Supplementary Material to "An Economic Model of Friendship: Homophily, Minorities and Segregation"

Sergio Currarini* Matthew O. Jackson† Paolo Pin‡

October 2007

1 Introduction

This material supplements the paper "An Economic Model of Friendship Formation: Homophily, Minorities and Segregation" with an analysis of two important issues. First, we provide a treatment of the friendship formation model when preferences are satiated in the number of friends. We show that our qualitative results are still holding in this case, and that, in addition, inbreeding homophily can be generated for all groups even in the absence of bias in the technology of the meeting process. This point is of some importance, since it points to a larger role for choice and preferences in generating the observed patterns of friendships in U.S. high schools.

Second, we study in detail a model with discrete friendships, for which the version of the "law of large numbers" used in the paper with a continuum of agents no longer applies. For this case, we study the richer strategic framework that describes the matching process, and show that the structure of equilibrium strategies "replicates" the one we have derived in the continuum case, providing further justification to our model.

*Dipartimento di Scienze Economiche, Università di Venezia and School for Advanced Studies in Venice (SSAV). Email: s.currarini@unive.it

†Department of Economics, Stanford University and the Santa Fe Institute. Email: jacksonm@stanford.edu, http://www.stanford.edu/∼jacksonm/

‡Abdus Salam International Center for Theoretical Physics, Trieste and Università di Venezia. Email: pin@unive.it

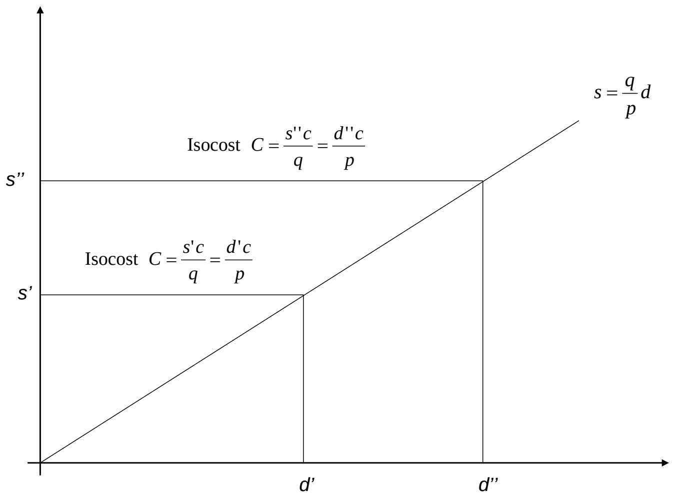

Figure 1: Isocost rectangles

2 Satiated Preferences

We study the endogenous matching model when preferences are satiated in friendships. Given the assumption of diminishing marginal returns to friends, satiated preferences are such that:

i) there exists a "bliss point" (s_b, d_b) such that $U(s_b, d_b) \geq U(s, d)$ for all (s, d);

ii) for each $s' \geq 0$ there exists $d' \geq 0$ such that $U_d(s', d) \leq 0$ for all $d \geq d'$; for each $d' \geq 0$ there exists $s' \geq 0$ such that $U_d(s, d') \leq 0$ for all $s \geq s'$.

The major change in the analysis with satiated preferences is the possibility that agents optimally reject friendship because of their negative marginal value. Strategies are therefore richer than in the case of non satiated preferences, and include four possible actions at each point in time: "accept only own type", "accept only different type", "accept any type" and "stop searching". It is no longer true in general that the probability of finding a friend of different type that wants to form a link is $1-q$, and we denote such probability by $p \leq 1-q$.

The minimum cost of achieving a point (s, d) is given by (see Figure 1):

$$C(s, d) = c \max \left[\frac{s}{q}, \frac{d}{p} \right].$$

Effectively, one follows a strategy of "any" until whichever of (s, d) is reached first, and then continues to search just for the types of friends that are needed to reach the point. In the continuum, note that the ordering of whether, for instance, “any” is followed first and then “same”, or vice versa no longer matters. We therefore define the following sets:

$$\begin{aligned} A_1 &= \left\{(s,d) : \frac{s}{q} > \frac{d}{p}\right\}; \\ A_2 &= \left\{(s,d) : \frac{s}{q} < \frac{d}{p}\right\}; \\ A_3 &= \left\{(s,d) : \frac{s}{q} = \frac{d}{p}\right\}. \end{aligned}$$

Points in A_1 are most efficiently reached by a combination of “any” and ”same” strategies, those in A_2 are most efficiently reached by a combination of “any” and ”different” strategies, and those in A_3 are most efficiently reached by following only an ”any” strategy. Note that it is never the case that rational agents would follow a strategy that includes both ”same” and ”different”.

Agents solve the following problem:

$$\max_{(s,d)} U(s,d) - C(s,d). \tag{1}$$

If U is strictly quasiconcave, there is a unique cost minimizing point for each possible level of utility. The following proposition establishes necessary conditions for cost minimizing points to be also net utility maximizers.

