Assortative matching and underconnectivity in networks *

Andreas Bjerre-Nielsen†

Abstract

People choose friendships with people similar to themselves, i.e. they sort by resemblence. Economic studies have shown when sorting is optimal and constitute an equilibrium, however, this presumes lack of beneficial spillovers. We investigate formation of economic and social networks where agents may form or cut ties. We combine a setup with link formation where agents have types that determine the value of a connection. We provide conditions for sorting in friendships, i.e. that agents tend to partner only with those with those sufficiently similar to themselves. Conditions are provided with and without beneficial spillovers from indirect connections. We show that sorting may be suboptimal, yet a socially stable outcome, despite otherwise obeying the conditions for sorting in Becker (1973). We analyze policy tools to mitigate suboptimal sorting.

Keywords: network formation; underconnectivity; assortative matching; network externalities; one-sided matching.

JEL classification: C71, C78, D61, D62, D85.

1 Introduction

A ubiquitous finding in studies of social relations is the tendency to form more ties with people similar to one-self, i.e. the pattern known as *sorting* or homophily cf. the meta-study by McPherson et al. (2001). Pioneered by Becker (1973) economic research has contributed to the understanding of sorting by providing mathematically sufficient conditions for when marriage- and labor markets and groups get sorted and when this is optimal. Yet, no attention has been devoted whether the condition for optimality of Becker (1973) on sorting are valid in the context of network externalities. Our investigation yields new insights on network formation by rational

*This paper has grown out of "Sorting in Networks: Adversity and Structure", (https://arxiv.org/abs/1503.07389). The author is grateful to Peter Norman Sørensen who supervised his PhD. A large thanks to Jan Eeckhout, Matthew Jackson, Jesper Rüdiger, Hans Keiding, Bartosz Redlicki, John Kennes, Thomas Jensen and seminar participants at CoED 2015, CoopMAS 2015, EDGE 2015, University of Copenhagen, Stanford and Universitat Pompeu Fabra. This paper is a revised version of a chapter in the author's PhD dissertation at University of Copenhagen from 2016. The work was funded by University of Copenhagen as part of the author's PhD.

†University of Copenhagen, andreas.bjerre-nielsen@econ.ku.dk

agents: when is sorting stable and/or suboptimal if beneficial network externalities are either present or not? In addition, we demonstrate how to enact policies to curb sorting when excessive, and thus provide higher welfare. These new insights can help policymakers to design better schools or corporations.

A modest tendency to sort may be beneficial by forming communities that with high synergies and shared values. Conversely, too much sorting can be detrimental, e.g. by slowing down the dissemination of information, see Golub and Jackson (2012). As recent decades have seen a rise in residential income segregation and assortative mating by education it is a great concern for policy makers how to tackle the problems that stems from increased sorting.[1] The insights from our analysis are essential to understand the indirect effects of sorting. For instance, specific smaller programs aimed at bridging students from different backgrounds can be effective at mitigating the inefficacy of the social structure that would arise without interventions. It is also relevant for understanding indirect consequences of dividing school cohorts by ability, as is the current practice in many countries (cf. OECD (2012)); such a division impacts not only the direct social peers of the students but also severs the bridge for transmission of positive externalities.

Our framework explores a setting with agents choosing partners under three core assumptions. First, agents are heterogeneous in type for creating value in partnerships (which can be interpreted as a peer effect). Type may refer to productive and non-productive capabilities such as skill, social aptitude or interests and ethnicity. Second, there are possible externalities from friends of friends as in the 'connections model' of Jackson and Wolinsky (1996) which captures spillovers of ideas and favors (which does not contain types). Third, agents choose a limited number of partners reflecting constraints of time and effort.[2] Note our framework could also model corporations or self-governing organizations forming bilateral partnerships among themselves. In this setting, we investigate robustness of network structures in the following sense: no agents can form and/or delete links from network and be better off when allowing for transfers of utility. Most of our results require only pairwise (Nash) equilibrium and thus only requires stability against bilateral deviations. One result requires the strong (Nash) equilibrium where any coalition of agents can form links between themselves.

Our central contribution is to show the general existence of networks sorted by type that are stable yet suboptimal in terms of welfare and that this inferior situation can be alleviated by implementing a welfare enhancing network. Our paper does this by building a parsimonious framework uniting the frameworks of Becker (1973) and Jackson and Wolinsky (1996). Along with the main results we also make a number of smaller contributions toward the understanding of conditions for sorting. The results established in this paper contribute to the game theoretic knowledge on formation of network as well as homophily and assortative matching.

Our generalization of sorting are relevant for further investigations of homophily in social

[1] See Reardon and Bischoff (2011) on segregation and Schwartz and Mare (2005) on marriage sorting.

[2] Limited partners is also consistent with empirical research; Ugander et al. (2011) shows this for the entire Facebook network and Miritello et al. (2013) in phone calls for millions of people.

networks. Moreover, we provide a clear intuition how pathological sorting constitutes an undesirable outcome despite being the only stable configuration of a social network under conceivable circumstances. The intuition is analogue to a classic public goods problem: there is a lack in the provision of links across types. This indicates that sorting, a generic network structure commonly observed, is likely to suffer from inefficiencies in network formation. Finally, to mitigate the sorting issue we demonstrate the validity of a policy instrument that provides incentives to link across types.

Contributions and literature In what follows we explain our results and discuss them in relation to the literature. We begin with the vast literature on optimal networks under externalities. The field has a long tradition and begins with the general formulation under the quadratic assignment problem Koopmans and Beckmann (1957). The field gained traction when Katz and Shapiro (1985); Farrell and Saloner (1986) showed stable but inefficient outcomes can arise in two-sided markets with externalities. The field was revolutionized by Jackson and Wolinsky (1996) demonstrating there is an incompatibility in networks between the stability and efficiency; Bloch and Jackson (2007) extends these results to show the tension is preserved when allowing for more coordination and more flexible transfers between agents.

Unlike the above papers we do not provide any new generic insights. Instead we show existence of a general class of inefficient networks which has the property of being sorted in type and constitute the unique equilibrium under plausible conditions. Specifically we demonstrate that when Becker (1973)'s complementarity condition holds and that there are sufficiently many agents of each type as well as only a moderate level of externalities then the set of pairwise (Nash) stable networks equals a certain set of sorted networks; this set of sorted networks is shown to be non-empty and to consist only of networks that are perfectly sorted where any two agent within a given type is connected (not linked) and each agent uses all its quota of links. We also show that any network in this set is inefficient (under moderate externalities) due to a lack of connectivity across types. For more information, see Theorem 1 for the case of non-constant decay (or Theorem 3 in Appendix B for constant decay). In addition, we show this problem of sub-optimality can be mitigated through policy, see Proposition 3.

A seminal mathematical work on sorting and segregation is Schelling (1969, 1971). Schelling's use models of residential segregation across space to show that only small preferences on composition of neighbors are necessary to yield high segregation. Although related the modeling differences between Schelling's spatial model and networks/matchings are stark; the latter ones have a more flexible setup allowing for connections between any individuals and network effects.

The only paper that investigates inefficient sorting in networks is de Martí and Zenou (2017); they also model type complementarity and positive externalities.[3] Their results show existence of sorted networks that are stable yet inefficient due to the lack of linking across types.[4] Although

[3]Note this paper was developed independently and without awareness of de Martí and Zenou (2017).

[4]In de Martí and Zenou (2017) results on stable sorting are found in Propositions 1.ii, 4.iii while Proposition 6

similar there are some profound and crucial differences that motivates our analysis. Compared with de Martí and Zenou (2017) our setup uses exogenous complementarity while they use endogenous complementarity. This difference, and a few technical ones,[5] allows us to considerably strengthen results: we establish that sorting constitutes the unique set of stable networks; we show there exist a globally efficient network that is sorted but has connectivity between groups and demonstrate that this network is implementable through a simple policy, see Proposition 3; finally, our results are neither limited to only five agents of each type, nor sub-structures of within-type networks being either stars or cliques, nor is it limited to only two types.

The fundamental difference in how complementarity between our setup and de Martí and Zenou (2017) means that our results should be interpreted differently. Our results are relevant in cases when we do *not* expect endogenous complementarity, e.g. in skills, personal interests or to some extent geography. In addition, the modeling of complementarity also entails that the source of inefficiency is different. In our setup suboptimality stems from misaligned incentives (due sorting be the unique stable outcome) which is in line with the incompatibility of efficiency and stability of Jackson and Wolinsky (1996); no one wants to volunteer to build the bridge between communities. On the contrary, in de Martí and Zenou (2017) suboptimality may be due to a lack of coordination on implementing another stable network that is welfare improving.[6] Finally our framework is carefully chosen to directly build on Becker (1973), and as a consequence our results have strong implications for the literature on assortative matching.

