, a network is heterophilous when $\phi_i \leq \frac{1}{2}$ for $i \in \{A, B\}$.

The two dimensions of heterogeneity provide a parsimonious way of capturing a great variety of economic environments. To fix ideas, I provide two examples of very different environments which will turn out to have very different implications for welfare, as shown in Section 5.

Example 1: Productivity Differentials. When a firm and a type- i worker meet they draw the match specific productivity from some continuous and log-concave distribution $F_i(\cdot)$ where the distribution of A -type workers first order stochastically dominates that of B -types. In equilibrium, the match is formed if the draw is above an endogenous cutoff, a type- A worker is more likely to make a draw above his cutoff than a B -type and, conditional on forming a match, his productivity is higher on average than that of an employed B -type. This example is consistent with a labor market where worker heterogeneity is due to productivity differences.

Example 2: Discrimination. When a firm and a worker meet they draw match quality which is good with probability \bar{p} and bad with probability $1 - \bar{p}$ for both worker types. A good match produces \bar{y} and a bad match produces $\underline{y} < b$ for both types. Bad matches are never formed (as they are unprofitable), good matches with A -types are always formed and good matches with B -types are formed with probability $\zeta < 1$. These assumptions lead to $y_A = y_B = \bar{y}$ and $p_B = \bar{p}\zeta < \bar{p} = p_A$. The assumption that $\zeta < 1$ can be interpreted as discrimination in the labor market since the heterogeneity in the probability of forming a match is not driven by productivity differentials.

I provide a rough outline of two ways of rationalizing $\zeta < 1$ that are consistent, respectively, with statistical and taste discrimination (a more detailed analysis is beyond the scope of this paper). First, suppose that match quality is uncertain when the firm and worker meet and the firm draws a binary signal about match quality. For A -type workers the signal is always accurate and for B -type workers it is always accurate if match quality is bad but it is accurate with probability ζ if match quality is good.¹⁰ In equilibrium a match is

¹⁰Statistical discrimination has often been modeled to arise due to less accurate productivity signals by

only formed after a good signal, so long as the ζ is not too high. Second, suppose that a share $1 - \zeta$ of recruiters incur a high disutility from hiring B -workers regardless of match quality and, therefore, B -workers are hired only if match quality is good *and* they meet a non-discriminatory recruiter.¹¹

2.2.2 Market and referrals

Vacancies are created in two ways: (1) a new firm enters and starts searching through the market; (2) an existing firm expands which occurs at exogenous rate ρ , where $\rho < \delta$. Following an expansion, the new position is sold off, keeping firms' employment at one worker. A firm and a worker meet either through search in the market or through a referral, which occurs when a firm expands and asks its current employee to refer a link. A meeting function determines the rate of meeting through the market and firms' expansion rate determines the rate of meeting through referrals.

When a firm that employs a worker expands, the worker refers one of his links at random. If the referred worker is unemployed, he meets with the firm and they form a match with the type-specific probability. If the match is not formed or the referred worker is employed then the referral opportunity is lost and search in the market begins.

Two remarks is in order. First, although the referring worker does not select a particular worker type to refer to the firm, he does choose the worker types that he is linked with. In other words, the worker's strategic considerations about who to refer are incorporated in the network formation stage rather than the referral stage and this is, essentially, an assumption about timing rather than substance.¹²

Second, the firm benefits from the referral in (potentially) two ways: it samples a worker

the discriminated group, e.g. Aigner and Cain (1977).

¹¹Discrimination might take place at any level of the recruiting process and, therefore, does not need to be identified with the firm's owner or the firm's policy.

¹²The model can be reformulated so that the network-formation and referral decisions are made, and associated costs are borne, during the labor market stage. Such a reformulation complicates the analysis but does not qualitatively affect the allocation so long as a worker cannot adjust his network in response to his time-varying employment status (this possibility is discussed in the Conclusions).

immediately without having to go through time-consuming search in the market which can only add to the firm's value; furthermore, the worker is drawn from a different type distribution than the unemployment pool which might, and in equilibrium will, turn out to be better than the type distribution among the unemployed. Given a choice, therefore, the firm prefers to use a referral rather than forgo the opportunity.

A referral from a type- i worker arrives at an unemployed type- i worker with probability $\phi_i u_i$ and an unemployed type- k ($\neq i$) worker with probability $(1 - \phi_i) u_k$, where u_k is the unemployment rate of type- k workers. The composition of worker types in the referral pool will generally differ from that in the unemployment pool.

Denoting the value of employing a type- i worker by J_i and the value of a vacancy by V , the surplus generated by an expansion when employing a type- i worker is equal to:

$$X_i = V + \phi_i u_i p_i (J_i - V) + (1 - \phi_i) u_k p_k (J_k - V). \quad (3)$$

The flow value to a match between a firm and a type- i worker is $y_i + \rho \gamma X_i$, where the incumbent match receives share $\gamma \in [0, 1]$ of the expansion's surplus. Notice that a firm which enters the labor market anew receives value V , the value of a vacancy. The second and third term of equation (3) reflect the additional value of a referral.

When unemployed, worker j of type i receives a referral when the employer of one of his links expands and worker j is chosen among the referrer's links. Worker j has n_{ii}^j links with type- i workers and n_{ik}^j links with type- k workers. Each link of type i is employed with probability $1 - u_i$ and gets the opportunity to refer at rate ρ . A referrer of type i has $n_{ii} + n_{ik}$ links and each of them is equally likely to receive the referral. And similarly for type- k links. Therefore, our worker is referred to a job at rate

$$\alpha_{Ri}^j = \frac{\rho n_{ii}^j (1 - u_i)}{n_{ii} + n_{ik}} + \frac{\rho n_{ik}^j (1 - u_k)}{n_{kk} + n_{ki}}.$$

Using network symmetry ($n_{ii}^j = n_{ii}$, $n_{ik}^j = n_{ik}$), consistency ($n_{ik} = n_{ki}$) and the definition of the homophily rate from equation (2) leads to:

$$\alpha_{Ri} = \rho\phi_i(1 - u_i) + \rho(1 - \phi_k)(1 - u_k). \quad (4)$$

It is worth remarking on two features of equation (4). First, the referral rate does not directly depend on the size of the network but, instead, on the homophily rates of the two worker types, ϕ_A and ϕ_B . This feature greatly simplifies the analysis and is the outcome of this model's specific assumptions. Relaxing these assumption is left for future work and is discussed in the Conclusions.

Second, a worker's referral rate increases in the homophily rate of his own type and decreases in the homophily rate of the other type: type-*A* unemployed workers receive share ϕ_A of referrals generated by type-*A* employed workers and share $1 - \phi_B$ of referrals generated by type-*B* employed workers; and similarly for type-*B* unemployed workers. Therefore, network structure (hierarchy or homophily) crucially affects the distributional impact of referral use and it is crucial to have an equilibrium model of the network structure in order to study that impact.

Three types of agents search in the market: measure v vacancies, measure u_A type-*A* unemployed workers and measure u_B type-*B* unemployed workers. A Cobb-Douglas function determines the flow of meetings in the market between vacancies and workers of either type:

$$m(v, u_A, u_B) = \mu v^\eta (u_A + u_B)^{1-\eta},$$

where $\mu > 0$ and $\eta \in (0, 1)$.

The rate at which a vacancy meets with a type i worker through the market is

$$\alpha_{Fi} = \frac{m(v, u_A, u_B)}{v} \frac{u_i}{u_A + u_B} = \mu \left(\frac{u_A + u_B}{v} \right)^{1-\eta} \frac{u_i}{u_A + u_B}.$$

The rate at which a type i worker meets a firm through the market is

$$\alpha_{Mi} = \frac{m(v, u_A, u_B)}{u_A + u_B} = \mu\left(\frac{v}{u_A + u_B}\right)^\eta.$$

Since this rate does not depend on the worker's type, the i -subscript is henceforth dropped.

The steady state conditions equate each type's flows in and out of unemployment:

$$u_A(\alpha_M + \alpha_{RA})p_A = (1 - u_A)\delta, \quad (5)$$

$$u_B(\alpha_M + \alpha_{RB})p_B = (1 - u_B)\delta. \quad (6)$$

I assume that the vacancy cost is low enough for $u_A \leq \frac{1}{2}$ and $u_B \leq \frac{1}{2}$, which is the most relevant case.

2.2.3 Value functions

I now describe the agents' value functions. First, consider a firm. When vacant, it searches in the market, meets with a type- i worker at rate α_{Fi} and forms a match with probability p_i . When producing, the firm's flow payoffs are $y_i + \rho\gamma X_i - w_i$ where w_i denotes the wage. The match is destroyed at rate δ . The firm's values of a vacancy (V) and employing a type- i worker (J_i) satisfy

$$rV = -k + \alpha_{FA}p_A(J_A - V) + \alpha_{FB}p_B(J_B - V),$$

$$rJ_i = y_i + \rho\gamma X_i - w_i + \delta(V - J_i).$$

Consider a worker of type i . When unemployed, his flow utility is b . Job opportunities appear at rate $\alpha_M + \alpha_{Ri}$ and a match is formed with probability p_i . When employed, the worker's flow utility equals his wage and the match is destroyed at rate δ . The worker's value

of unemployment (U_i) and employment (W_i) equal

$$\begin{aligned} rU_i &= b + (\alpha_M + \alpha_{Ri})p_i(W_i - U_i), \\ rW_i &= w_i + \delta(U_i - W_i). \end{aligned}$$

The wage solves the Nash bargaining problem

$$w_i = \operatorname{argmax}_w (W_i - U_i)^\beta (J_i - V)^{1-\beta}. \quad (7)$$

where β denotes the worker's bargaining power.

Finally, the steady state utility of a type i agent is:

$$\Lambda_i = u_i U_i + (1 - u_i) W_i.$$

2.2.4 Off-equilibrium network

I now determine the payoffs to worker j of type i whose network is (n_{ii}^j, n_{ik}^j) and might differ from that of the other type i workers. Worker j is measure zero and therefore his network does not affect any aggregate quantity such as the values of other agents, unemployment rates or vacancy creation.