Proposition 1 *The solution (s^*, d^*) to (1) satisfies the following condition:*

$$qU_s(s^*, d^*) + pU_d(s^*, d^*) = c. \tag{2}$$

Moreover,

a. (s^*, d^*) lies in A_1 only if

$$\frac{c}{q} = U_s(s^*, d^*) \tag{3}$$
$$0 = U_d(s^*, d^*). \tag{4}$$

b. (s^*, d^*) lies in A_2 only if

$$\frac{c}{p} = U_d(s^*, d^*) \tag{5}$$
$$0 = U_s(s^*, d^*). \tag{6}$$

c. (s^*, d^*) lies in A_3 only if:

$$\begin{aligned} U_s(s^*, d^*) &> 0; & (7) \\ U_d(s^*, d^*) &> 0. & (8) \end{aligned}$$

Proof. If (s^*, d^*) lies in A_1, then it solves the following problem:

$$\max_{(s,d)} U(s,d) - \frac{c}{q}s$$

whose solution trivially requires conditions (3)-(4). Similar arguments apply to point b. As for point c, the relative maximization problem is:

$$\max_{s} U(s, \frac{p}{q}s) - \frac{c}{q}s,$$

yielding (2) as a necessary first order condition. Note also that in order for $(s^*, d^*) \in A_3$ to be a solution to (1), conditions (7)-(8) must hold. In fact, if $U_s(s^*, d^*) < 0$ and $s^* > 0$ and $d^* > 0$, we can decrease s by a small amount ds, leaving costs unaffected and increasing utility. Same arguments hold if $U_d(s^*, d^*) < 0$.■

In order to prepare for the next section in which a concept of steady-state equilibrium is introduced (a generalization of the one given in the paper)[1], we now show that for all q_i and p_i there exists a unique optimal strategy (s_i, d_i).

We start by denoting by C_1 and C_2 the set of cost minimizing solutions satisfying conditions (4) and (6), respectively. Let C_3 denote the set of points that satisfy condition (2). It is useful to derive the slopes of the above curves using an application of the implicit function theorem.

We say that $U(s, d)$ satisfies the "substitutes" property if $U_{sd} < 0$. The following inequalities are derived for the case of substitutes:

$$\left.\frac{ds}{dd}\right|_{C_1} = -\frac{U_{dd}}{U_{sd}} < 0; \qquad (9)$$

$$\left.\frac{ds}{dd}\right|_{C_2} = -\frac{U_{sd}}{U_{ss}} < 0; \qquad (10)$$

$$\left.\frac{ds}{dd}\right|_{C_3} = -\frac{qU_{sd} + pU_{dd}}{qU_{ss} + pU_{sd}} < 0. \qquad (11)$$

In the next two lemmas and in the resulting proposition we consider only the case of substitutes. Moreover we assume that $U_{sd}(s, d) > U_{dd}(s, d)$ and $U_{sd}(s, d) > U_{ss}(s, d)$ for all (s, d). This is a standard assumption, which ensures the quasiconcavity of U when $U_{sd} < 0$.

[1] Following similar steps as in Appendix A one could show that our concept of equilibrium can be obtained as a limit of a well defined matching process in discrete time.

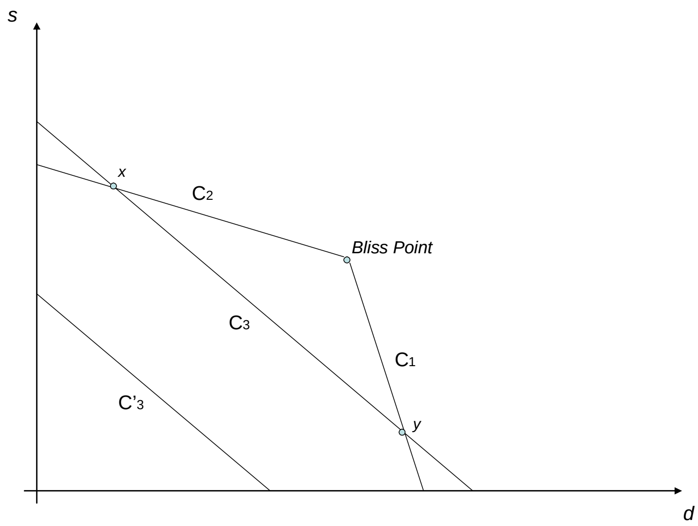

Figure 2: The *loci* C_1, C_2 and C_3.

Lemma 1 *Let $p > 0$ and $q > 0$. Then*

$$\left.\frac{ds}{dd}\right|_{C_2} > \left.\frac{ds}{dd}\right|_{C_3} > \left.\frac{ds}{dd}\right|_{C_1}$$

and

$$\left.\frac{ds}{dd}\right|_{C_2} > -1 > \left.\frac{ds}{dd}\right|_{C_1}.$$

Lemma 2 *Let $p > 0$ and $q > 0$. The curves C_1 and C_3 intersect at most once. The same holds for the curves C_2 and C_3. Moreover, denoting by x and y such intersections, $s_x < s_y$ and $d_x > d_y$.*

Proof. The arguments are mostly graphical. See Figure 2.

We can now fully characterize optimal solutions to (1) for the case of substitutes.