The most relevant research on exogenous complementarity in networks is Johnson and Gilles (2000); Jackson and Rogers (2005); Galeotti et al. (2006); the first assumes agents all have a unique type with linking costs proportional to their distance while the two latter use an islands type of model (where agents have same type). Johnson and Gilles (2000) shows existence of a pairwise stable equilibrium with local connectivity between adjacent types, possibly with local cliques where all agents within a given range are connected. Jackson and Rogers (2005) shows that clustering and short paths are robust features among both pairwise stable networks efficient networks with full linkage among same type. Galeotti et al. (2006) investigates minimally connected networks in a setup with one-sided link formation.[7] None of the above papers investigate the problem with the network being 'underconnected' explored in this paper.

Of lower importance we also make a number of specific contributions towards the understand-

contains result on inefficiency.

[5]One noteworthy difference is that we model cost implicitly through an identical quota on the number of links and that we use value complementarity. These two differences entails that our study directly extends the assortative matching framework to a networks setting. de Martí and Zenou (2017) models cost explicitly with complementarity such that there are low linking costs with same types. Another difference is that we use the slightly stronger equilibrium concept than their use of pairwise stability; this change is necessary when modeling cost through a degree quota.

[6]Results in de Martí and Zenou (2017) do not rule out there can exist stable networks (e.g. some amount of connectivity between groups) that are more efficient than a sorted network. This follows as de Martí and Zenou (2017) have multiplicity in equilibria and they only establish relative inefficiency between two networks (complete network and perfectly sorted network of cliques).

[7]Note that one-sided link formation is based on the setup of Bala and Goyal (2000) which only requires the weaker equilibrium concept, Nash stability, as links do not need mutual acceptance.

ing of assortative matching. We establish sufficient conditions for stable networks to exhibit sorting in type. In the simple case without externalities our Proposition 1 extends the classic work to a one-sided matching setting with many partners. In the case of externalities we provide various results, including a characterization, as outlined above. Finally our Proposition 4 establishes that if the agent population is large then sorting is the unique strongly stable outcome (i.e. the core) when Becker's complementarity condition hold and that externalities satisfy asymptotic independence in social connections.

The work related to stability of sorting has a strong tradition for two-sided matching e.g. labor and dating markets starting with Becker (1973). The research on one-sided assortative matching, i.e. the basis for our setup, has been limited to formation of clubs (equivalent to cliques in networks) under various technologies cf. Farrell and Scotchmer (1988), Kremer (1993), Durlauf and Seshadri (2003), Legros and Newman (2002), Pycia (2012), Baccara and Yariv (2013). All the research on one-sided assortative matching finds conditions for sorting which correspond to type complementarity in Becker (1973). Yet, none of these papers allow for linking between groups nor consider network externalities which are the extensions we treat. Note that Proposition 1 can be seen as a direct extension of the classic result by Becker (1973) for networks without externalities - our contribution is to propose a new measure of sorting which is tractable in equilibrium.

We round off the literature review with noting there is another strand of literature on homophily in network formation, see Currarini et al. (2009, 2010); Bramoullé et al. (2012). Their approach, however is different: we use a one period model with strategic link formation while they rely on matching sequences that are dynamic and stochastic. Currarini et al. (2009, 2010) investigate show how differences in community sizes play a role in explaining empirical phenomena, including homophily, assortativity. Bramoullé et al. (2012) investigates the conditions for long run integration of a network.

Paper organization The paper proceeds as follows: Section 2 introduces the model; Section 3 investigates sorting under no externalities; Section 4 analyzes the setting with externalities, focusing on sorting and its potential sub-optimality, and; Section 5 concludes in a discussion of assumptions. All proofs are found in Appendix A.

2 Model

Let $N = \{1, .., n\}$ constitute a set of agents. Each agent $i \in N$ is endowed a fixed measure of *type*, $x_i \in X$ where $X \subset \mathbb{R}$ is the set of realized levels of types for agents in N. Let $\bar{x} = \max X$ and $\underline{x} = \min X$. Let agents' type be sorted descending in their label and let $\mathcal{X} = (x_1, x_2, ..., x_n)$ where $x_l \geq x_{l+1}$ for $l = 1, .., n-1$.

Linking and networks Two agents $i, j \in N$ may *link* if mutually accepted; a link may be broken by both agents. A link between i and j is denoted $ij \in \mu$ where the set μ consists of links and is called a *network*. The set of all networks is denoted $M = \{\mu | \, \mu \subseteq \mu^c\}$ where μ^c is the *complete* network in which all agents are linked.

A *coalition* of agents is a group $t \subseteq N$ such that $t \in T$ where T is the superset of N excluding the empty set. For a given group t define $\mathcal{X}(t)$ as the vector of types ordered descending for agents in t. A *coalitional move* is a set of actions implemented by a coalition that moves the network from one state to another. A move from μ to $\tilde{\mu}$ is *feasible* for coalition t if: added links, $\tilde{\mu}\backslash\mu$ are formed only between members of coalition t; deleted links, $\mu\backslash\tilde{\mu}$ must include a member of coalition t.

Network measures The *neighborhood*, ν, is the set agents whom an agent links to: $\nu_i(\mu) = \{j \in N : ij \in \mu\}$. The number of neighbors is called *degree* and denoted $k_i(\cdot)$ for i. A *path* is a subset of links $\{i_1 i_2, i_2 i_3, ..., i_{l-1} i_l\} \subseteq \mu$ where no agent is reached more than once. The *distance* between two agents i, j in a network is the length of the shortest path between them - this is denoted $p_{ij} : M \to \mathbb{N}_0$; when no path exists then the distance is infinite.

Utility The utility accruing to agent i is denoted u_i. An agent's utility equals benefits less costs, expressed mathematically as $u_i = b_i - c_i$. The aggregate utility is denoted $U(\cdot)$.

We model costs of linking indirectly through an opportunity cost of linking. We do this through a *(degree) quota* on links, κ, which is the maximum number of links for any agent, i.e. for $i \in N$ it holds $k_i(\cdot) \leq \kappa$. We say there is *no linking surplus* when all agents have a degree equal to the degree quota, i.e. $\forall i \in N : k_i = \kappa$.

Benefits to agent i is a weighted sum consisting of two elements; network and individual value:

$$b_i(\mu) = \Sigma_{j \neq i} w_{ij}(\mu) \cdot z_{ij} \tag{1}$$

The network factor, $w_{ij}(\mu)$ is a function of network distance. The individual link value is z_{ij} which measures the personal value to i of linking to j - the value is a function of the two partners' type $z_{ij} = z(x_i, x_j)$. The function z is assumed twice differentiable as well as taking positive and bounded values.[8] Let the *total link value* be defined as the value of linking for the pair, i.e. $Z_{ij} = z_{ij} + z_{ji}$.

In order to derive results a restriction of payoffs is necessary. The essential feature of the total link value for sorting is complementarity in type:[9]

Definition 1. *The link value has* ***supermodularity*** *if* $\frac{\partial^2}{\partial x \partial y} Z(x, y) > 0$ *- this entails:*

$$Z(x, \tilde{x}) + Z(y, \tilde{y}) > Z(x, y) + Z(\tilde{x}, \tilde{y}), \qquad x > \tilde{y},\, \tilde{x} > y. \tag{2}$$

The network components is further restricted in the analysis under externalities in Section 4.

[8] The upper bound rules out an infinite number of links in equilibrium.

[9] Complementarity between type corresponds to cheaper link between same/similar types used in the models of Johnson and Gilles (2000); Jackson and Rogers (2005); Galeotti et al. (2006).

The game framework This paper explores a static setting of one period. Agents' information about payoffs of other agents is complete. Together the players, action, utility and information constitutes a game we will now outline the stability concept for.

Any pair of agents can transfer 'utility' between them. Let a *net-transfer* from agent j to agent i be denoted as $\tau_{ij} \in \mathbb{R}$ such that $\tau_{ij} = -\tau_{ji}$ which implies non-wastefulness of utility. The matrix of net transfers is denoted τ. For each agent i its net-payoff is defined as $s_i = u_i(\mu) + \Sigma\tau_{ij}$. In addition it is assumed that the net-transfer between two agents with a link can only be changed if this change is mutually agreed or the link is broken.

Stability We define network stability using the concept of coalitional moves. A coalition t is *blocking* a network μ with net-transfers τ if there is a feasible coalitional move from network μ to network $\tilde{\mu}$ with $\tilde{\tau}$ where all members in t have higher net-payoff after the move.