Worker j 's network affects the referrals that he receives and the referrals that he generates. His arrival rate of job opportunities through a referral is given by:

$$\alpha_{Ri}^j = \frac{n_{ii}^j(1 - u_i)\rho}{n_{ii} + n_{ik}} + \frac{n_{ik}^j(1 - u_k)\rho}{n_{kk} + n_{ki}}.$$

The proportion of time that he spends unemployed is determined by

$$u_i^j(\alpha_M + \alpha_{Ri}^j)p_i = (1 - u_i^j)\delta \Rightarrow u_i^j = \frac{\delta}{\delta + (\alpha_M + \alpha_{Ri}^j)p_i}.$$

I will assume that, when forming his network, a worker does not take into account the differential value that he might have on his employer due to the referrals that he generates: $X_i^j = X_i$. The assumption has a second-order effect on the equilibrium to the extent that the benefits from forming a network arise mostly from the opportunity of receiving referrals rather than from the increased wage that the firm is willing to pay to someone who can generate referrals. This assumption is made for convenience and will tend to underplay linking with type- A workers, who generate higher surplus.

The value of unemployment for worker j and of the firm employing him are given by:

$$\begin{aligned} rU_i^j &= b + (\alpha_M + \alpha_{Ri}^j)p_i(W_i^j - U_i^j) \\ rJ_i^j &= y + \rho\gamma X_i - w_i^j + \delta(V - J_i^j) \end{aligned}$$

Notice that the worker's referral rate directly affects his value of being unemployed and therefore his outside option when bargaining with his employer.

The steady state utility of worker j of type i is:

$$\Lambda_i^j = u_i^j U_i^j + (1 - u_i^j) W_i^j \quad (8)$$

2.3 Definition of Equilibrium

The equilibrium in the labor market for a given network is defined as follows.

Definition 2.1 *A Labor Market Equilibrium given network (n_{AA}, n_{AB}) and (n_{BB}, n_{BA}) is the steady state measures of unemployed workers $\{u_A, u_B\}$ and the measure of vacancies v such that:*

- *The labor market is in steady state as described by (5) and (6).*
- *The surplus is split according to (7).*
- *There is free entry of firms: $V = 0$.*

The equilibrium is defined as follows.

Definition 2.2 *An Equilibrium is (e_{AA}, e_{AB}) and (e_{BB}, e_{BA}) which solve (1) subject to the symmetry restriction where the labor market payoffs are given by equation (8).*

3 Equilibrium and Theoretical Implications

I begin the analysis of the labor market for any symmetric network, solve for the equilibrium network and, finally, derive the model's implications for inequality.

3.1 The labor market

The surplus of a match between a firm and a type- i worker equals:

$$S_i = W_i - U_i + J_i - V.$$

Nash bargaining implies:

$$\begin{aligned} W_i - U_i &= \beta S_i, \\ J_i - V &= (1 - \beta) S_i. \end{aligned}$$

Combine the above with equation (3) and the free entry condition to arrive at:

$$S_i = \frac{y_i - b + \rho\gamma(1 - \beta)(1 - \phi_i)u_k p_k S_k}{r + \delta + (\alpha_M + \alpha_{Ri})p_i\beta - \rho\gamma(1 - \beta)\phi_i u_i p_i}. \quad (9)$$

Notice that S_i depends on S_k due to the possibility that a type- i worker may refer a type- k link in the case of an expansion. If $\phi_i = 1$ then i types only refer workers of the same type and the term multiplying S_k drops out.

The value of a vacancy is given by

$$rV = -k + \alpha_{FA}p_A(1 - \beta)S_A + \alpha_{FB}p_B(1 - \beta)S_B. \quad (10)$$

The first result proves that an equilibrium exists and that a worker's unemployment rate depends only on p_i and does not depend on productivity, which follows immediately from observing the steady state equations (5) and (6).

Proposition 3.1 *A Labor Market Equilibrium exists. The steady state unemployment rate of a type- i worker depends on his probability of matching when meeting a firm (p_i) and does not depend on his productivity on the job (y_i).*

3.2 Equilibrium network

This section characterizes the equilibrium in two steps. First, I show that a worker's choice of network formation effort is unique. Second, I aggregate the optimal choice of all workers to characterize the equilibrium network structure.

Consider the problem of worker j of type i . The first order conditions of the worker's problem with respect to effort on networking with his own type are:¹³

$$\begin{aligned} \frac{d\mathcal{L}_i^j}{de_{ii}^j} &= \frac{\partial \Lambda_i^j}{\partial \alpha_{Ri}^j} \frac{d\alpha_{Ri}^j}{de_{ii}^j} - ce_{ii}^j \\ &= \frac{\partial \Lambda_i^j}{\partial \alpha_{Ri}^j} \frac{\rho(1 - u_i)}{n_{ii} + n_{ik}} - ce_{ii}^j \end{aligned} \quad (11)$$

The first ratio of equation (11) describes how the steady state utility changes with the referral rate while the second ratio describes how effort affects the referral rate.

¹³I only consider the case where $E_{ii} > 0$ and $E_{ik} > 0$. Equilibria without networks exist but are of no particular interest for this study.

Similarly:

$$\frac{d\mathcal{L}_i^j}{de_{ik}^j} = \frac{\partial \Lambda_i^j}{\partial \alpha_{Ri}^j} \frac{\rho(1-u_k)}{n_{kk} + n_{ki}} \frac{E_{ki}}{E_{ki} + E_{ik}} - ce_{ik}^j \quad (12)$$

We have:

Proposition 3.2 *Worker j 's optimal effort $(e_{ii}^{j*}, e_{ik}^{j*})$ in the network formation stage is unique.*

The Proposition proves that steady state utility Λ_i^j is strictly increasing and strictly concave in a worker's referral rate α_{Ri}^j and, therefore, the worker's optimal effort level is unique. The concavity result is intuitive: a worker's steady state utility increases in the rate at which he meets with job opportunities; a high referral rate leads to less time spent unemployed and therefore lower benefit from additional increases in α_{Ri}^j , yielding concavity.

I now state the main result regarding the existence and characterization of Equilibrium.

Proposition 3.3 *An Equilibrium exists. Furthermore:*

1. *The network structure is characterized by one variable:*

- $\phi_A = 1 - \phi_B \equiv \phi^*$ and $\frac{\alpha_{RA}}{\alpha_{RB}} = \frac{\phi^*}{1-\phi^*}$

2. *The network structure depends on relative (un)employment which, in turn, depends on the probability of forming a match:*

- $\left(\frac{\phi^*}{1-\phi^*}\right)^2 = \frac{1-u_A}{1-u_B}$
- $p_A > p_B$ leads to $u_A < u_B$, $\phi^* > \frac{1}{2}$ and $n_{AA} + n_{AB} > n_{BB} + n_{BA}$
- $p_A = p_B$ leads to $u_A = u_B$, $\phi^* = \frac{1}{2}$ and $n_{AA} + n_{AB} = n_{BB} + n_{BA}$

3. *The network structure does not depend on workers' productivity.*

Proposition 3.3 provides a sharp characterization of equilibrium. Part 1 proves that network structure can be reduced to a single variable, ϕ^* , which determines the referral rate of type- A workers, and whose complement $(1 - \phi^*)$ determines the referral rate of type- B workers. Part 2 shows that ϕ^* is increasing in the relative employment of type- A workers which, in turn, is determined by the relative probability of being hired when meeting a firm. Part 3 shows that network structure does not depend on productivity.

To understand these results it is useful to first reiterate the agents' incentives for forming costly links: forming one more link generates additional referrals and the magnitude of this increase depends on the potential link's employment rate. As a result all workers, independently of their own type, prefer to link with high-employment A -types.¹⁴ Consequently type- A workers spend less effort "linking-down" than do B -types "linking-up", which leads to greater rationing of across-type links for B -workers. Furthermore, type- B workers do not respond to this rationing by creating more links with other B -types because such links are less valuable. This explains why B -types have more links with A - than with B -types ($n_{BB} \leq n_{AA}$ or, equivalently, $\phi_B \leq \phi_A$) which leads to a hierarchial network and a (weakly) lower referral rate for B -types than for A -types.

Notice that hierarchy is an outcome of the many-to-many matching structure of the referral network. If, instead, each worker could only create a single connection and the A types decided to only link with other A -types, then the B -types would, by necessity, link with each other, leading to a homophilous network. Burdett and Coles (1999) study a model of marriage, where it is natural to restrict to a single connection, and show that a homophilous (positive assortative) structure arises. This observation suggests that the assumption on whether linking (matching) is 1-1 or many-to-many is crucially important for the resulting matching structure.

Recalling that a potential link's employment rate depends only on p_i and does not depend

¹⁴Notice that this is not a straightforward implications of assuming $p_A \geq p_B$. The result states that there is no equilibrium with a favorable network structure for B -workers leading to $u_B < u_A$.

on y_i , it follows that a link's attractiveness does not depend on productivity which, therefore, does not affect network structure. The distinction between productivity and employment rate becomes extremely relevant when considering the efficiency of equilibrium, in Section 5.

3.3 Inequality and referrals

I now examine the effect of referrals on labor market outcomes. The use of referrals affects the arrival rate of job opportunities differentially for the two worker types, thereby affecting their relative employment rates and, consequently, their relative wages. I compare the referral model's employment outcomes with those of an alternative model which is identical except that all workers have the same network regardless of type (i.e. $\phi_A = \phi_B = \frac{1}{2}$) and, therefore, all workers face the same arrival rate of job opportunities.