Proposition 2 *If $(s^*, d^*) \in A_1$ satisfy conditions (3)-(4) then (s^*, d^*) is the unique solution to problem (1). If $(s^*, d^*) \in A_2$ satisfy conditions (5)-(6) then (s^*, d^*) is the unique solution to problem (1). Finally, if $(s^*, d^*) \in A_3$ satisfy conditions (2), (7) and (8) then (s^*, d^*) is the unique solution to problem (1).*

Proof. Let $(s^*, d^*) \in A_1$ satisfy conditions (3)-(4). Then by lemma 2 (s^*, d^*) is unique and $(s^*, d^*) = C_1 \cap C_3$. By lemma 1, if (s, d) satisfies conditions (5)-(6), then $(s, d) \notin A_2$.

Consider then any point $(s,d) \in A_2$. This point is always dominated by the minimum cost point $x \in A_3$ in the intersection between the indifference curve passing through (s,d) and A_3. Also, let $(s,d) \in A_3$ satisfy condition (2). It must be that $U_d(s,d) < 0$, since it lies on the right of curve C_1 on which $U_d(s,d) = 0$. By Proposition 1 it cannot be a solution to (1). Using symmetric argument, it can be shown that $(s^*, d^*) \in A_2$ satisfy conditions (5)-(6) is the unique solution to (1). Consider finally $(s^*, d^*) \in A_3$ satisfy conditions (2), (7) and (8). Suppose that (s^*, d^*) is dominated by some point $(s,d) \in A_1$. Consider the optimal point for problem (1) in the closure of A_1. This point must be a cost minimizing point (that is, it must lie on C_1), and must belong to A_3, otherwise it would satisfy conditions (3)-(4). However, all points on A_3 are dominated by (s^*, d^*), which implies a contradiction.■

2.1 Steady State Equilibrium

Our notion of steady state equilibrium is now modified in order to account for the richer set of possible optimal strategies. Given flows $N_1, \ldots, N_n$, *an equilibrium* is a collection of (s_i, d_i, M_i) for each i (along with indirectly determined (q_i, p_i, t_i)'s) such that:

(i) $t_i = \max\left[\frac{s_i}{q_i}, \frac{d_i}{p_i}\right]$,[2]

(ii) $M_i = N_i t_i$,

(iii) $q_i = \frac{M_i\left(\frac{s_i}{q_i}\right)/t_i}{\sum_j M_j}$, and

(iv) $p_i = \frac{\sum_{k \neq i} M_k\left(\frac{d_k}{p_k}\right)/t_k}{\sum_j M_j}$.

(i) represents the time that a type i agent needs to spend in the matching process.

(ii) represents the stock of agents of type i that will be in the matching process in steady state at any time.

(iii) represents the relative probability that a type i agent will find another type i agent $\frac{M_i}{\sum_j M_j}$ times the chance that such a same-type agent will be willing to form a friendship with i: $\left(\frac{s_i}{q_i}\right)/t_i$. Note that if $s_i, d_i \in A_1$, then $t_i = \frac{s_i}{q_i}$ and so this second factor is 1.

(iv) represents the probability that a type i agent will find an agent of a type $k \neq i$: $\frac{M_k}{\sum_j M_j}$, times the chance that such an agent will be willing to form a friendship with i: $\left(\frac{d_k}{p_k}\right)/t_k$; and then sums across k.

Let $\overline{M} = \sum_j M_j$. Using (ii), we rewrite (iii) and (iv) as

[2]If $p_i = 0$ or $d_i = 0$ then let 0/0=0.

(iii') $q_i = \frac{N_i \frac{s_i}{q_i}}{\overline{M}}$, and

(iv') $p_i = \frac{\sum_{k \neq i} N_k \frac{d_k}{p_k}}{\overline{M}}$.

We can deduce some constraints on equilibrium.

From (iii') it follows that

$$q_i^2 = \frac{N_i s_i}{\overline{M}} \quad , \tag{12}$$

and so (when $q_j \neq 0$):

$$\frac{q_i}{q_j} = \sqrt{\frac{N_i s_i}{N_j s_j}}. \tag{13}$$

Note that (13) implies that in a steady state if $N_i > N_j$, then $s_i \geq s_j$ implies that $q_i > q_j$.

We can also derive conditions for the consistency of cross-type links in equilibrium. From (iv'), by summing the equation for each p_k where $k \neq i$, we deduce that for any i

$$\frac{N_i d_i}{\bar{M} p_i} = \frac{\sum_{k \neq i} p_k - (n-2)p_i}{n-1} = \frac{\sum_k p_k - (n-1)p_i}{n-1}, \tag{14}$$

Since the left hand side is nonnegative, this implies that (when $n > 2$)

$$p_i \leq \frac{\sum_{k \neq i} p_k}{n-2} \quad . \tag{15}$$

Note that (14) also implies that

$$\frac{\left(\sum_k p_k - (n-1)p_i\right) p_i}{\left(\sum_k p_k - (n-1)p_j\right) p_j} = \frac{N_i d_i}{N_j d_j} \quad . \tag{16}$$