We employ two concepts of stability. The first is *strong stability*: this is satisfied for a network if there exist transfers such that no coalition (of any size) may have a feasible move which is profitable for all its members. The second concept, *pairwise (Nash) stability*,[10] is similar but has weaker requirements: it holds when there exist transfers such that no coalitions of at most two agents may block. The sets of stable networks are denoted respectively M^{s-stb} and M^{p-stb}. A further discussion of the stability concepts is found in Section 5.

Our pairwise definition of stability is stricter than that of Jackson and Wolinsky (1996). However, the stricter requirement enables substitution of links (simultaneous deletion and formation) which is a necessary requirement for establishing results in the matching literature.[11]

A noteworthy feature is that strong stability implies pairwise stability; thus any condition valid for pairwise stability also applies to strong stability. In addition without network externalities every pairwise stable network is also strongly stable, see Lemma 1. Note also that any strongly stable network requires efficiency (coalition of all agents can implement any network). Thus we can employ efficiency to derive structure in strongly stable networks.

3 Analysis: no externalities

This brief section analyzes the setting where network externalities are absent and thus indirect connections are irrelevant.

We begin with defining our measure of sorting. The concept of sorting that we employ is a generalization of the sorting when there is a single partner such as Becker's marriage market. The shape of sorting is such that a high type agent has partners which weakly dominate in type when compared partner-by-partner with the partners of a lower type agent. Note the comparison is done over the sorted set of partners type $\mathcal{X}$. The sorting pattern is mathematically defined as:

[10] This is also known as bilateral stability, cf. Goyal and Vega-Redondo (2007)

[11] For instance, pairwise stability coupled with and Becker (1973)'s condition for sorting, i.e. supermodularity, would not imply that sorting be the unique pairwise stable outcome in the marriage market as in Becker (1973).

Definition 2. ***Sorting in type*** *holds in μ if for all pair i, j such that $x_i > x_j$ it holds that:*

$$\mathcal{X}(\nu_i(\mu)/\{j\})_l \geq \mathcal{X}(\nu_j(\mu)/\{i\})_{l+l^*}, \qquad \forall l \in \{1, .., k^*\},$$

where $k^* = \min(k_i(\mu), k_j(\mu))$ and $l^* = \max(k_j(\mu) - k_i(\mu), 0)$.

Our first result is that sorting in type emerges under the same conditions as in Becker (1973) when network externalities are absent:

Proposition 1. *If there is supermodularity and no externalities then for any pairwise stable network there is sorting in type.*

The proof of this proposition follows by establishing that pairwise stable network must be strongly stable without externalities; then we use that strongly stable networks are efficient and show that sorting in type must hold under efficiency.

4 Analysis: externalities

We proceed to a more general context where indirect connections matter for utility. Whenever we allow for externalities we restrict our attention to forms of linking utility.

$$w_{ij}(\mu) = \begin{cases} \delta^{p_{ij}(\mu)-1}, & \text{constant decay,} \\ \mathbf{1}_{=1}(p_{ij}(\mu)) + \delta \cdot \mathbf{1}_{\in[2,\infty)}(p_{ij}(\mu)), & \text{hyperbolic decay,} \end{cases} \tag{3}$$

where $\mathbf{1}_{\in(1,\infty)}(l)$ is the Dirac measure/indicator function of whether $1 < l < \infty$.

The first and more general setting is where utility from connections decays over increasing distance at a constant exponential rate - this corresponds to benefits from linking in the 'connections-model' from Jackson and Wolinsky (1996). The other case is when externalities from indirect connections are discounted equally at any distance if there is a connection, i.e. a finite path length. This second case is referred to as hyperbolic decay and entails that there is no decay beyond that from distance one (linked) to distance two.

The introduction of externalities to our framework implies that the pairwise utility no longer depends only on the total link benefits. As a consequence sorting is not guaranteed to be neither stable nor efficient. The intuition for this is straight forward: externalities entails that the total welfare from sorting is internalized for the pair while the total welfare for indirect connections are not internalized. We see this by inspecting the utility functions. Suppose that $\hat{g}$ is a pairwise deviation such that agents i, j form a link. Then the pairwise total net utility from deviation can be expressed as follows under externalities:

$$u_i(\hat{g}) + u_j(\hat{g}) - u_i(g) - u_j(g) = Z_{ij} + \sum_{k \in \{i,j\},\, l \notin N \setminus \{i,j\}} z_{kl} \cdot \left(\delta^{p_{kl}(\hat{g})} - \delta^{p_{kl}(g)}\right) \tag{4}$$

From the analysis of the previous section we found that in the absence of externalities then sorting prevails; in the above this correspond to the net utility of deviating equaling Z_{ij}. Therefore we see that the total benefits to sorting are preserved for the pair.

The total benefits to all agents that accrues from agents i, j forming a link are:

$$\begin{aligned} U(\hat{\mu}) - U(\mu) &= u_i(\hat{\mu}) + u_j(\hat{\mu}) - u_i(\mu) - u_j(\mu) + \sum_{l \notin N \backslash \{i,j\}} [u_l(\hat{\mu}) - u_l(\mu)] \\ &= u_i(\hat{\mu}) + u_j(\hat{\mu}) - u_i(\mu) - u_j(\mu) + \sum_{l \notin N \backslash \{i,j\}, l' \in N, l' \neq l} Z_{ll'} \cdot (w_{ll'}(\hat{\mu}) - w_{ll'}(\mu)) \end{aligned} \tag{5}$$

Inspection of Equation 5 informs us that the pairwise utility of linking does not capture the aggregate gains from linking; moreover we see that the gains not captured correspond to the indirect benefits that others receive from the deviation. This implies there is a disparity between the pairwise incentives and total welfare: the pairwise incentives captures the full benefits of sorting but not the full gains to lower distances between agents.

4.1 Finite population

We begin with the situation where there are a finite number of agents. Before starting the main analysis of networks under externalities we will define type self-sufficiency. This holds when each type has more agents than the degree quota, see below. Self-sufficiency of types is important for perfect sorting where it is a necessary condition for feasibility when requiring no link surplus. We briefly investigate the situation when type self-sufficiency does not hold, i.e. $n_x \leq \kappa$ for one or more types $x \in X$. Some immediate conclusions are possible to derive:

Definition 3. *A type $x \in X$ is self-sufficient if $n_x > \kappa$.*

Proposition 2. *Suppose there is supermodularity and type self-sufficiency does not hold:*

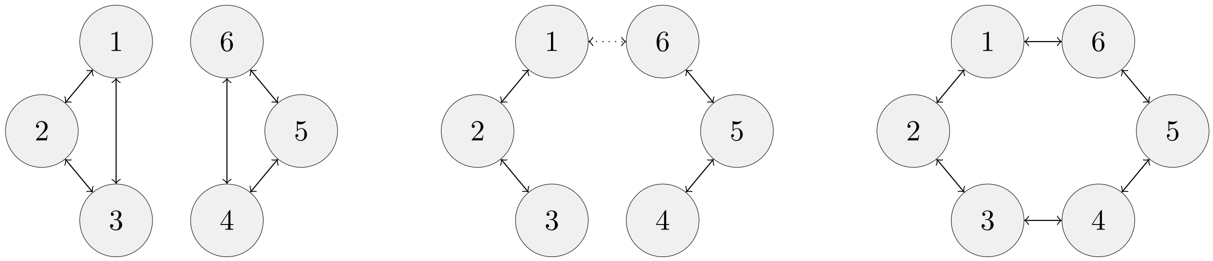
(i) if $n \leq \kappa + 1$ then $\{\mu^c\} = M^{p-stb} = M^{\max U}$;

(ii) if $n > \kappa + 1$ and there are two types where $n_{\bar{x}} = n_{\underline{x}}$, then a network where every agent has $n_{\bar{x}} - 1$ same-type links and $\kappa - n_{\bar{x}} + 1$ cross-type links is stable and efficient.

We move on to examining supoptimal sorting. We will show that a class of sorted networks are suboptimal when introducing network externalities. This sub-optimality holds despite fulfilling the Becker's complementarity, i.e. supermodularity. Our aim is to find equilibria of the kind where each type of agent only link among themselves, we naturally may call perfect sorted:

Definition 4. *A network μ is* ***perfect sorted (in type)*** *if $x_i = x_j$ for any $ij \in \mu$. Denote the set of perfectly sorted networks as M^{p-srt}.*

We note that the remainder of this subsection is restricted in two ways. Firstly, by confining our analysis to the setting where there is type self-sufficiency (i.e. $\forall x \in X : n_x > \kappa$). Secondly,



(A): Segregated network, μ. Pairwise stable if $\delta \leq \bar{\delta}(\hat{Z})$.