This alternative model forms a natural benchmark as it exhibits the same search and informational frictions as the referral model and, therefore, any difference in employment outcomes is due to the equilibrium nature of the referral network. Furthermore, subsuming the arrival rate of job opportunities through the market and referrals in an extended aggregate meeting function, means that the alternative model can be readily related to the standard search and matching model.¹⁵

Denote the arrival rate of job opportunities for a type- i worker in the referral model and the alternative model by α_i and $\hat{\alpha}_i$, respectively. Therefore:

$$\begin{aligned}\hat{\alpha}_A &= \hat{\alpha}_B = \alpha_M + \rho \frac{1}{2} (2 - u_A - u_B) \\ \alpha_A &= \alpha_M + \rho \phi^* (2 - u_A - u_B) \\ \alpha_B &= \alpha_M + \rho (1 - \phi^*) (2 - u_A - u_B)\end{aligned}$$

Proposition 3.3 implies that, when $p_B < p_A$, we have $\alpha_B < \hat{\alpha}_B = \hat{\alpha}_A < \alpha_A$ and the employment rate across types is more unequal in the referral model than in the alternative

¹⁵The functional form of the aggregate meeting function would, of course, differ from the usual ones.

model. The following proposition summarizes the preceding discussion.

Proposition 3.4 *If $p_A > p_B$ then the use of referrals exacerbates inequality.*

Though intuitive, this result is not necessarily immediate from an ex ante point of view. Since type- B workers face worse employment prospects in the labor market, the opportunity to invest in network formation could, in principle, reduce inequality in labor market outcomes. The opposite result arises in equilibrium because low(er) employment rates make B -types less attractive as links which reduces their access to referrals.

3.4 Determinants of the strength of hierarchy

The variable ϕ^* determines the strength of the network's hierarchical structure. If ϕ^* is near $\frac{1}{2}$, then both worker types have similar referral rates and the network is not very hierarchical. If ϕ^* is close to 1, then the A -types enjoy a much higher referral rate and the hierarchical structure is very strong. I now perform two comparative statics exercises to explore the circumstances that lead to stronger or weaker network hierarchy and, by extension, greater or lesser referral-induced inequality.

The strength of the network's hierarchical structure depends on the relative importance of the referral channel in generating job opportunities. If the arrival rate of job opportunities through the market is high, say because of greater vacancy entry or greater efficiency in the meeting process, then referrals play a less important role in job-finding and, therefore, the incentives for creating a hierarchical network are lower. The next proposition states the result.

Proposition 3.5 *Increasing α_M leads to lower ϕ^* in equilibrium.*

As shown in Proposition 3.3, the strength of the network's hierarchy depends on the relative employment rates of the two types which, in turn, depends on the relative probability

that a match is formed when a worker and firm meet. The next Proposition shows that a greater differential in the probability of forming a match leads to a more hierarchical network and, therefore, greater differential in referral rates between A and B -type workers. For simplicity assume that $p_A = 1 - p_B = p$ and abstract from changes in v . The proposition states the result:

Proposition 3.6 *Increasing p leads to higher ϕ^* in equilibrium.*

4 Testable predictions and evidence

This Section compares the model's main predictions on the structure of networks and the interaction between referrals and labor markets with the empirical evidence.

4.1 Hierarchy and homophily in model and data

The equilibrium network structure (hierarchy) differs from the sociological evidence of network homophily (for a survey see McPherson, Smith-Lovin and Cook, 2001) which has also motivated the early economics literature, e.g. Montgomery (1991). This is a key difference and it is important to discuss the conceptual difference between these approaches and to relate them to the available evidence. Three points are worth making.

First, for the purposes of studying the effect of referral networks on labor markets, it is important to distinguish between a social network and a referral network: the former encompasses the set of people that one interacts with socially, which has been extensively studied by sociologists and is homophilous along observable characteristics (McPherson, Smith-Lovin and Cook, 2001); the latter consists of the set of people that one refers to a job which might, in principle, differ from social networks because strategic considerations of future benefits might be salient (see the third point below).

Second, the homophily documented by the sociology literature concerns observable characteristics (after all, it is measured). In one application of this model, worker heterogene-

ity explicitly refers to unobservable differences in productivity (example 1 in Section 2.2). Therefore, there is no a priori contradiction in having a (social or referral) network which is, say, homophilous across (observable) educational attainment and simultaneously hierarchical across (unobservable) ability within a particular level of education.

Although it is inherently difficult to determine network structure along unobservable characteristics, there is some circumstantial evidence to suggest that referral networks are not necessary homophilous across unobserved productivity. It is a common practice among large firms to provide monetary incentives for their employees to refer candidates to apply for a job at the firm (indeed, many of the firm-level studies on referrals are based on data collected by human resources departments for purposes of remunerating referrers). These incentives are available to all workers: none of the studies report that better-than-average workers are somehow favored in such schemes. This observation suggests that firms believe that even their lower-ability employees are able to refer higher-ability applicants to the job and is inconsistent with an extreme version of homophily.

Third, the distinction between a social and referral network is most relevant in the “discrimination” application of this model (example 2 in Section 2.2) where the characteristics of referred workers are, by definition, observable. While there is, again, a paucity of work on correlating referrer-referred types, there is some evidence to suggest that the referral network might be hierarchical even when the social network is typically homophilous. The experimental study of Beaman, Keleher and Madruger (2015) examines the referrer-referred identity across gender. The authors find that women are more likely to refer men than other women for a job (57% of time) even though the sociological evidence points towards gender homophily in social networks (McPherson, Smith-Lovin and Cook, 2001).¹⁶ They also find that men refer other men even more frequently (77% of the time) which is consistent with

¹⁶The study took place in a developing country (Malawi) where gender homophily of social networks is, if anything, likely to be stronger. The job in question was one where a significant proportion of employees were women and therefore referring men is not due to the job characteristics.

a hierarchical referral network.¹⁷ They conclude that referrals tend to reinforce, rather than alleviate, gender inequalities even when the referrers are women, which is consistent with the present paper’s theoretical exploration. Of course, this evidence is far from conclusive but it points to the need for a more nuanced interpretation of the interaction between social networks, referral networks and labor markets.

Gender discrimination is only one dimension to the complicated issue of discrimination. To study racial or ethnic discrimination, the present model would need to be extended, as there is considerable evidence of segregation of economic networks (including referrals) along these lines. While interesting, and important, this is left for future work. See the Conclusions for a discussion.

4.2 Referrals and the labor market

In the model, when a firm meets a worker it is more likely that the worker is type-*A* if the meeting occurs through a referral rather than the market. The intuition is quite straightforward: regardless of own type, each worker is linked with more type-*A* workers than type-*B* workers. Therefore, the recipient of a referral is more likely to be a *A*-type worker regardless of the referrer’s type. The following proposition formalizes this intuition.

Proposition 4.1 *When a firm and a worker meet, it is more likely that the worker is of type *A* if the meeting is through a referral rather than through the market.*

Proposition 4.1 leads to several predictions, which are supported by the empirical literature or offer directions for future work. Overall, they show that the model provides a good positive description of the interaction between referrals and labor markets.

Prediction 1: A match is more likely to be formed if the worker and firm meet through a referral.

¹⁷Interpreting the non-trivial difference in proportions requires a richer environment than the present paper’s.

In their firm-level studies, Fernandez and Weinberg (1997), Castilla (2005), Brown, Setren and Topa (2015) and Burks et al (2015) examine all the applicants for a job (both successful and unsuccessful) and find that referred applicants are more likely to be hired after controlling for their observable characteristics, consistent with Prediction 1.

Prediction 2: A match is more productive in expectation if the worker and firm meet through a referral.

Consistent with Prediction 2, Castilla (2005), Pinkston (2012) and Burks et al (2015) Barr, Bojilov and Munasinghe (2016) find that the referred are more productive after controlling for worker observables (Castilla, Burks et al and Barr et al have direct measures of output by worker while Pinkston has subjective measures, reported by the employer). Hensvik and Skans (2015) use a very detailed Swedish data set to document that referred workers have higher scores in the armed forces ability test, conditional on observables.

Prediction 3: The wage is higher in expectation if the worker and firm meet through a referral.

Consistent with prediction 3, Simon and Warner (1992), Bayer, Ross and Topa (2008), Brown, Setren and Topa (2015), Hensvik and Skans (2015) and Dustmann (2016) find that referred workers receive higher wages than non-referred workers after controlling for worker observables and firm/place of employment fixed effects.^{18 19}

Prediction 4: The wage premium of a referred worker depends on the extent of unobservable heterogeneity.

¹⁸These wage differentials decline with tenure but a significant part persists over time. Simon and Warner (1992), Galenianos (2013), Brown, Setren and Topa (2015) and Dustmann et al (2016) develop models with gradual learning about match quality that can rationalize this pattern. The present paper's interpretation of ex ante worker heterogeneity refers to the persistent part of the wage differential and is therefore complementary to the learning interpretation.

¹⁹Pistaferrri (1999), Pellizzari (2010) and Bentolila, Michelacci and Suarez (2010) find zero or even negative effect of referrals on wages. These studies, however, do not control for firm fixed effects, unlike the studies cited above, which suggests that selection is important on the firm side. See Galenianos (2013) for a model where low-productivity firms use referrals more intensely.

Prediction 4 provides a direction for future research. The literature documents that different types of jobs or different industries use referrals at different rates, as documented in Topa (2011). Galenianos (2014) studies the implications on matching efficiency from heterogeneous referral use. A detailed analysis of the source of this heterogeneity is still missing.

5 Welfare Analysis

The use of referrals exacerbates inequalities in labor markets as it mostly benefits type- A workers, as shown in Section 3.3. This feature, however, is not a priori detrimental for aggregate welfare if it facilitates the employment of more productive workers. This Section examines the conditions under which the use of referrals improves aggregate welfare.

The social planner’s objective is to maximize steady state output subject to the search and informational frictions of the economic environment. The planner has two instruments at his disposal: the network structure and the entry of vacancies. The network structure is described by ϕ_A and ϕ_B . I restrict attention to networks where $\phi_A = 1 - \phi_B$, as this network structure obtains in equilibrium.²⁰ I denote the planner’s solution by ϕ^P and v^P .