Finally, from (12) and (14), we get

$$\frac{s_i}{d_i} = \frac{(n-1)q_i}{\sum_k p_k - (n-1)p_i} \cdot \frac{q_i}{p_i} \quad , \tag{17}$$

which means that, if i's optimum is in A_3, then $\frac{s_i}{d_i} = \frac{q_i}{p_i}$ and so

$$(s_i, d_i) \in A_3 \text{ implies } q_i = \frac{\sum_k p_k - (n-1)p_i}{n-1}, \tag{18}$$

which can be read also as $q_i + p_i = \frac{\sum_k p_k}{n-1}$. Similarly for the case where i's optimum is in A_1 (that is: $\frac{s_i}{d_i} > \frac{q_i}{p_i}$), we have that

$$(s_i, d_i) \in A_1 \text{ implies } q_i > \frac{\sum_k p_k - (n-1)p_i}{n-1} \tag{19}$$

In the case where i's optimum is in A_2

$$(s_i, d_i) \in A_2 \text{ implies } q_i < \frac{\sum_k p_k - (n-1)p_i}{n-1} \tag{20}$$

Next, note that by the definition of p_i

$$p_i \overline{M} = N_j d_j / p_j + \sum_{k \neq i,j} N_k d_k / p_k,$$

or

$$p_i p_j \overline{M} = N_j d_j + \sum_{k \neq i,j} N_k d_k p_j / p_k.$$

By a similar equation for j we deduce that

$$N_j d_j + \sum_{k \neq i,j} N_k d_k p_j / p_k = N_i d_i + \sum_{k \neq i,j} N_k d_k p_i / p_k.$$

Thus,

$$N_j d_j - N_i d_i = \sum_{k \neq i,j} N_k d_k (p_i - p_j) / p_k. \tag{21}$$

This implies

Proposition 3 *If $n = 2$, then $N_1 d_1 = N_2 d_2$. If $n > 2$, then $N_j d_j > N_i d_i$ if and only if $p_i > p_j$ (and $d_k > 0$ for some $k \neq i, j$).*

2.2 Equilibrium with Two Groups

In this section we study the simpler case of two groups, and analyze the three main empirical evidences outlined in the paper: relative homophily, more total friendships for larger groups and inbreeding homophily for all groups.

The next proposition shows that if neither type is in A3, then relative homophily holds.

Proposition 4 ***(relative homophily)*** *If $N_i > N_j$ and an equilibrium is interior, then $d_i < d_i$. Moreover, if s and d are substitutes and $s_i, d_i \in A_1 \cup A_2$ then $s_i > s_j$.*

Proof: The fact that $d_i < d_j$ follows from Proposition 3. If $s_i, d_i \in A_1$, then $U_d(s_i, d_i) = 0$. Then by the substitutes condition and the fact that $d_i < d_j$, it follows that $s_i > s_j$ or else we would have $U_d(s_j, d_j) < U_d(s_i, d_i) = 0$, which cannot be true given that type j is optimizing. If $s_i, d_i \in A_2$, then $U_s(s_i, d_i) = 0$. Then by the substitutes condition and the fact that $d_i < d_j$, it follows that $s_i > s_j$ or else we would have $U_s(s_j, d_j) < U_s(s_i, d_i) = 0$, which cannot be true given that type j is optimizing. ∎

We will restrict the analysis to the case in which both types' strategies are in A1 - that is $\frac{s_1}{d_1} > \frac{q_1}{p_1}$ and $\frac{s_2}{d_2} > \frac{q_2}{p_2}$. This could be easily obtained if there is a homophilous bliss point ($s^* > 0$ and $d^* = 0$) and a low cost c, as in the numerical example we are going to discuss.

PROPOSITION **5** ***(larger groups make more friends)*** *Let $N_1 > N_2$, and consider an interior equilibrium such that $\frac{s_1}{d_1} > \frac{q_1}{p_1}$ and $\frac{s_2}{d_2} > \frac{q_2}{p_2}$. Then, $(s_1 + d_1) > (s_2 + d_2)$.*

Proof. Given the slope of C_1, which is

$$\left.\frac{\partial s}{\partial d}\right|_{C_1} = -\frac{U_{dd}}{U_{sd}} < -1$$

and given the result of Proposition 4, we conclude that $(s_1 + d_1) > (s_2 + d_2)$. ∎

The next proposition shows that the larger group always display inbreeding homophily. Moreover, and differently from the case of satiated preferences, this can be true also for the smaller group.

PROPOSITION **6** ***(inbreeding homophily for all groups)***. *Let $N_1 > N_2$ and consider an interior equilibrium such that $\frac{s_1}{d_1} > \frac{q_1}{p_1}$ and $\frac{s_2}{d_2} > \frac{q_2}{p_2}$. Then type 1 always displays inbreeding homophily. Moreover, there exist satiated preferences for which type 2 also displays inbreeding homophily.*

Proof. Both types display inbreeding homophilous if:

$$H_i = \frac{s_i}{s_i + d_i} > w_i = \frac{N_i}{N_1 + N_2} \quad .$$

Note that since $H_i > q_i$ for both types, $q_1 + q_2 = 1$ (this because both types are assumed to be in A1) and $w_1 + w_2 = 1$, we have that at least one of the two types is inbreeding homophilous, possibly both. Note here the difference with the case of satiated preference, in which $H_i = q_i$, and all types cannot have inbreeding homophily at the same time.