(B): Pairwise move by agent 1, 6 who form a link together and remove links 13, 46 respectively.

(C): Connected network, $\tilde{\mu}$. Efficient if $\delta \geq \underline{\delta}(\hat{Z})$.

Figure 1: *Sorted network is stable but inefficient.* The above three networks depict Example 1. The network in (A) is pairwise (Nash) stable for some parameters and the network in (B) is the only kind of feasible deviation. The network in (C) is an efficient network.

this subsection exclusively investigates the case with hyperbolic decay as it provides for more intuitive and more immediate results without restrictions on the network. As noted earlier, a more general exposition is found in Supplementary Appendix B.

We begin our analysis by exhibiting a simple illustration of the problem. Let excessive sorting refer to a network with perfect sorting according to type where the segregated groups could collectively benefit from connecting. However, they fail to connect as incentives do not internalize externalities under pairwise network formation. The relevant set of networks are those networks which are perfectly sorted but within each type all agents are connected. Let networks which are perfectly sorted and connected among each type be denoted:

$$M^{p-srt:conn} = \{\mu \in M^{p-srt} \,|\, \forall i, j \in N, x_i = x_j : p_{ij}(\mu) < \infty\}$$

The are interested in showing there exists an open region of thresholds $(\underline{\delta}, \bar{\delta})$ such that if the level of externalities is within the region then any network in $M^{p-srt:conn}$ where there is no degree surplus is pairwise (Nash) stable for some transfers τ but inefficient. An introduction to this problem is found in Example 1 below in a stylized, simple manner. The example is graphically represented in Figure 1.

Example 1. *There are six agents; three of high type (1,2,3) and three of low type (4,5,6). Moreover, there is supermodularity, degree quota of two (κ=2) and constant decay. Define two networks: a network with perfect sorting, $\mu = \{12, 13, 23, 45, 46, 56\}$, see Figure 1.A; a network which is bridged of μ, defined as $\tilde{\mu} = \{12, 23, 34, 45, 56, 61\}$, see Figure 1.C. We show in this example that for a range of decay-factors that μ is pairwise stable yet suboptimal.*

In this setup there is a unique move which is both feasible and payoff relevant.[12] *This move*

[12]Under pairwise stability at most one link can be formed in a single move. Without transfers all formed links have value and thus deletion of a link always leads to a loss. Thus only coalitional moves where new links are formed can be valuable. For network μ all links to same types are already formed. Thus only forming a link with the other types.

consists of forming a link across types when both delete a link. Such a move could be agents 1,6 forming a link while deleting links to 3 and 4 which we denote $\hat{\mu} = \mu \cup \{16\} \backslash \{13, 46\}$ *and seen as Figure 1.B. Benefits for agents 1 and 6 from network* μ *and deviating from it are:*

$$\begin{aligned} u_1(\hat{\mu}) + u_6(\hat{\mu}) &= (1+\delta) \cdot [z(\bar{x}, \bar{x}) + z(\underline{x}, \underline{x})] + [1+2\delta] \cdot [z(\bar{x}, \underline{x}) + z(\underline{x}, \bar{x})], \\ &= (1+\delta) \cdot \tfrac{1}{2} \cdot [Z(\bar{x}, \bar{x}) + Z(\underline{x}, \underline{x})] + [1+2\delta] \cdot Z(\bar{x}, \underline{x}), \\ u_1(\mu) + u_6(\mu) &= 2 \cdot [z(\bar{x}, \bar{x}) + z(\underline{x}, \underline{x})] = Z(\bar{x}, \bar{x}) + Z(\underline{x}, \underline{x}). \end{aligned}$$

The condition for deviating to $\hat{\mu}$ *not being pairwise profitable is* $u_1(\mu) + u_6(\mu) > u_1(\hat{\mu}) + u_6(\hat{\mu})$*; this is sufficient for pairwise stability due to payoff symmetry in* μ *and no transfers.*

We now turn to deriving condition for when segregating is inefficient. The aggregate benefits over all agents of the two networks μ *and* $\tilde{\mu}$ *is expressed below in the two equations.*

$$\begin{aligned} U(\tilde{\mu}) &= (2+\delta) \cdot [Z(\bar{x}, \bar{x}) + Z(\underline{x}, \underline{x})] + [2+7\delta] \cdot Z(\bar{x}, \underline{x}), \\ U(\mu) &= 3 \cdot [Z(\bar{x}, \bar{x}) + Z(\underline{x}, \underline{x})]. \end{aligned}$$

Sorting is inefficient when: $U(\mu) < U(\tilde{\mu})$*. The two inequalities governing pairwise stability and inefficiency has the following positive solution:*

$$\begin{aligned} \underline{\delta}(\hat{Z}) &= \tfrac{\hat{Z}}{\hat{Z}+\frac{5}{2}}, \\ \bar{\delta}(\hat{Z}) &= \tfrac{\hat{Z}}{\hat{Z}+1}, \end{aligned} \qquad \hat{Z} = \tfrac{Z(\bar{x},\bar{x})+Z(\underline{x},\underline{x})}{2Z(\bar{x},\underline{x})} - 1,$$

where $\underline{\delta}$ *and* $\bar{\delta}$ *are thresholds for respectively when network* μ *becomes inefficient and unstable when* δ *increases.*

The example above demonstrates that sorting can be inefficient when there are network effects despite there being complementarity in type, i.e. supermodular link values. The inefficiency stems from a novel source - the pairwise formation of links. The intuition is that under pairwise deviation the two agents do not internalize the total value created for other agents number of indirect links between a high and a low agent. Note that the above example has a close correspondence to Propositions 1, 6 from de Martí and Zenou (2017) as their results also holds only for cliques with less than five of each type.[13]

We proceed with a generalization of the example above which holds for various structures of the subnetworks within types and for multiple types. As we proceed with our generalization we introduce two new concepts to convey the results.

Let the subset of agents who link in μ be denoted $N(\mu) = \{i \in N : \exists j \text{ s.t.} ij \in \mu\}$. A *component*, $\tilde{\mu}$ of network μ is a subnetwork (i.e. $\tilde{\mu} \subseteq \mu$) where for any agent $i \in N(\tilde{\mu})$ it holds that: agent i is connected in $\tilde{\mu}$ to all other agents $j \in N(\tilde{\mu})$, and; for all $j \in N$ if $ij \in \mu$ then

[13] See the literature review for a thorough discussion of similarities and differences.

$ij \in \tilde{\mu}$. A perfectly sorted network is *locally connected* if there is one component for each type; the set of networks which are perfect sorted and locally connected is denoted $M^{p-srt:conn}$.

Our next result is to extend the above example to a less restrictive setting. The extension holds for a class of sorted networks fulfilling certain conditions, see below.

Definition 5. *Let $\hat{M}$ be a set networks where links are only between same-type agents, all agents use the full quota of available links and all agents of same type are connected, i.e.*

$$\hat{M} = M^{p-srt:conn} \cap M^{no-surpl.}$$

Theorem 1. *Suppose there is supermodularity, hyperbolic decay, more than one partner is allowed and type self-sufficiency then:*

(i) any network in $\hat{M}$ is pairwise stable when $\delta \leq \bar{\delta}$, i.e. $\hat{M} \subseteq M^{p-stb}_{\delta \leq \bar{\delta}}$;

(ii) any network in $\hat{M}$ is inefficient when $\delta > \underline{\delta}$, i.e. $(\hat{M} \cap M^{\max U}_{\delta > \underline{\delta}}) = \emptyset$;

(iii) the set of networks $\hat{M}$ is non-empty when the count of agents of each type has even number greater than the degree quota, i.e. $\hat{M} \neq \emptyset$ if for all $x \in X : (\kappa \cdot n_x) \in 2\mathbb{N}$;

where threshold $\underline{\delta}, \bar{\delta}$ are bounded as follows:

$$\begin{aligned} \underline{\delta} &\leq \min_{x,\tilde{x} \in X} \left(\frac{\hat{Z}_{x,\tilde{x}}}{\hat{Z}_{x,\tilde{x}} + \frac{1}{2} n_x n_{\tilde{x}}} \right), & \hat{Z}_{x,\tilde{x}} &= \frac{Z(x,x)+Z(\tilde{x},\tilde{x})}{2Z(x,\tilde{x})} - 1 \\ \bar{\delta} &= \min_{x,\tilde{x} \in X} \left(\frac{\hat{Z}_{x,\tilde{x}}}{\hat{Z}_{x,\tilde{x}} + \max(n_x, n_{\tilde{x}}) - |n_x - n_{\tilde{x}}| \cdot \hat{z}_{x,\tilde{x}}} \right), & \hat{z}_{x,\tilde{x}} &= \frac{z(x,\tilde{x})}{Z(x,\tilde{x})}. \end{aligned}$$

The theorem above is applicable to numerous settings where social networks are formed. If for example schools or firms pre-sort individuals according to talent or otherwise this may lead to a sorted network which is stable but suboptimal. A specific example could be a school where sorting by academic performance type is common in many countries and is often known as tracking. In such cases the sorting induced by the institution could lead to a stable network with no linking across despite links across having potential to raise welfare.