In Galenianos (2014) I studied a similar referral model (but with homogeneous workers) and showed that vacancy entry is generically inefficient in equilibrium. The search process is subject to externalities typical of random search models and the efficient level of vacancy entry obtains only when the surplus-sharing parameters (β and γ) take exactly the “right” (non-generic) values. This logic transfers to the present model and, since the current focus is on inequality between the two types of workers rather than vacancy entry, I will focus the analysis on network structure and assume that vacancy creation is constrained efficient. As the results on the optimal network do not qualitatively depend on vacancy entry, this

²⁰Alternatively, the planner could choose ϕ_A and ϕ_B unconstrained. This will make it harder to compare the constrained optimum with the equilibrium outcome and draw conclusions about the source of inefficiency.

assumption is innocuous for the rest of the welfare analysis.

5.1 Optimal network

The planner solves the following problem:

$$\begin{aligned} \max_{\phi} \mathcal{W}(\phi) &= (1 - u_A)y_A + bu_A + (1 - u_B)y_B + u_Bb \\ \text{s.t. } (1 - u_A)\delta &= u_{AP}(\alpha_M + \alpha_{RA}) \\ \text{and } (1 - u_B)\delta &= u_{BP}(\alpha_M + \alpha_{RB}) \end{aligned}$$

where $\phi = \phi_A = 1 - \phi_B$ determines α_{RA} and α_{RB} . Denote the planner's solution by ϕ^P .

The proposition provides the main efficiency result.

Proposition 5.1 *The equilibrium network is generically inefficient: $\phi^P \neq \phi^*$.*

To explore the source of the inefficiency, it is useful to examine the necessary condition for optimality from the planner's problem:

$$-u'_A(\phi^P)(y_A - b) = u'_B(\phi^P)(y_B - b) \quad (13)$$

The steady state conditions determine $u'_i(\phi)$ and $u'_A(\phi) < 0$ and $u'_B(\phi) > 0$ since a higher ϕ increases the referral rate of A -types and reduces the referral rate of B -types.

The planner chooses ϕ^P to maximize the productivity-weighted employment rate of the two worker types and, therefore, the optimal network depends on workers' productivity. Recall that the equilibrium network does not share this feature as ϕ^* does not depend on productivity levels (Proposition 3.3) which is an important source of the equilibrium network's inefficiency.

The next proposition characterizes the optimal network in more detail and compares it with the equilibrium network.

Proposition 5.2 *The planner's solution has the following features:*

1. *If $y_A = y_B$ and $p_A = p_B$ then the equilibrium network is efficient: $\phi^P = \phi^* = \frac{1}{2}$.*
2. *If $y_A = y_B$ and $p_A > p_B$ then $\phi^P < \frac{1}{2} < \phi^*$.*
3. *Increasing y_A or reducing y_B leads to higher ϕ^P (if $\phi^P \in (0, 1)$) and does not affect ϕ^* .*

Part 1 of the Proposition is intuitive: if workers are identical, then the planner would like to maximize aggregate employment. Since the employment rate is strictly concave in the arrival rate of job opportunities, maximum employment requires a symmetric network. Inspecting equation (13) demonstrates that this is the unique solution.

Part 2 describes a major source of divergence between the planner and equilibrium solutions. If productivity is equal across types (or, more generally, productivity differentials are small) then the planner wants, again, to maximize employment. Since type- B workers form a match with lower probability than type- A workers, the planner compensates by increasing their referral rate ($\phi^P < \frac{1}{2}$) which reduces inequality in comparison to a symmetric network (see Section 3.3). In this case, however, the planner's optimal network goes in exactly the opposite direction from the decentralized solution: the equilibrium features a network that exacerbates inequality exactly because from the workers' point of view it is only the probability of getting a job that matters.

Finally, part 3 shows that in some settings the planner does prefer to increase inequality. When productivity differences are substantial, then the planner prefers to increase the employment rate of the A -types even at the expense of the B -types' employment. Therefore, the optimal network features greater hierarchy in favor of the type- A workers as productivity differentials increase. Since the equilibrium network does not depend on workers' productivity, this again creates a wedge between the optimal and the equilibrium solution.

Underlying the inefficiency result is a tension between incentives in link formation and the optimal recipient of referrals. Individual incentives inevitably lead to a network that

“favors the favored”: everyone wants to link with high-employment workers, who end up dominating the network. However, the optimal referral recipient need not be the type-As, as shown above, which is responsible for the inefficiency.

5.2 Discussion and policy implications

The model’s generality allows the welfare analysis to encompass very different economic environments, summarized in a stylized way in the two examples in Section 2.2, and to demonstrate that this generality is useful in identifying when referrals help or hinder welfare. I now review the implications of the welfare analysis.

When worker heterogeneity is driven by productivity differentials (as in example 1 from Section 2.2) the use of referrals improves welfare despite the increased inequality, though the constrained optimum is not generically reached. In such a case, the use of referrals has mostly benign effects as it reduces search and informational frictions.

When workers have similar productivity on the job but nevertheless face different probability of being hired, e.g. when discrimination is present, the use of referrals reduces welfare: the planner’s optimal network structure would reduce inequality ($\phi^P < \frac{1}{2}$) while the equilibrium network structure increases inequality ($\phi^* > \frac{1}{2}$). Example 2 from Section 2.2 presents a set of environments where this scenario occurs in equilibrium and which are consistent with discrimination in the labor market.

Using the model to determine which of the two examples best describes the actual data is much harder to do. The evidence that referred workers have higher probability of being hired and receiving higher wages are consistent with both examples (lower wages for the disadvantaged group are due to lower employment prospects, rather than lower productivity). Therefore, at the very least actual productivity measures are needed.

Castilla (2005), Burks et al (2015) and Barr, Bojilov and Munasinghe (2016) provide such data for their firm-level studies and show that referrals are mostly associated with

higher productivity. In a more detailed data-set, Hensvik and Skans (2015) have objective ability measures which suggest that referrals are, on average, given on merit at least for the population under study: men who completed the draft in Sweden. Whether this generalizes to women or minority populations is very much an open question.

Two recent papers suggest that discrimination against women (whether deliberate or subconscious) is present, even in skilled professions. Goldin and Rouse (2000) find that making symphony orchestra auditions “blind” by introducing a screen significantly increased the probability that a woman advances and is hired. Moss-Racusin et al (2012) gave identical CVs with randomized male or female names to be evaluated by science faculty for a laboratory manager position and find that CVs with male names are rated to be consistently more competent. Although these works are about very specific professions, they suggest that gender discrimination is present and, according to the model, in such cases the use of referrals has a very detrimental effect on welfare by magnifying that inequality.

The model suggests a number of possible actions to improve welfare in the discrimination case. The most direct action is to follow policies that improve the employment opportunities of the disadvantaged group, such as offering incentives for hiring B -workers or, as in the orchestra example, finding ways to reduce subconscious biases. Such a policy improves the employment opportunities for B -types which creates a positive feedback effect by also improving their network.

A less direct but still beneficial policy is to increase the importance of formal hiring methods by encouraging the creation of vacancies (increasing v) or improving the functioning of the market (increasing market matching efficiency, μ). Such policies might, of course, be desirable on their own right but they also have the effect of reducing the relative importance of hiring through referrals (see Proposition 3.5) and might therefore reduce the inefficiency created by referrals’ exacerbation of inequalities.

6 Conclusions

This is the first paper to study the interaction between labor markets, endogenous referral networks and worker heterogeneity. The model provides a good qualitative description of the effects of referral use: consistent with empirical evidence, referred workers are more likely to be hired, to receive higher wages and to be more productive. Furthermore, the use of referrals exacerbates inequality among workers.

The welfare analysis encompasses a great variety of economic environments, leading to subtle implications. In the case where worker heterogeneity is mainly due to productivity differentials, the use of referrals improves welfare. If, however, productivity is similar across types and, nevertheless, one worker type forms matches with higher probability, then the equilibrium network structure is qualitatively the opposite from the optimal structure, leading to severe inefficiencies. The latter case is likely to occur in the presence of discrimination in the labor market.

This paper provides an initial step in the study of interactions between labor markets, endogenous networks and heterogeneity and there are many possible directions for further work. A first direction is to study the evolution of a worker's referral network in relation to his employment status. Requiring that a worker's network needs constant effort to sustain (say, because it depreciates over time) and making that effort more costly when a worker is unemployed is a natural framework where the issue of unemployment duration dependence due to the long-term unemployed's loss of contact with the labor market can be studied.

A second direction is to examine the interaction of referrals with discrimination in more detail. An important assumption in the present model is that the two types of workers are socially integrated in the sense that they can easily be included in each other's referral networks, if they so wish. Under this assumption, the model can be used to examine gender discrimination, since men and women interact socially regardless of whether gender discrimination is present in the labor market. To study racial or ethnic discrimination, which are

typically combined with the affected groups' social segregation, the model can be extended by allowing for heterogeneous costs for linking across types. Increasing such a cost can lead to greater homophily across types. Furthermore, in that case the arrival rate of job opportunities through the market might also differ by type, creating a richer interaction between markets and referrals.

Additionally, network size does not currently play a crucial role but one might expect that a denser network might lead to more information transmission, thereby extending the advantage of well-networked type-*A* workers. Furthermore, a model where type-*A* workers refer more often might create a force for sorting of types across firms, which has not been explored in the literature.

APPENDIX

A Proofs

Proposition 3.1.

Proof. A worker's unemployment rate is determined by the steady state equation. These equations include a worker's probability of forming a match when meeting a firm (p_i) and do not include the worker's productivity (y_i).