From Proposition 3 ($N_1 d_1 = N_2 d_2$) we have that

$$w_1 = \frac{N_2(d_2/d_1)}{N_2(1 + d_2/d_1)} = \frac{d_2}{d_1 + d_2} \text{ and } w_2 = \frac{d_1}{d_1 + d_2} \quad .$$

Since

$$H_1 = \frac{s_1}{s_1 + d_1} \quad \text{and} \quad H_2 = \frac{s_2}{s_2 + d_2} \quad ,$$

we have that $H_1 > w_1$ if and only if $s_1 > d_2$, as well as $H_2 > w_2$ if and only if $s_2 > d_1$. Note that since we have proved that $s_1 + d_1 > s_2 + d_2$, we know that type 1 must display inbreeding homophily. We are therefore interested in establishing that condition $s_2 > d_1$ can be satisfied. We do this by means of a specific form of satiated preferences:

$$U(s, d) = -\alpha(s - M)^2 - \beta d^2 - \gamma(s - M)d$$

where $(M, 0)$ is the bliss point, with $M > 0$, $2\alpha > \gamma > 0$ and $2\beta > \gamma$ are parameters restrictions. Note that this utility has the substitutes property and satisfies diminishing marginal returns. It can be shown that for any positive N_1 and N_2, there is a cost $c > 0$ for which both groups have optima in A1.[3]

Steady state conditions for these preferences are:

$$\left\{ \begin{array}{rcl} U_s^i & = & \frac{c}{q_i} \\ U_d^i & = & 0 \\ N_1 d_1 & = & N_2 d_2 \\ q_i & = & \frac{N_i \frac{s_i}{q_i}}{N_1 \frac{s_1}{q_1} + N_2 \frac{s_2}{q_2}} \end{array} \right. \Rightarrow \left\{ \begin{array}{rcl} -2\alpha(s_i - M) - \gamma d_i & = & \frac{c}{q_i} \\ -2\beta d_i - \gamma(s_i - M) & = & 0 \\ \frac{d_1}{d_2} & = & \frac{N_2}{N_1} \\ \frac{q_1}{q_2} & = & \sqrt{\frac{N_1 s_1}{N_2 s_2}} \end{array} \right.$$

$$\Rightarrow \left\{ \begin{array}{rcl} \left(4\frac{\alpha\beta}{\gamma} - \gamma\right) d_i & = & \frac{c}{q_i} \\ \frac{q_1}{q_2} & = & \sqrt{\frac{d_2 s_1}{d_1 s_2}} \end{array} \right. .$$

If we divide the first equation for type 1 by the first equation for type 2, and substitute in the other equation, we obtain

$$\frac{d_1}{d_2} = \frac{q_2}{q_1} \Rightarrow \frac{d_1}{d_2} = \sqrt{\frac{d_1 s_2}{d_2 s_1}} \Rightarrow \frac{s_1}{d_2} = \frac{s_2}{d_1} \quad .$$

This means that also type 2 exhibits inbreeding homophily for any $N_1 \geq N_2 > 0$.■

3 Discrete Friendships

We study the structure of optimal strategies when agents come in discrete units. The main difference from the case of a continuum of friends is the presence of randomness when agents

[3] That is because, as c gets smaller, the C_3 curve (characterized by the equation $qU_s + pU_d = c$) gets arbitrarily closer to the bliss point.

accept any type of friend. In this case, strategies may be contingent on the realized type of the match, and optimal strategies may keep track of all such contingencies. We show in this section that, at least in the case of homophilous preferences, optimal strategies have a simple structure, which is similar to the one we would find in the case of a continuum of friends: agents accept any friend until some threshold level of different type friends is achieved, and then only accept same type friends until they stop searching.

3.1 One Friend

To get an idea of the main forces at work, we start with the simple case where agents desire at most one friend. Consider a case where $U(1,0) > U(0,1) > c$, $c > U_s(0,1) > U_s(1,0)$ and $c > U_d(0,1) > U_d(1,0)$.[4]

In this scenario, agents search for at most one friend and prefer to have a friend of their own type. There are two things to examine here. First, conditional on searching will a given type be willing to form a friendship with an agent of a different type if he or she happens to meet one, or will they only be willing to form a friendship with an agent of their own type? Second, is it worthwhile searching at all?

Without loss of generality for this case, set $U(1,0) = f > U(0,1) = 1 > c$.

Starting with the first question; a strategy of only accepting a match with one's own type yields a expected discounted payoff to an agent of type i of

$$V_i(own) = q_i f - c + (1 - q_i)V_i(own)$$

or

$$V_i(own) = \frac{q_i f - c}{q_i}.$$

An analogous calculation shows that a strategy of accepting a friendship with any (willing) type yields an expected discounted payoff of

$$V_i(any) = \frac{q_i f + p_i - c}{q_i + p_i}.$$

Thus, an agent of type i is willing to form a friendship with an agent of a different type if and only if

$$q_i \geq q_i f - c. \tag{22}$$

It follows that the optimal "one-friend" strategies are:

[4] By $U_s(s,d)$ we here mean the difference $U(s+1,d) - U(s,d)$, and the same for $U_d(s,d)$. Similarly we can define the second order direct and cross differences U_{ss} , U_{dd} and U_{sd}.