Note that the existence property in theorem (iii) requires either even link quota or even number of agents for each type. The reason is technical and is that if both of these are not met then the total demand for links of the same type is an uneven when there is no link surplus but each links takes up a capacity of two and thus must be an even number; the implication is that perfect sorting and no link surplus is incompatible when this even number condition is not met. We discuss the choice of equilibrium concept in the discussion found in Section 5.

An alternative version of the above theorem under constant decay can be found in Appendix B in Theorem 3. For both Theorems 1 and 3 one can view the conditions for stability of sorting as generalizing not only Example 1 but also Propositions 1.ii and 4.iii from de Martí and Zenou (2017). These special cases of sorting correspond to a situation with two types and all agents of a given type link with one another, i.e. when $n_{\bar{x}} = n_{\underline{x}} = \kappa + 1$ and $|X| = 2$. The other part of the two theorems is on inefficiency of sorting and generalizes Example 1 and Proposition 6 from de Martí and Zenou (2017). It advances their proposition by removing the restriction to two types and linking between all same type agents as well as doing away with the limitation of having very few agents.[14]

Two types

When there are only two types of agents then we can strengthen our results further by establishing that sorting is the unique pairwise stable outcome for moderate levels of externalities. In addition we can show that the threshold for inefficiency now governs which network whether $\hat{M}$ or another class of networks with linking across types. This other class of networks that are described by having the minimum necessary links to connected the two different types of agents:

Definition 6. *Let there be two types. The set of cross-type bridged networks consist of a perfect sorted network μ^s where two links across types are added:*

$$M^{bridge} = \{\mu \in M \,|\, \exists \hat{\mu} \in \hat{M} : \mu = \hat{m} \cup \{ij, i'j'\}, x_i = x_{i'}, x_j = x_{j'}, x_i = x_j\}$$

Define the subset of cross-type bridged networks where each subnetwork of type is connected and without linking surplus:

$$\bar{M} = \{\mu \in M^{bridge} \,|\, \forall x \in X : |\mu_x| = \frac{n_x \kappa}{2} - 1 \; and \max_{i \neq j} p_{ij}(\mu) < \infty\},\; \mu_x = \{ij \in \mu | x_i = x_j\}$$

It is important to understand that in our setup a cross-type bridged network requires two links to be established across types. This is a technical condition stemming from the fact that reducing the number of links among same type by one frees up the capacity to establish a link by two agents; as a consequence it is possible for two links across types to be established. Using both of these possible links is important for establishing efficiency. It will turn out to also be important in the investigation of policy, see in Proposition 3. This stems from the fact that when compensating one agent to establish a link across types the agent to which it has deleted a link has an incentive to form a new link which will potentially destabilize subnetwork within the subnetwork of their type. We discuss this assumption and how it relates to our choice of model in Section 5.

Theorem 2. *Suppose conditions for Theorem 1 are satisfied and suppose that $|X| = 2$,*

(i) if $\delta < \bar{\delta}$ then a network is in $\hat{M}$ iff. it is pairwise stable, i.e. $\hat{M} = M^{p-stb}_{\delta < \bar{\delta}}$;

(ii) if $\delta < \underline{\delta}$ then a network is in $\hat{M}$ iff. it is efficient, i.e. $\hat{M} = M^{\max U}_{\delta > \underline{\delta}}$.

[14] As noted in the literature review if there is an equal number of agents for each type then Proposition 6 in de Martí and Zenou (2017) is valid only for at most five agents of each type (i.e. a total of ten).

(iii) if $\delta > \underline{\delta}$ then a network is in $\bar{M}$ iff. it is efficient, i.e. $\bar{M} = M^{\max U}_{\delta > \underline{\delta}}$.

where the threshold for efficiency, $\underline{\delta} = \frac{\hat{Z}_{\underline{x},\bar{x}}}{\hat{Z}_{\underline{x},\bar{x}} + \frac{1}{2} n_{\underline{x}} n_{\bar{x}}}$.

A visualization of the computed thresholds of externalities when there are two types is found in Figure 2. The thresholds are computed for varying population size and varying strength of complementarity. The upper part of the figure keeps the population size fixed while lower ones keep the complementarity strength fixed. From inspection it is evident that the connection thresholds both approximately follow power-laws in the number of agents and strength of complementarity.

The remainder of this subsection will sketch a policy intervention that can mitigate the problem of suboptimal sorting by improving welfare through encouraging connection. These interventions can be seen more generally as a design problem where the policy maker intervenes to induce a network that produces higher welfare. The tool that the policy maker employs is providing incentives to agents for forming specific links. Define a *link-contingent contract* as a non-negative transfer $\mathcal{C}_{ij}$ to i for linking with another agent j. Denote the vector of link-contingent contract as $\mathcal{C}$.

Definition 7. *Let a network $\tilde{\mu}$ be* ***implementable*** *from μ, τ given $\mathcal{C}$ if there exist a sequence of tuples $(\mu_0, \tau_0), .., (\mu_l, \tau_l)$ where $\mu_0 = \mu$, $\mu_l = \tilde{\mu}$ and $\tau_l = \tau$ such that: for $q = 1, .., l$ from μ_{q-1} to μ_q is a feasible pairwise move which increases the pair's net-utility most given $\mathcal{C}$, and; $\tilde{\mu}$ is pairwise stable given τ_q and $\mathcal{C}$.*

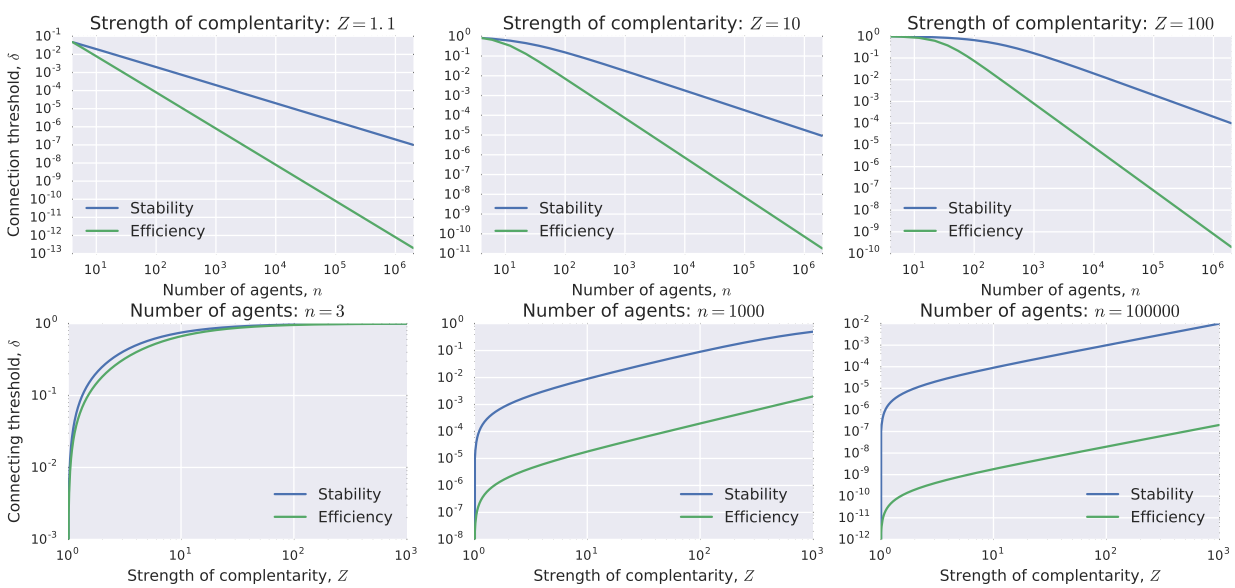


Figure 2: *Thresholds for connecting.* Visualization of thresholds for connecting from Theorem 1. The upper part varying sizes of populations and fixed strength of complementarity. The lower part has has varying strength of complementarity and fixed populations size. It is assumed there are two types with an equal number of agents.