Define:

$$\begin{aligned}
 H(u_A, u_B) &\equiv u_A \mu \left(\frac{v}{u_A + u_B} \right)^\eta + u_A \rho (\phi_A (1 - u_A) + (1 - \phi_B) (1 - u_B)) - \frac{\delta}{p_A} (1 - u_A) \\
 L(u_A, u_B) &\equiv u_B \mu \left(\frac{v}{u_A + u_B} \right)^\eta + u_B \rho (\phi_B (1 - u_B) + (1 - \phi_A) (1 - u_A)) - \frac{\delta}{p_B} (1 - u_B)
 \end{aligned}$$

and note that in a steady state $H(u_A, u_B) = L(u_A, u_B) = 0$ holds. Define $h(u_B)$ and $l(u_A)$ such that $H(h(u_B), u_B) = 0$ and $L(u_A, l(u_A)) = 0$.

Let $H_x(u_A, u_B) \equiv \partial H(u_A, u_B) / \partial x$ and similarly for $L(u_A, u_B)$ and observe that

$$\begin{aligned}
 H(0, u_B) &= -\frac{\delta}{p_A} < 0 \\
 H(1, u_B) &= \mu \left(\frac{v}{1 + u_B} \right)^\eta + \rho (1 - \phi_B) (1 - u_B) > 0 \\
 H_{u_A}(u_A, u_B) &= \mu \left(\frac{v}{u_A + u_B} \right)^\eta \left(1 - \frac{\eta u_A}{u_A + u_B} \right) + \rho (\phi_A (1 - u_A) + (1 - \phi_B) (1 - u_B)) \\
 &\quad + \frac{\delta}{p_A} - u_A \rho \phi_A > 0 \\
 H_{u_B}(v, u_A, u_B) &= -\frac{\eta u_A}{u_A + u_B} \mu \left(\frac{v}{u_A + u_B} \right)^\eta - u_A \rho (1 - \phi_B) < 0 \\
 h'(u_B) &= -\frac{H_{u_B}}{H_{u_A}} > 0
 \end{aligned}$$

The first two equations mean that $h(u_B)$ exists and is in $(0, 1)$ for any $u_B \in [0, 1]$. The third equation proves that $h(u_B)$ is a function (not a correspondence) and the fourth equation proves that $h'(u_B) > 0$. And similarly for $l(u_A)$.

Define $T_1(u_A) = h(l(u_A))$. A steady state is a fixed point of $T_1(u_A)$. The results above prove that $T_1(0) > 0$ and $T_1(1) < 0$ and therefore a fixed point exists. To prove the fixed point is unique it suffices to show that $T_1'(u_A) < 1$. We have:

$$T_1'(u_A) = h'(l(u_A))l'(u_A) = \frac{H_{u_B}(u_A, l(u_A))L_{u_A}(u_A, l(u_A))}{H_{u_A}(u_A, l(u_A))L_{u_B}(u_A, l(u_A))}$$

Notice that:

$$\begin{aligned} H_{u_A}(u_A, u_B) + L_{u_A}(u_A, u_B) &= (1 - \eta)\alpha_M + \alpha_{RA} + \frac{\delta}{p_A} - \rho(\phi_A u_A + (1 - \phi_A)u_B) > 0 \\ L_{u_B}(u_A, u_B) + H_{u_B}(u_A, u_B) &= (1 - \eta)\alpha_M + \alpha_{RB} + \frac{\delta}{p_B} - \rho(\phi_B u_B + (1 - \phi_B)u_A) > 0 \end{aligned}$$

and therefore $T_1'(u_A) < 1$ and the steady state is uniquely defined.

Note that:

$$\begin{aligned} H_v &= \frac{\eta u_A}{v} \mu\left(\frac{v}{u_A + u_B}\right)^\eta > 0 \\ L_v &= \frac{\eta u_B}{v} \mu\left(\frac{v}{u_A + u_B}\right)^\eta > 0 \end{aligned}$$

which imply that $\frac{dh(u_B)}{dv} < 0$ and $\frac{dl(u_A)}{dv} < 0$. Therefore u_A and u_B are decreasing in v .

The value of a vacancy is given by:

$$rV = \alpha_{FA}(1 - \beta)S_A + \alpha_{FB}(1 - \beta)S_B \tag{14}$$

where the S_i 's can be written as follows:

$$\begin{aligned} S_A &= \frac{(y_A - b)D_{L1} + (y_B - b)D_{H2}}{D_{H1}D_{L1} - D_{H2}D_{L2}}, \\ S_B &= \frac{(y_B - b)D_{H1} + (y_A - b)D_{L2}}{D_{H1}D_{L1} - D_{L2}D_{H2}} \end{aligned}$$

where

$$\begin{aligned} D_{i1} &= r + \delta + (\alpha_M + \alpha_{Ri})p_i\beta - \rho\gamma(1 - \beta)\phi_i u_i p_i, \\ D_{i2} &= \rho\gamma(1 - \beta)(1 - \phi_i)u_k p_k. \end{aligned}$$

Note that D_{i1} is increasing in v and D_{i2} is decreasing in v and therefore S_i is decreasing in v .

Furthermore the steady state equations imply that

$$\begin{aligned} v \rightarrow 0 &\Rightarrow (u_A, u_B) \rightarrow (1, 1) \Rightarrow \alpha_{Fi} \rightarrow \infty, \\ v \rightarrow \infty &\Rightarrow (u_A, u_B) \rightarrow (0, 0) \Rightarrow \alpha_{Fi} \rightarrow 0. \end{aligned}$$

These observations mean that a vacancy's value is above k for v near zero and below k for v very large and, therefore, a labor market equilibrium exists. ■

Proposition 3.2.

Proof. The second and cross-derivatives of the worker's objective function are:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}_i^j}{\partial (e_{ii}^j)^2} &= \frac{\partial^2 \Lambda_i^j}{\partial (\alpha_{Ri}^j)^2} \left(\frac{\rho(1 - u_i)}{n_{ii} + n_{ik}} \right)^2 - c \\ \frac{\partial^2 \mathcal{L}_i^j}{\partial (e_{ik}^j)^2} &= \frac{\partial^2 \Lambda_i^j}{\partial (\alpha_{Ri}^j)^2} \left(\frac{\rho(1 - u_k)}{n_{kk} + n_{ki}} \frac{E_{ki}}{E_{ki} + E_{ik}} \right)^2 - c \\ \frac{\partial^2 \mathcal{L}_i^j}{\partial (e_{ii}^j e_{ik}^j)} &= \frac{\partial^2 \Lambda_i^j}{\partial (\alpha_{Ri}^j)^2} \frac{\rho(1 - u_i)}{n_{ii} + n_{ik}} \frac{\rho(1 - u_k)}{n_{kk} + n_{ki}} \frac{E_{ki}}{E_{ki} + E_{ik}} \end{aligned}$$

Proving that $(\partial^2 \Lambda_i^j)/(\partial (\alpha_{Ri}^j)^2)$ is negative suffices to show that the maximization problem has a unique equilibrium.

Recall that worker j 's unemployment rate and match surplus are defined by:

$$\begin{aligned} u_i^j &= \frac{\delta}{\delta + (\alpha_M + \alpha_{Ri}^j)p_i} \\ S_i^j &= \frac{y_i - b + \rho\gamma X_i}{r + \delta + (\alpha_M + \alpha_{Ri}^j)p_i\beta} \end{aligned}$$

Rewrite worker j 's steady state utility as follows:

$$\begin{aligned}
\Lambda_i^j &= u_i^j U_i^j + (1 - u_i^j) W_i^j \\
&= (1 - u_i^j)(W_i^j - U_i^j) + \frac{1}{r} \left(b + (\alpha_M + \alpha_{Ri}^j) p_i (W_i^j - U_i^j) \right) \\
&= \frac{b}{r} + \left((1 - u_i^j) + \frac{(\alpha_M + \alpha_{Ri}^j) p_i}{r} \right) \beta S_i^j
\end{aligned}$$

Letting

$$\alpha_i^j = (\alpha_M + \alpha_{Ri}^j) p_i$$

and going through some algebra leads to:

$$\Lambda_i^j = \frac{b}{r} + \frac{\beta}{r} (y_i - b + \rho \gamma X_i) \frac{\alpha_i^j (r + \delta) + (\alpha_i^j)^2}{\delta(r + \delta) + \alpha_i^j (r + \delta + \delta\beta) + (\alpha_i^j)^2 \beta}$$

Differentiating Λ_i^j with respect to α_{Ri}^j we have:

$$\begin{aligned}
\frac{\partial \Lambda_i^j}{\partial \alpha_{Ri}^j} &= \frac{\beta (y_i - b + \rho \gamma X_i) p_i}{r (\delta(r + \delta) + \alpha_i^j (r + \delta + \delta\beta) + (\alpha_i^j)^2 \beta)^2} \left[(r + \delta + 2\alpha_i^j) (\delta(r + \delta) + \alpha_i^j (r + \delta + \delta\beta) + (\alpha_i^j)^2 \beta) \right. \\
&\quad \left. - (\alpha_i^j (r + \delta) + (\alpha_i^j)^2) (r + \delta + \beta\delta + 2\alpha_i^j \beta) \right] \\
&= \frac{\beta (y_i - b + \rho \gamma X_i) p_i}{r (\delta(r + \delta) + \alpha_i^j (r + \delta + \delta\beta) + (\alpha_i^j)^2 \beta)^2} \left[(r + \delta)^2 \delta + 2\alpha_i^j \delta (r + \delta) + (\alpha_i^j)^2 (r(1 - \beta) + \delta) \right] > 0
\end{aligned}$$

Define:

$$D = \delta(r + \delta) + \alpha_i (r + \delta + \delta\beta) + (\alpha_i^j)^2 \beta$$

Taking the second derivative of Λ_i^j with respect to α_{Ri}^j leads to:

$$\begin{aligned}
\frac{\partial^2 \Lambda_i^j}{\partial (\alpha_{Ri}^j)^2} &= \frac{\beta(y_i - b + \rho\gamma X_i)p_i}{rD^3} \left[(2\delta(r + \delta) + 2\alpha_i(r(1 - \beta) + \delta))(\delta(r + \delta) + \alpha_i(r + \delta + \delta\beta) + \alpha_i^2\beta)^2 \right. \\
&\quad \left. - ((r + \delta)^2\delta + \alpha_i 2\delta(r + \delta) + \alpha_i^2(r(1 - \beta) + \delta))2D_i((r + \delta + \delta\beta) + \alpha_i 2\beta) \right] \\
&= \frac{2\beta(y_i - b + \rho\gamma X_i)p_i}{rD^2} \left[-\delta(r + \delta)^2(r + \beta\delta) - \alpha_i^j \delta(r + \delta)(\beta(r + \delta) + (r + \delta)2\beta) \right. \\
&\quad \left. - (\alpha_i^j)^2 3\delta\beta(r + \delta) - (\alpha_i^j)^3(r(1 - \beta) + \delta)\beta \right] < 0
\end{aligned}$$

Therefore, an agent's steady state utility is a strictly increasing and strictly concave function of his job finding rate.