- not to search if $c \geq q_i f + p_i$;
- to search and accept any friend if $q_i f + p_i \geq c \geq q_i(f-1)$, and
- to search and only accept own types if $q_i(f-1) \geq c$.

Here we see some intuitive comparative statics, in line with the result we found in the analysis with a continuum. Agents with higher levels of p_i and q_i will search while agents with low enough levels will not. Also, agents who have high enough values of q_i will search but will only form friendships with their own types. This means that we can find parameter ranges where larger groups (higher q_i's) are unwilling to form cross-type friendships, while smaller groups are willing to form cross-group friendships.

3.2 Two Friends

Let us now examine a setting where agents are willing to form at most two friendships.

Consider a situation where $U(2,0) > U(1,1) > U(0,2)$ and $U_s(s_i, d_i) \leq 0$ and $U_d(s_i, d_i) \leq 0$ when $s_i + d_i \geq 2$. So, agents value at most two friends and prefer to form friendships with agents of their own type.

Let us also suppose that q_i, p_i and c are such that agents are always willing to form two friendships.

It is clear that in this case, an agent, when searching, will always be willing to form a friendship with an own type, since the value of an own–type friendship always dominates the value of a friendship with a different type. The question becomes whether or not an agent is willing to form a friendships with agents of different types.

In order to characterize the strategies, we start with a simple observation. Strategies are not affected by a failed meeting, so that if an agent sets with a given strategy and does not form any friendships in a given period, then it is optimal to still pursue that same strategy starting in the next period. This allows us to look for optimal strategies of a form that only change contingent on forming a friendship.

So, the types of strategies that we need to consider can be described by either being willing to form any friendship available in a given period or only willing to form friendships with same types. This leads to six possible strategies, listed by the strategies followed on the first friend and then the second friend:[5]

[5] We do not need to consider strategies of the form Same on the first friend and then contingency afterwards since the first friend is obviously of a same type under such a strategy.

- Same, Same,
- Same, Any,
- Any, Same,
- Any, Any,
- Any, and then Same if first friend is of the same type and Any if first friend is of a different type,
- Any, and then Any if first friend is of the same type and Same if first friend is of a different type.

The first thing to note is that the strategies (Same, Any) and (Any, Same) induce exactly the same distributions over possible friendships. This translates into a more general result that we discuss in more detail below.

In all the following analysis we will make use of the assumptions we used in previous section: $U_{ss} \leq U_{sd} < 0$ and $U_{dd} \leq U_{sd}$.

Note that under these assumption, then it does not make sense to follow a strategy of Same if first friend is of the same type and Any if first friend is of a different type, unless it happens to be that one is indifferent between these two strategies regardless of the outcome on the first friendship. This is true since it is the marginal value of a different type which becomes relatively more valuable if the first friend is of the same type, and vice versa. This is another more general result that we discuss below: contingent strategies should move in a direction opposite of the last friendship formed (so if it is a same friend, then to Any, and if it is a different, then to Same). When we refer to contingent, we then refer to such a strategy.

Thus, in looking for an optimal strategy it is enough to consider the following strategies:

- Same, Same,
- Any, Same
- Any, Any,
- Any, Contingent,

where contingent means following Any if first friend is of the same type and Same if first friend is of a different type.

Next, from our analysis of the one-friend case, we know that the decision of whether to search for Any or Same on a second friend depends on the relative size of c/q_i compared to $U(2,0) - U(1,1)$ if the first friend was of the same type and $U(1,1) - U(0,2)$ if the first friend was of a different type.

It is clearly possible to have both of these larger than c/q_i, both smaller than c/q_i, or the first larger and the second smaller. Thus, conditional on having followed an Any strategy on the first friend, it is possible to have Same, Any or Contingent be optimal on the second friend. We then need to check whether Any can be an optimal strategy on the first friend with each of the scenarios in the second stage.

We now argue that (Any, Same) is never a uniquely optimal strategy. If (Any, Same) is a better strategy than (Same, Same), then since (Same, Any) is equivalent to (Any, Same), it is also is a better strategy than (Same, Same). This implies that it is not worthwhile holding out for a same friendship on the second friend conditional on the first friend being same. Thus, $c/q_i > U(2,0) - U(1,1)$. Now, for (Any, Same) to be a better strategy than (Any, Contingent), it has to be that Same is a better strategy than Any contingent on having a same friend on the first friend. This implies that $c/q_i < U(2,0) - U(1,1)$, and so we have reached a contradiction.

Thus, the three possible optimal strategies (outside of cases of indifference) are

- Same, Same,
- Any, Contingent,
- Any, Any.