Proposition 3. *Suppose that conditions for suboptimal sorting from Theorem 1 are valid and there are two types then a policy maker can implement an efficient network from $\bar{M}$ when $\delta \in (\underline{\delta}, \bar{\delta})$ from a network in $\hat{M}$.*

Note that the individual compensation for paid to agents for connecting to others may not be equal. In particular the pay may also depend on the types. This is the case when there is both supermodularity and monotonicity in Z then agents may receive compensation increasing their type. This would be the case in a sorted suboptimal network with no transfers where components have same number of agents and are isomorphic to another.

4.2 Externalities - infinite population

We finalize this section by investigating what pattern of linking is exhibited when the count of agents becomes asymptotic infinite. In this large matching market we examine asymptotic perfect sorting, i.e. when measured share of links to same type agents converges to one. Note that we use a stronger equilibrium concept, strong stability, which allows for coordination between coalitions of any size to coordinate.

Definition 8. *Let* ***asymptotic perfect sorting*** *hold for a sequence of networks sets $\tilde{M}_n$ if for any network $\mu \in \tilde{M}_n$ where $n \to \infty$ it holds that $|\{ij \in \mu : x_i = x_j\}|/|\mu| \simeq 1$.*

Define *asymptotic independence* as $\delta < (\kappa - 1)^{-1}$. For large matching markets the sufficient conditions for asymptotic perfect sorting to emerge in strongly stable networks are:

Proposition 4. *If there is supermodularity, a degree quota and constant decay with asymptotic independence then there is asymptotic perfect sorting for strongly stable networks.*

The result above demonstrates that the availability of many agents for linking induces perfect sorting in strongly stable networks. It demonstrates the same prediction as the conclusion of Becker (1973) for the marriage market model but holds in the presence of externalities with constant decay.

For derivation of the result we use that an implication of strong stability is that efficiency must hold. Therefore by establishing that efficiency requires perfect sorting it follows that it most hold for any strongly stable network. Although efficiency is a unique property for strong stability (and does not hold for weaker concepts) it can be argued that strong stability should be seen as a refinement with desirable properties which makes it more likely when it exists.[15]

We conclude this section by noting that we may interpret the result on sorting in infinite population differently; there is no loss of sorting when there are many agents.

[15] In some circumstances the existence of contracts where an agent may subsidize or penalize another agent's link formation with alternative agents may imply that strong stability even if contracts were limited to being pairwise specified, cf. Bloch and Jackson (2007).

5 Concluding discussion

We have extended the assortative matching framework to a setting of networks. We have show that in general Becker (1973) condition for sorting are still essential for stability but that they are insufficient for efficiency (when there is a finite population). We have sketched a policy that can help overcome this issue.

We have chosen to model cost implicitly via a degree quota in order to have comparability with the matching literature. We expect, however, that our results should easily translate to the standard connections model of Jackson and Wolinsky (1996). In this other setup we expect that the intuition should transfer when limiting the number sub-networks within types to be either cliques or stars, as in de Martí and Zenou (2017). One advantage of translating the setup to a linear cost framework of the standard networks literature would be that the technical assumption of either even degree or an even number of agents for each type would not be necessary. Under hyperbolic decay one would also get a more natural efficient policy solution requiring only a single agent of each type to bridge the bridge their respective subnetworks.

Our analysis is based on strict assumptions which we now review. We avoid discussing sorting under search as there is a large literature e.g. Shimer and Smith (2000). One severe caveat with our analysis, and stable networks in general, is that these networks may not exist. The classic example is the room-mate problem, cf. Gale and Shapley (1962). Furthermore, the gross substitutes conditions from Kelso and Crawford (1982) which ensures existence of stable matchings in related settings are not satisfied in our setting with externalities.[16] Nevertheless, by changing equilibrium concept we expect to that some of the lack of existence could be solved. One approach is using farsighted stability as in Chwe (1994); Dutta et al. (2005). Another approach is using some approximative equilibrium concept e.g. cost of stability (the necessary payments to induce stability) from Bachrach et al. (2009).

There are also a number of restrictive assumptions on payoff. The most crucial assumptions are payoff separability and fixed structure of externalities. However, the results should be robust, for instance to introduction of utility from triads etc. which is common in the economic literature on networks. Another critical assumptions is supermodularity along with perfect transferability. Nevertheless, as mentioned in the introduction, these two assumptions can be replaced by monotonicity in individual link value and perfect non-transferability, which is also in line with some research on peer effects.[17]

References

Baccara, M., Yariv, L. 2013. Homophily in peer groups. American Economic Journal: Microeconomics, 5, 69–96.

[16] The lack of gross substitutes is due to the fact that a change in one active link can imply a change the value of other links. This fact will violate gross substitutes.

[17] Or more broadly by generalized increasing in differences from Legros and Newman (2007).

Bachrach, Y., Elkind, E., Meir, R., Pasechnik, D., Zuckerman, M., Rothe, J., Rosenschein, J. S. 2009. The cost of stability in coalitional games. In International Symposium on Algorithmic Game Theory, 122–134, Springer.

Bala, V., Goyal, S. 2000. A noncooperative model of network formation. Econometrica, 68, 1181–1229.

Becker, G. S. 1973. A Theory of Marriage: Part I. Journal of Political Economy, 81, 813–846.

Bloch, F., Jackson, M. O. 2007. The formation of networks with transfers among players. Journal of Economic Theory, 133, 83–110.

Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., Rogers, B. W. 2012. Homophily and long-run integration in social networks. Journal of Economic Theory, 147, 1754–1786.

Chwe, M. S.-Y. 1994. Farsighted coalitional stability. Journal of Economic theory, 63, 299–325.

Currarini, S., Jackson, M. O., Pin, P. 2009. An economic model of friendship: Homophily, minorities, and segregation. Econometrica, 77, 1003–1045.

Currarini, S., Jackson, M. O., Pin, P. 2010. Identifying the roles of race-based choice and chance in high school friendship network formation. Proceedings of the National Academy of Sciences, 107, 4857–4861.

Durlauf, S. N., Seshadri, A. 2003. Is assortative matching efficient? Economic Theory, 21, 475–493.

Dutta, B., Ghosal, S., Ray, D. 2005. Farsighted network formation. Journal of Economic Theory, 122, 143–164.

Farrell, J., Saloner, G. 1986. Installed base and compatibility: Innovation, product preannouncements, and predation. The American economic review, 940–955.

Farrell, J., Scotchmer, S. 1988. Partnerships. The Quarterly Journal of Economics, 103, 279–297.

Gale, D., Shapley, L. S. 1962. College admissions and the stability of marriage. The American Mathematical Monthly, 69, 9–15.

Galeotti, A., Goyal, S., Kamphorst, J. 2006. Network formation with heterogeneous players. Games and Economic Behavior, 54, 353–372.

Golub, B., Jackson, M. O. 2012. How Homophily Affects the Speed of Learning and Best-Response Dynamics. The Quarterly Journal of Economics, 127, 1287–1338.

Goyal, S., Vega-Redondo, F. 2007. Structural holes in social networks. Journal of Economic Theory, 137, 460–492.

Jackson, M. O., Rogers, B. W. 2005. The economics of small worlds. Journal of the European Economic Association, 3, 617–627.

Jackson, M. O., Wolinsky, A. 1996. A Strategic Model of Social and Economic Networks. Journal of Economic Theory, 71, 44–74.

Johnson, C., Gilles, R. P. 2000. Spatial social networks.. Review of Economic Design, 5.

Katz, M. L., Shapiro, C. 1985. Network externalities, competition, and compatibility. The American economic review, 75, 424–440.

Kelso, A. S. J., Crawford, V. P. 1982. Job matching, coalition formation, and gross substitutes. Econometrica: Journal of the Econometric Society, 1483–1504.

Klaus, B., Walzl, M. 2009. Stable many-to-many matchings with contracts. Journal of Mathematical Economics, 45, 422–434.

Koopmans, T. C., Beckmann, M. 1957. Assignment problems and the location of economic activities. Econometrica: journal of the Econometric Society, 53–76.

Kremer, M. 1993. The O-ring theory of economic development. The Quarterly Journal of Economics, 108, 551–575.

Legros, P., Newman, A. F. 2002. Assortative Matching in a Non-Transferable World. SSRN Scholarly Paper ID 328460, Social Science Research Network, Rochester, NY.

Legros, P., Newman, A. F. 2007. Beauty Is a Beast, Frog Is a Prince: Assortative Matching with Nontransferabilities. Econometrica, 75, 1073–1102.

de Martí, J., Zenou, Y. 2017. Segregation in friendship networks. The Scandinavian Journal of Economics, 119, 656–708.

