Appendices: The Formation of Collaboration Networks among Individuals with Heterogeneous Skills

Appendix A: Pairwise Stability and Efficiency. Briefly, a network is pairwise stable if no individual would prefer to terminate an existing link, and if no pair of individuals would prefer to add a link. Although this definition is usually used in undirected networks, it works equally well in the current context. Formally, a directed collaboration network, g, is *pairwise stable* if

(1) for all $ij \in g$, $u_i(g) \geq u_i(g - ij)$ and $u_j(g) \geq u_j(g - ij)$
(2) for all $ij \notin g$, if $u_j(g + ij) > u_j(g)$ then $u_i(g + ij) < u_i(g)$

Together, these two conditions ensure that links are mutual. That is, if a network is pairwise stable, then both ends agree to maintain the link.

Theorem. *Any complementarity network, $g \in \Gamma(\Psi)$, is pairwise stable. In other words, $\forall ij \in g$ $u_i(g) \geq u_i(g - ij)$ and $u_j(g) \geq u_j(g - ij)$ and for all $ij \notin g$, if $u_j(g + ij) > u_j(g)$ then $u_i(g + ij) < u_i(g)$.*

Proof. First, consider whether any individual wishes to unilaterally remove a link, $ij \in g$. Removing this incoming link costs individual j her share of the payoff from solving $i's$ problem $\left(\frac{1}{|C_i|+1} \geq 0\right)$, and thus she will never choose to do so. Individual i chose a minimal set of collaborators that allowed her to solve her problem, so removing this outgoing link means that she can no longer solve the problem. This would cost her the payoff $\left(\frac{1}{|C_i|+1} \geq 0\right)$ from solving the problem, and thus she will also never choose to do so. Finally, note that no individual would ever want to add an outgoing link to a cost-minimizing collaboration network, because having chosen a minimal set of collaborators, any additional link would require her to further split her prize. □

Theorem. *Any cost minimizing collaboration network, $g \in \Gamma(\Psi)$, is strongly efficient. In other words, $\sum_i u_i(g) \geq \sum_i u_i(g') \; \forall g' \in G$.*

Proof. Because all value is generated from solving problems, the maximum possible value in the network is N. Since solving problems is incentive compatible and there is no loss, the problem solvers always extract the maximum value from the network. □

Appendix B: General Proof of Theorem 2. An individual with the set $A \cup B$ will be able to help anyone needing any subset of those skills. Let $\delta(C)$ be the demand for a particular set of skills, C. In the general case, $\delta(C) = \Psi(S\backslash C) + \sum_{D:\Psi(C\cup D)=0} \Psi(S\backslash(C \cup D))$. The fraction who can supply the set C is $\sigma(C) = \sum_{D\subseteq S\backslash C} \Psi(C \cup D)$. Note that $\delta(C)$ and $\sigma(C)$ depend only on the particulars of the problem (S), the distribution of skills (Ψ), and the subset of skills (C). Thus, any individual with the skill set $A \cup B$ has expected degree

$$E[d(A \cup B)] = \sum_{C\subseteq A\cup B} \frac{\delta(C)}{\sigma(C)}$$

We can divide the problems that an individual with $A \cup B$ can solve into three groups:

(1) Requires only skills from set A: $C \subseteq A$

(2) Requires only skills from set B, including at least one found only in B : $\{C \,|\, C \subseteq B \, and \, \exists \, b \in C \, st \, b \in B\backslash A\}$

(3) Requires at least one skill from each set that can only be found in that set: $\{C \,|\, C \subseteq A \cup B, \, where \, \exists \, a, b \in C \, s.t. \, a \in A\backslash B \, and \, b \in B\backslash A\}$

Using this partition, we can write

$$\begin{aligned} E[d(A \cup B)] &= \sum_{C\subseteq A} \frac{\delta(C)}{\sigma(C)} + \sum_{C\subseteq B \, and \, C\cap B\neq\emptyset} \frac{\delta(C)}{\sigma(C)} + \sum_{C\subseteq A\cup B \, and \, C\cap A, \, C\cap B\neq\emptyset} \frac{\delta(C)}{\sigma(C)} \\ &= E[d(A)] + \sum_{C\subseteq B \, and \, C\cap B\neq\emptyset} \frac{\delta(C)}{\sigma(C)} + \phi \end{aligned}$$

which implies that

$$\begin{aligned} E[d(A \cup B)] + E[d(A \cap B)] &= E[d(A)] + \sum_{C\subseteq B \, and \, C\cap B\neq\emptyset} \frac{\delta(C)}{\sigma(C)} + \phi + E[d(A \cap B))] \\ &= E[d(A)] + \left(\sum_{C\subseteq B \, and \, C\cap B\neq\emptyset} \frac{\delta(C)}{\sigma(C)} + \sum_{C\subseteq A\cap B} \frac{\delta(C)}{\sigma(C)} \right) + \phi \\ &= E[d(A)] + E[d(B)] + \phi \\ &\geq E[d(A)] + E[d(B)] \end{aligned}$$

Appendix C: Discussion of the Gini Coefficient in the discrete case. The Gini coefficient measures the area between the Lorenz curve of a distribution (in this case, the

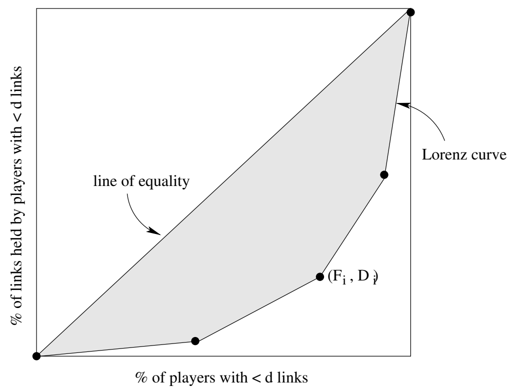

FIGURE 1. An example of the Gini coefficient for a discrete distribution, $\Delta(y)$. In this case, the random variable y takes on one of five values, $y_0 ... y_4$. The Gini coefficient is the area of the shaded region between the line of equality and the Lorenz curve.

distribution of expected degree), and the line of equality. In the case of a discrete distribution with values $y_0 ... y_N$ where $y_i < y_{i+1}$, the Lorenz curve is a piecewise function connecting points (F_i, D_i) where $F_i = \sum_{k=0}^{i} \Delta(y_k)$ is the fraction of individuals with strictly less than y_i links, and $D_i = \frac{\sum_{k=0}^{i} \Delta(y_k) y_k}{\sum_{k=0}^{N} \Delta(y_k) y_k}$ is the fraction of the total number of links they hold. See Figure 1 for an example. The gini coefficient for a discrete distribution is given by $G = 1 - \sum_{i=1}^{N} D_i (F_i - F_{i-1})$. Lower values of the gini coefficient indicate a more equal distribution of links across individals in the population, and higher values indicate a more skewed distribution of links. The coefficient is which is 0 when the distribution is perfectly equal (ie: the bottom x% of the population holds exactly x% of the links) and 1 when all of the links are held by a single individual.