These strategies are optimal according to the following comparisons.[6]

[6] Note that the expected cost of following a strategy where there is a probability p of making a friend in any given period has a geometric distribution, and hence the expected number of periods until a friendship is formed is $1/p$ and the anticipated search cost incurred until the friendship is formed is thus $-c/p$.
The expected payoff to a strategy of (Any, Any) is

$$\frac{p_i^2 U(0,2) + 2q_i p_i U(1,1) + q_i^2 U(2,0)}{(p_i+q_i)^2} - \frac{2c}{p_i+q_i}. \tag{23}$$

The expected payoff to a strategy of (Any, Contingent) is

$$\frac{\left(p_i^2 + 2q_i p_i\right) U(1,1) + q_i^2 U(2,0)}{(p_i+q_i)^2} - c\left(\frac{1}{p_i+q_i} + \left(\frac{p_i}{p_i+q_i}\right)\frac{1}{q_i} + \left(\frac{q_i}{p_i+q_i}\right)\frac{1}{p_i+q_i}\right). \tag{24}$$

- (Same, Same) is optimal if $U(2,0) - U(1,1) \geq c/q_i$
- (Any, Contingent) is optimal if $U(1,1) - U(0,2) \geq c/q_i \geq U(2,0) - U(1,1)$,[7]
- (Any, Any) is optimal if $c/q_i \geq U(1,1) - U(0,2)$.

We remark that p_i does not enter any of these calculations! This follows from the fact that the relative calculation is always as to whether or not to accept a friendship with a different type when it is available in a given match, or to hold out for a friendship with a same type. The probability of matching with a same type is given by q_i.

It follows easily from the above calculations that we can find situations where

- the most populous types only form same-type friendships,
- a middle range of types in terms of population size forms first-friend relationships with any types but second friendships contingently, and
- the least populous types are willing to form any friendship possible on both friends.

This is illustrated in the following example.

Example 1 *Three Groups Following Different Strategies*

Consider the case in which $U(2,0) = 4$, $U(1,1) = 3$ and $U(0,2) = 1$. Let $c = .25$. Let $q_1 = .3$, $q_2 = .2$, and $q_1 = .1$. Group one only forms own-type friendships, group 2 forms any friendship on the first friend and follows a contingent strategy on the second friend, and group 3 forms any friendships on both friends.

The expected payoff to a strategy of (Any, Same) or (Same, Any) is

$$\frac{p_i U(1,1) + q_i U(2,0)}{p_i + q_i} - c\left(\frac{1}{q_i} + \frac{1}{p_i + q_i}\right). \tag{25}$$

The expected payoff to a strategy of (Same, Same) is

$$U(2,0) - 2\frac{c}{q_i}. \tag{26}$$

One can directly check that (Same, Same) is weakly better than (Any, Contingent) if and only if $U(2,0) - U(1,1) \geq c/q_i$.
Similarly, (Any, Contingent) is weakly better than (Any, Any) if and only if $U(1,1) - U(0,2) \geq c/q_i$.
Note also that the first condition implies the second. The claimed optimality results follow directly.

[7] This range is nonempty by the assumption that $U_{ss} < U_{sd}$ and $U_{dd} < U_{sd}$.

3.3 Beyond Two Friends: General Characterizations of Steady State Strategies

For now, let us restrict attention to a case where $U_s(s_i, d_i) \geq U_d(s_i, d_i)$ for all s_i, d_i. Here, additional same-type friends are always at least as valuable as different-type friends. Without this condition strategies where one only considers friendships with different types can be optimal.

Note that once we move beyond two friends, we may have strings of contingencies in a strategy. For instance, it is conceivable to have a strategy where one follows a strategy of "Any" on the first friendship, and then follow a strategy of "Same" until some stopping point after if the first friend is different, and "Any" until some stopping point after if the first friend is same. One could also follow nested sorts of contingent strategies of the form: follow "Any" on the first friendship, and then follow "Any" again as long as only same friendships have been made unless hitting a stopping point, and once a different friendship has been made then follow a strategy of "Same" until stopping.

While complicated variations are conceivable, optimal strategies follow some specific forms. A basic intuition is that it is not advantageous to switch back and forth between contingent and non-contingent strategies, or "Any" and "Same" strategies.

There are a number of intuitions from Section 3.2 that extend beyond the two-friend case.

First, it is important to observe that we do not need to take into account how many periods it takes to realize a portion of a given strategy. For instance, if an agent is following a given strategy at the beginning of a period and does not form a friendship in that period, then it is optimal to continue following that strategy at the beginning of the next period. Thus, we can describe a strategy by a tree. The tree is formed of nodes which describe the strategy that is followed until a new friendship is formed.

So for instance following an "Any" action, either a d or s friend could be found. Thus, a node with an "Any" strategy has two successor nodes. A node corresponding to a "Same" action or a "Different" action is followed by a single successor node. An action of "Stop" is followed by a terminal node.

Figure 3 represents a strategy where an agent would like to form at most three friendships, at most one of which is a different friend. The agent plays the action "Any" until a different friend is found and then stops at three friendships, but stops once three friends have been found even if they are all of the "same" type.

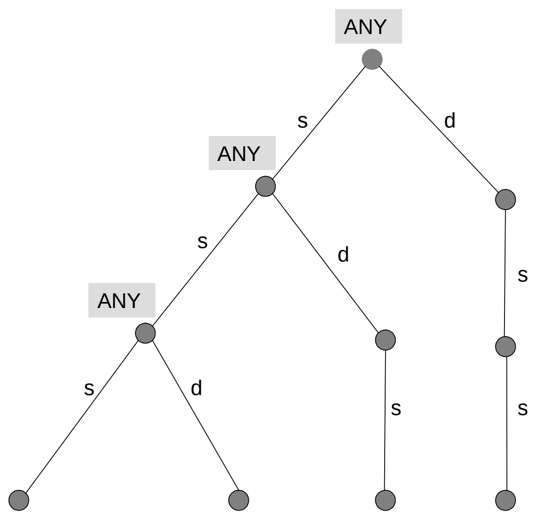

Figure 3: Tree representation of an equilibrium strategy.