McPherson, M., Smith-Lovin, L., Cook, J. M. 2001. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology, 27, 415–444.

Miritello, G., Moro, E., Lara, R., Martinez-Lopez, R., Belchamber, J., Roberts, S. G., Dunbar, R. I. 2013. Time as a limited resource: Communication strategy in mobile phone networks. Social Networks, 35, 89–95.

OECD. 2012. Equity and Quality in Education-Supporting Disadvantaged Students and Schools. OECD Publishing.

Pycia, M. 2012. Stability and preference alignment in matching and coalition formation. Econometrica, 80, 323–362.

Reardon, S. F., Bischoff, K. 2011. Income Inequality and Income Segregation 1. American Journal of Sociology, 116, 1092–1153.

Schelling, T. C. 1969. Models of segregation. The American Economic Review, 59, 488–493.

Schelling, T. C. 1971. Dynamic models of segregation. Journal of mathematical sociology, 1, 143–186.

Schwartz, C. R., Mare, R. D. 2005. Trends in educational assortative marriage from 1940 to 2003. Demography, 42, 621–646.

Shimer, R., Smith, L. 2000. Assortative matching and search. Econometrica, 68, 343–369.

Ugander, J., Karrer, B., Backstrom, L., Marlow, C. 2011. The anatomy of the facebook social graph. arXiv preprint arXiv:1111.4503.

A Proofs

Lemma 1. *In the absence of network externalities then the set of strongly stable networks is equivalent to the set of pairwise stable networks, i.e.,* $M_{\delta=0}^{p-stb} = M_{\delta=0}^{s-stb}$.

Proof. By definition it holds that $M_{\delta=0}^{s-stb} \subseteq M_{\delta=0}^{p-stb}$, thus we need to show that $M_{\delta=0}^{p-stb} \subseteq M_{\delta=0}^{s-stb}$ to prove the claim. This claim is shown using similar to arguments to Klaus and Walzl (2009)'s Theorem 3.i.

Let μ with associated contracts τ be a network which is blocked by a coalition. It will be shown that for every coalition $t \in T$ that blocks, within the coalition there is a subset of no more than two members that also wishes to block the network. Let $\tilde{\mu}$ be the alternative network that the blocking coalitions implements through a feasible coalition move and τ be the transfers associated with $\tilde{\mu}$.

It is always possible to partition the set of deleted links $\mu\backslash\tilde{\mu}$ into two: (i) a subset denoted $\hat{\mu}$ where for each link ij that can be deleted where one of the two partners can benefit, i.e. it holds that either $z_{ij} + \tau_{ij} - [c_i(\mu) - c_i(\mu\backslash ij)] < 0$ or $z_{ji} + \tau_{ji} - [c_j(\mu) - c_i(\mu\backslash ij)] < 0$; (ii) a subset denoted $\breve{\mu}$ where for each link ij neither of the previous two inequalities are satisfied.

Suppose that the first partition is non-empty, i.e. $\hat{\mu} \neq \emptyset$. However, as deleting links can be done by a single agent on its own then the move only takes needs the coalition of that agent to delete the link. Thus any part of a coalitional move that only involves profitably removing links can be performed in parts by a coalition with a single agent - therefore this move is also a pairwise block.

Thus it remains to be shown that the remaining part of coalitional move also can be performed as a pairwise block, i.e. when forming $\tilde{\mu}\backslash\mu$ and deleting $\breve{\mu}$. This part of the coalitional move must entail forming links as no links can be deleted profitably. The set of formed links $\tilde{\mu}\backslash\mu$ can be partitioned into a number of $|\tilde{\mu}\backslash\mu|$ feasible submoves of adding a single link while deleting links by each of the agents i and j who form a link. The feasibility for each of the partitioned moves is always true when there is a cost function as moves are unrestricted. It is now argued that each of the partitioned moves are feasible when there is a degree quota. If the network $\mu \cup ij$ is feasible then the move of simply adding the link is feasible. If $\mu \cup ij$ is not feasible, then agents i and j can delete at most one link each and if both μ and $\tilde{\mu}$ are feasible then this also feasible as the degree quota is kept.

For the coalitional move to $\tilde{\mu}$ it must be that at least at least one link among the implemented links $\tilde{\mu}\backslash\mu$ has a strictly positive value that exceeds the loss from deleting at most one link for each of two agents forming the link. This follows as it is known that deleting one or more links cannot add any value and thus must have weakly negative value and that by definition the total value to the blocking coalition must be positive. As every one of the partitioned moves is feasible, it follows that for every coalitional move there are two agents who can form link while potentially destroying current links and both be better off. In other words, for every coalition that blocks, there is a pairwise coalition that blocks. □

Proof of Proposition 1. Suppose the claim is false. Let q be the lowest index for which the condition fail: for all $l < q$ it holds that $\mathcal{X}(\nu_i(\mu)/\{j\})_l \geq \mathcal{X}(\nu_j(\mu)/\{i\})_{l+l^*}$ where $l^* = \max(k_j(\mu) - k_i(\mu), 0)$. Thus there are two agents i', j' such that:

$$\begin{aligned} x_{j'} &= \mathcal{X}(\nu_j(\mu))_q, \quad j' \in (\nu_j(\mu)\backslash(\nu_i(\mu) \cup \{i\})), \\ x_{i'} &< \mathcal{X}(\nu_j(\mu))_q, \quad i' \in (\nu_i(\mu)\backslash(\nu_j(\mu) \cup \{j\})). \end{aligned}$$

Recall $k^* = \min(k_i(\mu), k_j(\mu))$. The argument why there must exist an agent i' in $\nu_j(\mu)$ but not in $(\nu_j(\mu) \cup \{j\})$ is that $|\{\iota \in \nu_i(\mu) : x_\iota < x_{j'}\}| > |\{\iota \in \nu_j(\mu) : x_\iota < x_{j'}\}|$. This follows as by construction it holds that $|\{\iota \in \nu_i(\mu) : x_\iota < x_{j'}\}| = k^* - q + 1$ and $|\{\iota \in \nu_j(\mu) : x_\iota < x_{j'}\}| \leq k^* - q$.

The agents are such that $x_i > x_j, x_{i'} < x_{j'}$ as well as $ij', ji' \notin \mu$. However, this fact implies that there is a violation of strong stability: agents i, i', j, j' can deviate by destroying $\{ij, i'j'\}$ and forming $\{ij', i'j\}$ and thus increase payoffs due to supermodularity (cf. Equation 2). From Lemma 1 it follows that pairwise stability is also violated if strong stability is violated. ■

Proof of Proposition 2: Condition (i) follows from the fact that it is possible for every agent to be linked with one another. Moreover every link adds value. Thus as a consequence every link can be formed and will add value both for the pair forming and it at the aggregate level; thus the unique pairwise and efficient outcome must be the complete network.

We move on to proving condition (ii). Suppose μ is a network where every agent has $n_{\bar{x}} - 1$ same-type links and $\kappa - n_{\bar{x}} + 1$ cross-type links.

Efficiency of μ follows from three facts. Firstly, μ the maximum distance of 2 between any two agents as all same-type links are active and all agents have at least one cross-type link; thus the potential benefits from indirect connections are maximized (both for constant and hyperbolic decay). Secondly, the number of same type links are maximized for all agents and this will maximize the benefits from direct links; thus there must exactly $n_{\bar{x}} - 1$ same type links. Finally, there can be no link surplus because violation there exist a network where every agent has $n_{\bar{x}} - 1$ same-type links and $\kappa - n_{\bar{x}} + 1$ cross-type links and thus has no link surplus; this must have strictly higher aggregate utility as every direct link increases utility.