As a result, the worker's optimal effort level is given by the first order conditions of his optimization problem. ■

Proposition 3.3.

Proof. Using the definition for how effort leads into link creation:

$$\frac{e_{ii}^j}{e_{ik}^j} = \frac{E_{ik} + E_{ki}}{E_{ki}} \frac{n_{ii}^j}{n_{ik}^j} \tag{15}$$

Equate the first order conditions with respect to e_{ii}^j and e_{ik}^j with zero, rearrange and take their ratio:

$$\frac{e_{ii}^j}{e_{ik}^j} = \frac{1 - u_i}{1 - u_k} \frac{n_{ki} + n_{kk}}{n_{ik} + n_{ii}} \frac{E_{ik} + E_{ki}}{E_{ki}} \tag{16}$$

Combining equations (15) and (16):

$$\begin{aligned}
\frac{E_{ik} + E_{ki}}{E_{ki}} \frac{n_{ii}^j}{n_{ik}^j} &= \frac{1 - u_i}{1 - u_k} \frac{n_{ki} + n_{kk}}{n_{ik} + n_{ii}} \frac{E_{ik} + E_{ki}}{E_{ki}} \\
\Rightarrow \frac{n_{ii}^j}{n_{ik}^j} &= \frac{1 - u_i}{1 - u_k} \frac{n_{ki} + n_{kk}}{n_{ik} + n_{ii}}
\end{aligned}$$

Recall that:

$$\begin{aligned}
n_{ii}^j &= \phi_i^j (n_{ii}^j + n_{ik}^j) \\
n_{ik}^j &= (1 - \phi_i^j) (n_{ii}^j + n_{ik}^j) \\
\Rightarrow \frac{n_{ii}^j}{n_{ik}^j} &= \frac{\phi_i^j}{1 - \phi_i^j}
\end{aligned}$$

Combining the above together with the equilibrium symmetry condition $\phi_i^j = \phi_i$ for all j we have:

$$\begin{aligned}
\frac{\phi_A}{1 - \phi_A} &= \frac{1 - u_A}{1 - u_B} \frac{n_{BB} + n_{BA}}{n_{AA} + n_{AB}} \\
\frac{1 - \phi_B}{\phi_B} &= \frac{1 - u_A}{1 - u_B} \frac{n_{BB} + n_{BA}}{n_{AA} + n_{AB}}
\end{aligned} \tag{17}$$

Consistency requires $n_{AB} = n_{BA}$ which implies that:

$$\begin{aligned}
(n_{AA} + n_{AB})(1 - \phi_A) &= (n_{BB} + n_{BA})(1 - \phi_B) \\
\Rightarrow \frac{n_{BB} + n_{BA}}{n_{AA} + n_{AB}} &= \frac{1 - \phi_A}{1 - \phi_B} = \frac{1 - \phi_A}{\phi_A}
\end{aligned} \tag{18}$$

Combining equations (17) and (18) proves that $(\frac{\phi_A}{1 - \phi_B})^2 = \frac{1 - u_A}{1 - u_B}$.

The equilibrium is characterized by (v, u_A, u_B, ϕ) , where $\phi = \phi_A = 1 - \phi_B$, which satisfy $(\frac{\phi_A}{1 - \phi_B})^2 = \frac{1 - u_A}{1 - u_B}$, the steady state and free entry conditions. Given v (e.g. at its equilibrium value) this is determined by the root of the following equation:

$$T_2(\phi) = \phi^2(1 - u_B(\phi)) - (1 - \phi)^2(1 - u_A(\phi))$$

where $u_A(\phi)$ and $u_B(\phi)$ are defined by the steady state conditions $H(u_A, u_B, \phi) = L(u_A, u_B, \phi) = 0$ (with a slight abuse of notation).

I show that $T_2(\phi) = 0$ has a unique solution at ϕ^* . Note that:

$$T_2(0) = -(1 - u_A(0)) < 0$$

$$T_2(1) = 1 - u_B(1) > 0$$

$$\begin{aligned} T_2'(\phi) &= 2\phi(1 - u_B(\phi)) + 2(1 - \phi)(1 - u_A(\phi)) - \phi^2 u_B'(\phi) + (1 - \phi)^2 u_A'(\phi) \\ &= \frac{(1 - \phi)^2}{\phi} \left(2 \frac{\phi^2 (1 - u_B(\phi))}{(1 - \phi)^2} + \phi u_A'(\phi) \right) + \frac{\phi^2}{1 - \phi} \left(2 \frac{(1 - \phi)^2 (1 - u_A(\phi))}{\phi^2} - (1 - \phi) u_B'(\phi) \right) \end{aligned}$$

It suffices to show that $T_2'(\phi) > 0$ when $T(\phi) = 0$. Using $T(\phi) = 0$ we can rewrite:

$$T_2'(\phi) = \frac{(1 - \phi)^2}{\phi} \left[2(1 - u_A(\phi)) + \phi u_A'(\phi) \right] + \frac{\phi^2}{1 - \phi} \left[2(1 - u_B(\phi)) - (1 - \phi) u_B'(\phi) \right] \quad (19)$$

Using implicit differentiation we have:

$$\begin{aligned} u_A'(\phi) &= -\frac{L_{u_B} H_\phi - H_{u_B} L_\phi}{H_{u_A} L_{u_B} - H_{u_B} L_{u_A}} \\ u_B'(\phi) &= -\frac{H_{u_A} L_\phi - L_{u_A} H_\phi}{H_{u_A} L_{u_B} - H_{u_B} L_{u_A}} \end{aligned}$$

We examine the square brackets in equation (19) separately:

$$2(1 - u_A) + \phi u_A' = \frac{1}{\Delta} \left(L_{u_B} (2(1 - u_A) H_{u_A} - u_A \alpha_{RA}) - H_{u_B} (2(1 - u_A) L_{u_A} + \alpha_{RA} u_B) \right)$$

where $\Delta = H_{u_A} L_{u_B} - H_{u_B} L_{u_A} > 0$.

Recalling that $L_{u_B} + H_{u_B} > 0$ it suffices to show that:

$$\begin{aligned} &2(1 - u_A) H_{u_A} - u_A \alpha_{RA} + 2(1 - u_A) L_{u_A} + \alpha_{RA} u_B > 0 \\ \Rightarrow &\alpha_M 2(1 - u_A) (1 - \eta) + \alpha_{RA} (2(1 - u_A) - u_A + u_B) + 2(1 - u_A) \left(\frac{\delta}{p_A} - \rho \phi u_A - \rho (1 - \phi) u_B \right) > 0 \end{aligned}$$

which is positive when $u_A \leq \frac{1}{2}$.

Similarly:

$$2(1 - u_B) - (1 - \phi) u_B' = \frac{1}{\Delta} \left(H_{u_A} (2(1 - u_B) L_{u_B} - u_B \alpha_{RB}) - L_{u_A} (2(1 - u_B) H_{u_B} + \alpha_{RB} u_A) \right)$$

which is positive because $H_{u_A} + L_{u_A} > 0$ and

$$\alpha_M 2(1 - u_B)(1 - \eta) + \alpha_{RB} (2(1 - u_B) - u_B + u_A) + 2(1 - u_B) \left(\frac{\delta}{p_B} - u_B \rho (1 - \phi) - u_A \rho \phi \right) > 0$$

which is positive if $u_B \leq \frac{1}{2}$. Therefore, given v , (ϕ, u_A, u_B) are unique.

Furthermore notice that:

$$T_2\left(\frac{1}{2}\right) = \frac{1}{4}(u_A - u_B)$$

When $\phi = \frac{1}{2}$ we have $\alpha_{RA} = \alpha_{RB}$ and therefore $u_A < u_B \Leftrightarrow p_A > p_B$. Therefore:

$$\begin{aligned} p_A = p_B &\Rightarrow u_A = u_B \Rightarrow T_2\left(\frac{1}{2}\right) = 0 \Rightarrow \phi^* = \frac{1}{2} \\ p_A > p_B &\Rightarrow u_A < u_B \Rightarrow T_2\left(\frac{1}{2}\right) < 0 \Rightarrow \phi^* > \frac{1}{2} \end{aligned}$$

This completes the proof. ■

Proposition 3.6.

Proof. Notice that

$$\frac{\partial T_2(\phi)}{\partial p} = \frac{(1 - \phi)^2}{1 - u_B} \left((1 - u_B) \frac{du_A}{dp} - (1 - u_A) \frac{du_B}{dp} \right)$$

and also $\frac{\partial T_2(\phi)}{\partial p} < 0 \Leftrightarrow \frac{d\phi^*}{dp} > 0$.