PROPOSITION **7** *An optimal strategy is always equivalent to a strategy represented by a tree that has the following properties:*

(i) Pick any path in the tree, and let $\eta_0, \ldots \eta_K$ be the nodes along that path (with η_0 being the root node corresponding to the starting action). There exists k' such that "Any" is played at all nodes η_k where $k < k'$, and then either "Same" is played at all non terminal nodes η_k where $k \geq k'$, or "Different" is played at all non terminal nodes η_k where $k \geq k'$.

(ii) If "Same" is played following s then "Different" is not played following d.

The following properties hold if same-type friends dominate different-type friends, so that $U_s(s,d) \geq U_d(s,d)$ for all (s,d):

(iii) An action of "Different" is never played.

(iv) Following a node that corresponds to a play of "Any", if "Same" is played at a successor node, it is only played following the outcome of d. This, together with (iii), implies that if an action of "Any" follows a realization of d, then either it also follows a realization of s or else "Stop" follows the realization of s.

Thus, when same-type friends dominate different-type friends, an optimal strategy is always equivalent to a strategy that plays "Any" until some number of d friendships have been formed or some limit is hit (that could depending on the fraction of d's found so far), and then follow "Same" thereafter.

Proof: Consider a tree T and any node η where "Same" is played. Create a new tree as follows. Cut the tree at the node η and move the sub-tree that starts from successor node to η to take the place of η. Next, replace each terminal node in that sub-tree with one action of "Same" and then replace the terminal node. This leaves the realizations of the strategy completely unchanged and thus is equivalent. The same is true of "Different". So we can find a tree equivalent to an optimal strategy that has the following property:

(i′) Pick any path in the tree, and let $\eta_0, \dots \eta_K$ be the nodes along that path (with η_0 being the root node corresponding to the starting action). There exists k' such that "Any" is played at all nodes η_k where $k < k'$, and then either "Same" or "Different" is played at each non terminal node η_k where $k \geq k'$.

We then conclude (i) by noting that a strategy of "Different" and then "Same" is dominated by a strategy of "Any" and then "Different" following s and "Same" following d.[8]

Next, note that a strategy of only "Differents" is more valuable following a realization of s than d, and similarly a strategy of only "Sames" is more valuable following a realization of d than s (this because of diminishing marginal returns). Thus, if "Same" is chosen over different after s it is chosen after d as well.

Point (iii) follows easily from the condition that $U_s(s,d) \geq U_d(s,d)$ for all (s,d); which implies that a strategy of "Any" is at least as good as "Different".

Having found for any optimal strategy an equivalent strategy that has a tree satisfying (ii) and (iii), we work with this tree in what remains. To prove (iv) we proceed in steps.

First, consider a node η where "Any" is played followed by only "Same" on both succeeding subtrees. Let s_0, d_0 be the friend composition at η. Suppose that $x \geq 1$ "Sames" are played following a realization of s, while $y \geq 1$ "Sames" are played following a realization of d. Let C_S denote the anticipated cost of finding a friend under a "Same" action and C_A under an "Any" action.[9] This implies that "Any" is at least as good as simply following

[8]Note that if an optimal strategy plays both "Different" and "Same", then it must be that $q_i > 0$ and $p_i > 0$.

[9]Note that $C_S = \frac{c}{q_i}$ and $C_A = \frac{c}{q_i + p_i}$.

"Same" for $x+1$ more friends.

$$\frac{p_i}{p_i+q_i}\left[U(s_0+y,d_0+1)-yC_S\right]+\frac{q_i}{p_i+q_i}\left[U(s_0+x+1,d_0)-xC_S\right]-C_A \geq U(s_0+x+1,d_0)-(x+1)C_S. \quad (27)$$

Alter the strategy as follows. If s is realized after η, then at the next node play "Any" and then conditional on that outcome play $y-1$ "Sames" if d and then $x-1$ "Sames" if s. Conditional on s being realized after η, the expected continuation payoff from following this strategy is:

$$\frac{p_i}{p_i+q_i}\left[U(s_0+y,d_0+1)-(y-1)C_S\right]+\frac{q_i}{p_i+q_i}\left[U(s_0+x+1,d_0)-(x-1)C_S\right]-C_A, \quad (28)$$

while the expected continuation payoff from the original strategy is

$$U(s_0+x+1,d_0)-xC_S. \quad (29)$$

Note that by (27) it follows that (28) $\geq$ (29). We conclude that we can find a strategy satisfying (i)-(ii), and such that a node where "Any" is played is followed by at most one successor node involving a strategy of "Same". Since "Same" is more valuable after a realization of a different than of a same friend, we obtain the first part of (iv). The second part of (iv) is immediate. ∎