We have:

$$\begin{aligned} \frac{du_A}{dp} &= -\frac{L_{u_B} H_p - H_{u_B} L_p}{\Delta} \\ \frac{du_B}{dp} &= -\frac{H_{u_A} L_p - L_{u_A} H_p}{\Delta} \end{aligned}$$

where $\Delta = H_{u_A} L_{u_B} - H_{u_B} L_{u_A}$ and

$$\begin{aligned} H_p &= \frac{\delta(1 - u_A)}{p^2} \\ L_p &= -\frac{\delta(1 - u_B)}{(1 - p)^2} \end{aligned}$$

Therefore:

$$\begin{aligned}
\frac{\partial T(\phi)}{\partial p} &= -\frac{(1-\phi)^2}{\Delta(1-u_B)} \left[(1-u_B)(L_{u_B}H_p - H_{u_B}L_p) - (1-u_A)(H_{u_A}L_p - L_{u_A}H_p) \right] \\
&= -\frac{(1-\phi)^2}{\Delta(1-u_B)} \left[\frac{\delta(1-u_B)}{(1-p)^2} \left(\alpha_M \left(1 - \frac{\eta u_A(2-u_A-u_B)}{u_A+u_B} \right) + \rho\phi(2-u_A-u_B)(1-u_A) \right) \right. \\
&\quad \left. + \frac{\delta(1-u_A)}{p^2} \left(\alpha_M \left(1 - \frac{\eta u_B(2-u_A-u_B)}{u_A+u_B} \right) + \rho(1-\phi)(2-u_A-u_B)(1-u_B) \right) \right]
\end{aligned}$$

The terms multiplied by ρ are positive. Furthermore:

$$\begin{aligned}
&\frac{1-u_B}{(1-p)^2} \left(1 - \frac{\eta u_A(2-u_A-u_B)}{u_A+u_B} \right) + \frac{1-u_A}{p^2} \left(1 - \frac{\eta u_B(2-u_A-u_B)}{u_A+u_B} \right) > \\
&(1-u_B) \left(1 - \frac{\eta u_A(2-u_A-u_B)}{u_A+u_B} \right) + (1-u_A) \left(1 - \frac{\eta u_B(2-u_A-u_B)}{u_A+u_B} \right) = \\
&(2-u_A-u_B) \left(1 - \frac{\eta u_A(1-u_B)}{u_A+u_B} - \frac{\eta u_B(1-u_A)}{u_A+u_B} \right) > \\
&(2-u_A-u_B)(1-\eta) > 0
\end{aligned}$$

and therefore $\frac{\partial T_2(\phi)}{\partial p} < 0$ and $\frac{d\phi^*}{dp} > 0$. ■

Proposition 3.5.

Proof. Notice that

$$\frac{\partial T_2(\phi)}{\partial \alpha_M} = -\phi^2 \frac{du_B}{d\alpha_M} + (1-\phi)^2 \frac{du_A}{d\alpha_M}$$

If the sign is positive then $\frac{d\phi^*}{d\alpha_M} < 0$ and vice versa.

Recall that $\frac{\phi^2}{(1-\phi)^2} = \frac{1-u_A}{1-u_B}$ when $T_2(\phi) = 0$ and therefore:

$$\frac{\partial T_2(\phi)}{\partial \alpha_M} = \frac{(1-\phi)^2}{1-u_B} \left((1-u_B) \frac{du_A}{d\alpha_M} - (1-u_A) \frac{du_B}{d\alpha_M} \right)$$

We have:

$$\begin{aligned}
\frac{du_A}{d\alpha_M} &= -\frac{L_{u_B}H_{\alpha_M} - H_{u_B}L_{\alpha_M}}{\Delta} \\
\frac{du_B}{d\alpha_M} &= -\frac{H_{u_A}L_{\alpha_M} - L_{u_A}H_{\alpha_M}}{\Delta}
\end{aligned}$$

where $\Delta = H_{u_A}L_{u_B} - H_{u_B}L_{u_A}$ and

$$H_{\alpha_M} = u_A$$

$$L_{\alpha_M} = u_B$$

Therefore:

$$\begin{aligned}\frac{du_A}{d\alpha_M} &= -\frac{1}{\Delta} \left(\frac{\delta u_A}{p_B u_B} + u_B u_A \rho (2\phi - 1) \right) \\ \frac{du_B}{d\alpha_M} &= -\frac{1}{\Delta} \left(\frac{\delta u_B}{p_A u_A} - u_B u_A \rho (2\phi - 1) \right)\end{aligned}$$

Hence

$$\begin{aligned}\frac{\partial T_2(\phi)}{\partial \alpha_M} &= -\frac{(1-\phi)^2}{\Delta(1-u_B)} \left[(1-u_B) \left(\frac{\delta u_A}{p_B u_B} + u_B u_A \rho (2\phi - 1) \right) - (1-u_A) \left(\frac{\delta u_B}{p_A u_A} - u_B u_A \rho (2\phi - 1) \right) \right] \\ &= -\frac{(1-\phi)^2}{\Delta(1-u_B)} \left[\alpha_M (u_A - u_B) + \rho (2 - u_A - u_B) (u_A (1-\phi)(1-u_B) - u_B \phi (1-u_A)) \right] > 0\end{aligned}$$

and we have $\frac{d\phi^*}{d\alpha_M} < 0$. ■

Proposition 4.1.

Proof. In a meeting through the market the probability that the worker is of type A is:

$$P[A|\text{market}] = \frac{u_A}{u_A + u_B}$$

Denote the rate that referral are generated from A -type workers and are received by A - and B -type unemployed workers by R_{AA} and R_{AB} , respectively. We have:

$$R_{AA} = \rho(1-u_A)\phi_A u_A$$

$$R_{AB} = \rho(1-u_A)(1-\phi_A)u_B$$

The rate at which referrals are generated from type- B workers is defined by R_{BA} and R_{BB} in an equivalent way.

In a meeting through referrals, the probability that the referred worker is of a type A is:

$$\begin{aligned} P[A|\text{referral}] &= \frac{R_{AA} + R_{BA}}{R_{AA} + R_{AB} + R_{BA} + R_{BB}} \\ &= \frac{[(1 - u_A)\phi_A + (1 - u_B)(1 - \phi_B)]u_A}{[(1 - u_A)\phi_A + (1 - u_B)(1 - \phi_B)]u_A + [(1 - u_A)(1 - \phi_A) + (1 - u_B)\phi_B]u_B} \end{aligned}$$

Noting that

$$(1 - u_A)\phi_A + (1 - u_B)(1 - \phi_B) \geq (1 - u_A)(1 - \phi_A) + (1 - u_B)\phi_B$$

proves that $P[A|\text{referral}] > P[A|\text{market}]$. ■

Proposition 5.1.

Proof. The problem becomes:

$$\max_{\phi} \mathcal{W}(\phi) = y_A + y_B - u_A(\phi)(y_A - b) - u_B(\phi)(y_B - b)$$

where $u_A(\phi)$ and $u_B(\phi)$ are defined by the steady state conditions.

Differentiating and going through some algebra ($\Delta = H_{u_A}L_{u_B} - H_{u_B}L_{u_A}$):

$$\begin{aligned} \mathcal{W}'(\phi) &= -u'_A(\phi)\bar{y}_A - u'_B(\phi)\bar{y}_B \\ &= -\frac{\rho(2 - u_A - u_B)u_A u_B}{\Delta} \left[\frac{(y_A - b)\delta}{p_B u_B^2} - \frac{(y_B - b)\delta}{p_A u_A^2} - (\bar{y}_A - \bar{y}_B) \left(\frac{2\eta\alpha_M}{u_A + u_B} + \rho \right) \right] \end{aligned}$$

Let $T_3(\phi)$ denote the term inside the square brackets and note that the planner's solution is given by $T_3(\phi^P) = 0$. I show that $T'_3(\phi) < 0$ when $T_3(\phi)$, which suffices to show that $T_3(\phi)$ has at most one root.

Differentiating with respect to ϕ :

$$\begin{aligned} T'_3(\phi) &= -\frac{2(y_A - b)\delta}{p_B u_B^3} \frac{du_B}{d\phi} + \frac{2(y_B - b)\delta}{p_A u_A^3} \frac{du_A}{d\phi} + (y_A - y_B) \frac{2\eta\alpha_M(1 + \eta)}{(u_A + u_B)^2} \left(\frac{du_A}{d\phi} + \frac{du_B}{d\phi} \right) \\ &= 2 \frac{du_A}{d\phi} \left[\frac{(y_B - b)\delta}{p_A u_A^3} + (y_A - y_B) \frac{\eta\alpha_M(1 + \eta)}{(u_A + u_B)^2} \right] - 2 \frac{du_B}{d\phi} \left[\frac{(y_B - b)\delta}{p_A u_A^2 u_B} + (y_A - y_B) \frac{\rho}{u_B} \right. \\ &\quad \left. + (y_A - y_B) \frac{\eta\alpha_M}{u_B(u_A + u_B)} \left(2 - \frac{(1 + \eta)u_B}{u_A + u_B} \right) \right] \end{aligned}$$

where we used the condition for $T_3(\phi) = 0$. Notice that both terms in the square brackets are positive. Furthermore:

$$\begin{aligned}\frac{du_A}{d\phi} &= -\frac{u_A\rho(2-u_A-u_B)}{\Delta} \left[\alpha_M \left(1 - \frac{2\eta u_B}{u_A+u_B} \right) + \alpha_{RB} + \frac{\delta}{p_B} - u_B\rho \right] < 0 \\ \frac{du_B}{d\phi} &= \frac{u_B\rho(2-u_A-u_B)}{\Delta} \left[\alpha_M \left(1 - \frac{2\eta u_A}{u_A+u_B} \right) + \alpha_{RA} + \frac{\delta}{p_A} - u_A\rho \right] > 0\end{aligned}$$

and we have proved that $T_3'(\phi) < 0$ when $T_3(\phi) = 0$.

As a result there is a unique ϕ^P that solves the planner's problem. We have:

$$\begin{aligned}T_3(0) &\leq 0 && \Rightarrow \phi^P = 0 \\ T_3(0) &> 0 \quad \text{and } T_3(1) \geq 0 && \Rightarrow \phi^P = 1 \\ T_3(0) &> 0 \quad \text{and } T_3(1) < 0 && \Rightarrow \phi^P \in (0, 1) \text{ and } T(\phi^P) = 0\end{aligned}$$

If $y_A = y_B$ and $p_A = p_B$ then $T_3(\phi) = 0 \Leftrightarrow u_A = u_B$. Therefore the planner's solution is $\phi^P = \frac{1}{2} = \phi^*$. For general values of y_i and p_i , note that $T_3(\phi) \neq T_2(\phi)$ and therefore $\phi^P \neq \phi^*$. ■

Proposition 5.2.

Proof. If $y_A = y_B$ and $p_A = p_B$ then $T_3(\phi) = 0 \Leftrightarrow u_A = u_B$. Therefore the planner's solution is $\phi^P = \frac{1}{2} = \phi^*$ which proves part 1. It is easy to see that:

$$\frac{dT_3(\phi)}{dy_A} > 0 > \frac{dT_3(\phi)}{dy_B}$$

and therefore when ϕ^P is interior we have:

$$\frac{d\phi^P}{dy_A} > 0 > \frac{d\phi^P}{dy_B}$$

which proves part 3.

We examine the case where $y_A = y_B = y$. The planner's problem is given by:

$$\max_{\phi} \mathcal{W} = 2y - (u_A(\phi) + u_B(\phi))(y - b)$$

The solution is given by the root to:

$$\begin{aligned}\mathcal{W}'(\phi) &= -(u'_A(\phi) + u'_B(\phi))(y - b) \\ &= -\frac{\rho(2 - u_A - u_B)u_A u_B}{\Delta} \left[\frac{\delta(y - b)}{p_B u_B^2} - \frac{\delta(y - b)}{p_A u_A^2} \right]\end{aligned}$$

When $y_A = y_B$ the planner's solution satisfies:

$$T_3(\phi) = u_A^2 p_A - u_B^2 p_B = 0$$

Recall that $T_3(\phi)$ crosses the x-axis from above, in the case of an interior solution and that when $p_A = p_B$ then $T_3(\phi) = 0$ only if $u_A = u_B$ which happens when $\phi^P = \frac{1}{2}$. I show that $T_3(\frac{1}{2}) < 0$ when $p_A > p_B$ which proves that $\phi^P < \frac{1}{2}$. When $\phi = \frac{1}{2}$ we have $\alpha_{RA} = \alpha_{RB}$. Let $\alpha = \alpha_M + \alpha_{RA} = \alpha_M + \alpha_{RB}$. The steady state conditions can be rearranged as follows:

$$\begin{aligned}u_A &= \frac{\delta}{\alpha p_A + \delta} \\ u_B &= \frac{\delta}{\alpha p_B + \delta}\end{aligned}$$

Then:

$$\begin{aligned}T_3\left(\frac{1}{2}\right) &= p_A u_A^2 - p_B u_B^2 \\ &= p_A \left(\frac{\delta}{\alpha p_A + \delta}\right)^2 - p_B \left(\frac{\delta}{\alpha p_B + \delta}\right)^2\end{aligned}$$

Note that $T_3(\frac{1}{2}) < 0$ if:

$$\begin{aligned}p_A(\alpha p_B + \delta)^2 &< p_B(\alpha p_A + \delta)^2 \\ p_A \alpha^2 p_B^2 + 2p_A \alpha p_B \delta + p_A \delta^2 &< p_B \alpha^2 p_A^2 + 2p_B \alpha p_A \delta + p_B \delta^2 \\ (p_A - p_B)\delta^2 &< p_A p_B \alpha^2 (p_A - p_B) \\ \delta^2 &< p_A p_B \alpha^2\end{aligned}$$

Combining the steady state conditions we have:

$$\delta^2 = p_A \alpha p_B \alpha \frac{u_A}{1 - u_A} \frac{u_B}{1 - u_B}$$

which proves that the above inequality holds since $u_A < u_B \leq \frac{1}{2}$. This completes the proof. ■

References

- [1] Aigner, Dennis J. and Glen G. Cain (1977). “Statistical theories of discrimination in labor markets.” *Industrial and Labor Relations Review* 30 (2), 175-187.
- [2] Barr, Tavis, Raicho Bojilov and Lalith Munasinghe (2016). “Learning about the employer-employee match when workers refer job candidates: referrals and search efficiency.” Working paper.
- [3] Bayer, Patrick, Stephen L. Ross and Giorgio Topa (2008). “Place of work and place of residence: informal hiring networks and labor market outcomes.” *Journal of Political Economy* 116 (6), 1150-1196.
- [4] Beaman, Lori, Niall Keleher and Jeremy Magruder (2015). “Do job networks disadvantage women? Evidence from a recruitment experiment in Malawi.” Working paper.
- [5] Bentolila, Samuel, Claudio Michelacci and Javier Suarez (2010). “Social contacts and occupational choice.” *Economica* 77, 20-45.
- [6] Bertrand, Marianne and Sendhil Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination.” *American Economic Review* 94(4), 991-1013.
- [7] Brown, Meta, Elizabeth Setren and Giorgio Topa (2016). “Do Referrals Lead to Better Matches? Evidence from a Firms Employee Referral System.” *Journal of Labor Economics* 34 (1), 161-209.
- [8] Burdett, Ken and Melvyn G. Coles (1997). “Marriage and Class.” *Quarterly Journal of Economics* 112 (1), 141-168.
- [9] Burks, Stephen V., Bo Cowgill, Mitchell Hoffman and Michael Housman (2015). “The value of hiring through employee referrals.” *Quarterly Journal of Economics* 130(2), 805-839.

- [10] Calvo-Armengol, Antoni (2004). "Job contact networks." *Journal of Economic Theory* 115, 191-206.
- [11] Calvo-Armengol, Antoni and Matthew Jackson (2004). "The effects of social networks on employment and inequality." *American Economic Review* 94, 426-454.
- [12] Castilla, Emilio (2005). "Social networks and employee performance in a call center." *American Journal of Sociology* 110 (5), 1243-1283.
- [13] Dustmann, Christian, Albrecht Glitz, Uta Schoenberg and Herbert Brucker (2016). "Referral-based job search networks." *Review of Economic Studies* 83, 514-546.
- [14] Fernandez, Roberto and Nancy Weinberg (1997). "Sifting and sorting: personal contacts and hiring in a retail bank." *American Sociological Review* 62, 883-902.
- [15] Galeotti, Andrea and Luca Paolo Merlino (2014). "Endogenous job contact networks." *International Economic Review* 55 (4), 1201-1226.
- [16] Galenianos, Manolis (2013). "Learning about match quality and the use of referrals." *Review of Economic Dynamics* 16, 668-690.
- [17] Galenianos, Manolis (2014). "Hiring through referrals." *Journal of Economic Theory* 152, 304-323.
- [18] Goldin, Claudia and Cecilia Rouse (2000). "Orchestrating impartiality: the impact of "blind" auditions on female musicians." *American Economic Review* 90 (4), 715-741.
- [19] Granovetter, Mark (1973). "The strength of weak ties." *American Journal of Sociology* 78 (6), 1360-1380.
- [20] Granovetter, Mark (1995). *Getting a job: a study of contacts and careers*. 2nd edition Cambridge, MA. Harvard University Press.

- [21] Hensvik, Lena and Oskar Skans (2015). "Social networks, employee selection and labor market outcomes." *Journal of Labor Economics*, forthcoming.
- [22] Ioannides, Yannis and Linda Datcher Loury (2004). "Job information networks, neighborhood effects, and inequality." *Journal of Economic Literature* XLII, 1056-1093.
- [23] Igarashi, Yoske (2016). "Distributional effects of hiring through networks." *Review of Economic Dynamics* 20, 90-110.
- [24] Kugler, Adriana (2003). "Employee referrals and efficiency wages." *Labour Economics* 10, 531-556.
- [25] McPherson, Miller, Lynn Smith-Lovin and James M. Cook (2001). "Birds of a feather: Homophily in social networks." *Annual Review of Sociology* 27, 415-444.
- [26] Montgomery, James D. (1991). "Social networks and labor market outcomes." *American Economic Review* 81, 1408-1418.
- [27] Mortensen, Dale T. and Tara Vishwanath (1994). "Personal contacts and earnings: it is who you know!" *Labour Economics* 1, 187-201.
- [28] Moss-Racusin, Corinne A., John F. Dovidio, Victoria L. Brescoll, Mark J. Graham and Jo Handelsman (2012). "Science faculty's subtly gender biases favor male students." *PNAS* 109 (41), 16474-16479.
- [29] Pellizzari, Michele (2010). "Do friends and relatives really help in getting a good job?" *Industrial and Labor Relations Review* 63 (3), 494-510.
- [30] Pinkston, Joshua (2012). "How Much Do Employers Learn from Referrals?" *Industrial Relations* 51(2), 317-341.
- [31] Pistaferri, Luigi (1999). "Informal networks in the Italian labor market." *Giornale degli Economisti* 58 (3-4), 355-375.

- [32] Simon, Curtis and John Warner (1992). "Matchmaker, matchmaker: The effect of old boy networks on job match quality, earning and tenure." *Journal of Labor Economics*, 10 (3), 306-330.
- [33] Topa, Giorgio (2001). "Social Interactions, Local Spillovers and Unemployment." *Review of Economic Studies* 68, 261-295.
- [34] Topa, Giorgio (2011). "Labor Markets and Referrals." *The Handbook of Social Economics*. Amsterdam. Elsevier North Holland.
- [35] Weinberg, Bruce A., Patricia Reagan B. and Jeffrey J. Yankow (2004). "Do neighborhoods affect hours worked? Evidence from longitudinal data." *Journal of Labor Economics* 22 (4) 891-924.