Stability of μ follows from reviewing the feasible deviations. Let there be no transfers between any agents. Firstly, deleting one or more links is profitable as it lowers the agents own welfare. Secondly, forming a link requires deletion of one or more links by both agents. Deleting more links than one will lower the utility this only the deviations with deletion of a single link are relevant to consider - this corresponds to substitution of a link. Substituting either a same type link for another same type link or a cross type link for another cross type provides no change of utility to the pair of agents deviating. Substituting a cross type link for a same type link is not feasible. Substituting a same type link for a cross type link will lower the utility as the indirect benefits are unchanged but the direct benefits must be lower on aggregate due to supermodularity. ■

Lemma 2. *For every κ, n such that $n > \kappa$ and $n \cdot \kappa$ is even there exists a network $\mu_{n,\kappa}$ where all agents have exactly κ neighbors. Moreover, if $\kappa \geq 2$ then $\mu_{n,\kappa}$ is connected.*

Proof. Suppose n is even. Let % be the modulus operator. We can construct the following networks.

$$\begin{aligned}
\hat{\mu}_{n,\kappa} &= \left\{ij : i \in \{1, ..., \tfrac{n}{2}\}, \, j \in \{(\tfrac{n}{2} + i \,\%\, \tfrac{n}{2}), ..., (\tfrac{n}{2} + [i + \kappa - 1] \,\%\, \tfrac{n}{2}\})\right\}, \ \kappa \leq \tfrac{n}{2}, \\
\tilde{\mu}_{n,\kappa} &= \begin{cases} \hat{\mu}_{n,\kappa}, & \kappa \leq \frac{n}{2}, \\ \mu_c \backslash \hat{\mu}_{n,n-\kappa-1}, & \kappa > \frac{n}{2}. \end{cases}
\end{aligned}$$

Letting $\mu_{n,\kappa} = \tilde{\mu}_{n,\kappa}$ is sufficient for n is even. When n is odd we know that κ is even and thus we can use the following amended procedure instead:

$$\iota_{n,\kappa}(\iota) = \begin{cases} \frac{n-1}{2} + \iota, & \kappa \leq \frac{n-1}{2} \\ \frac{n-1}{2} + (\iota + \kappa) \,\%\, \frac{n-1}{2}, & \kappa > \frac{n-1}{2} \end{cases}$$

$$\mu_{n,\kappa} = \tilde{\mu}_{n-1,\kappa} \backslash \left\{ij : i \in \{1, .., \tfrac{\kappa}{2}\}, \, j = \iota_{n,\kappa}(i)\right\} \cup \left\{ij : i = n, \, j \in (\cup_{\iota\{1,..,\frac{\kappa}{2}\}} \{\iota, \iota_{n,\kappa}(\iota)\})\right\}$$

We now show that if $\kappa \geq 2$ it follows that $\tilde{\mu}_{n,\kappa}$ is connected. Assume that n is even and suppose $\kappa \leq \frac{n}{2}$; for any $i \in N : i < \frac{n}{2}$ where $i' = i + 1$ and let $j = \frac{n}{2} + i + 1$ where $ij, i'j \in \tilde{\mu}_{n,\kappa}$; thus for all $i, i' \in \{1, .., \frac{n}{2}\}$ it holds that $p_{ii'}(\tilde{\mu}_{n,\kappa}) < \infty$. In addition, as for any $i \in N : i \leq \frac{n}{2}, j = \frac{n}{2} + i$ it holds that $ij \in \tilde{\mu}_{n,\kappa}$ it follows that $\tilde{\mu}_{n,\kappa}$ is connected. If instead $\kappa > \frac{n}{2}$ then by construction $ii' \in \tilde{\mu}_{n,\kappa}$ if either $\max(i, i') \leq \frac{n}{2}$ or $\min(i, i') > \frac{n}{2}$ as $ii' \notin \hat{\mu}_{n,n-\kappa-1}$. Moreover, for $i \in N : i < \frac{n}{2}$ and $j = \frac{n}{2} + (i + \kappa) \,\%\, \frac{n}{2}$ it holds that $ij \notin \hat{\mu}_{n,n-\kappa-1}$; thus $ij \in \tilde{\mu}_{n,\kappa}$. Therefore $\tilde{\mu}_{n,\kappa}$ must be connected.

Assume instead that n is odd. By the above argument there are at least two connected subnetworks consisting of agents in $\cup_{\iota\{1,..,\frac{\kappa}{2}\}}\{\iota,\iota_{n,\kappa}(\iota)\}$ and agents who are connected through agent, n, i.e. $N\backslash(\cup_{\iota\{1,..,\frac{\kappa}{2}\}}\{\iota,\iota_{n,\kappa}(\iota)\})$. If $\kappa\leq\frac{n-1}{2}$ where $i=\frac{\kappa}{2}$, $i'=\frac{\kappa}{2}+1$ and $j=\frac{n-1}{2}+\frac{\kappa}{2}+1$ then $ij,i'j\in\tilde{\mu}_{n,\kappa}$ and thus $\tilde{\mu}_{n,\kappa}$ is connected. If $\kappa>\frac{n-1}{2}$ where $i=\frac{\kappa}{2}$, $i'=\frac{\kappa}{2}+1$ and $j=\frac{n-1}{2}+(\iota+\kappa+1)\,\%\,\frac{n-1}{2}$ then $ij,i'j\in\tilde{\mu}_{n,\kappa}$ and thus $\tilde{\mu}_{n,\kappa}$ is connected. □

Lemma 3. *Suppose that* $\min_{x\in X}n_x>\kappa$, $\kappa\geq 2$. *If* $\exists i\in N$ *such that:*
a) $|\{i'\in\nu_i(\mu):x_{i'}=x_i\}|\leq n_x-2$;
b) $\min_{i'\in N_x\backslash\nu_i(\mu)}k_{i'}(\mu)=\kappa$, *and;*
c) $\max_{i'\in N_x\backslash\nu_i(\mu)}|\{i''\in\nu_{i'}(\mu):x_{i''}\neq x\}|=0$;
then $\exists i',i''\in\mu$ *such that* $i',i''\notin\nu_i(\mu)$ *and* $p_{i'i''}(\mu\backslash\{i'i''\})<\infty$

Proof. Suppose that for $i\in N$ the conditions a)-c) are met but the lemma is not true. If $i'\in N_x$ and $ii'\notin\mu$ then there must exist some $i''\in N_x$ such that $i'i''\in\mu$ and $i''\notin\nu_i(\mu)$ due to conditions a)-c). If $p_{i'i''}(\mu\backslash\{i'i''\})<\infty$ then the proof is terminated so we must assume $p_{i'i''}(\mu\backslash\{i'i''\})=\infty$.

As $p_{i'i''}(\mu\backslash\{i'i''\})=\infty$ then the network $\mu\backslash\{i'i''\}$ has two components, $\mu',\mu''\subseteq\mu\backslash(\{i'i''\}$, where in each component μ' or μ'' there are at least $\kappa+1$ agents of type x (as for any $\iota\in(\nu_{i'}(\mu)\cup\nu_{i''}(\mu))$ it holds that $x_\iota=x$).

Agent i can at most be connected to one of i',i'' in $\mu\backslash(\{i'i''\})$ as otherwise i',i'' would be connected in $\mu\backslash\{i'i''\}$. Denote the in subnetwork of $\{\mu',\mu''\}$ where i is part of as $\tilde{\mu}$ and define $\tilde{N}=\{\iota\in N_x\backslash\nu_i(\mu):\exists\iota'\in N:\iota\iota'\in\tilde{\mu}\}$.

Let $\iota_0\in\arg\max_{\iota\in i',i''}p_{\iota i}$ and iteratively $\iota_l\in\nu_{\iota_{l-1}}(\mu),l\in\mathbb{N}$. Moreover, there must be a unique path in $\mu\backslash\{i'i''\}$ between any two agents $\iota,\iota'\in\tilde{N}$ as otherwise $i\iota,i\iota'\notin\mu$ but $p_{\iota\iota'}(\tilde{\mu}\backslash\{i'i''\})<\infty$; by changing the labels we could denote $i'=\iota$ and $i''=\iota'$ and we would have shown the existence of the desired pair of agents.

The fact here is a unique path between any two agents in $\tilde{N}$ entails that at level l or below there are $\sum_{q=0}^{l}(\kappa-1)^q$ agents; thus $n_x\geq\sum_{q=0}^{l}(\kappa-1)^q$. Let l be the minimal q such that $\forall\in\iota\in\tilde{N}:p_{i\iota}\leq q$; as n_x is finite such a q must exist. In addition, as there is a unique path between agents in μ then any agent $\iota\in\tilde{N}:p_{\iota\iota_0}=l$ has only one link, and thus its degree is less than κ (as $\kappa\geq 2$). This violates the condition that all $i'\in N$ where $x_{i'}=x$ has $k_{i'}=\kappa$. □

Proof Theorem 1.i. We begin with showing property (i). Suppose $\mu\in M^{p-srt:conn}\cap M^{no-surpl.}$ - we will demonstrate there are thresholds on δ such that μ has pairwise stability. We're only interested in the minimal thresholds such that for all values of externalities below those then stability holds. Thus it is sufficient to evaluate the deviations from the network where the net gains are highest.

The losses from breaking a link $ij\in\mu$ can be shown to have bounded from below such that: $\geq\delta\cdot(1-Z(x,x))$. Suppose that $n_x=\kappa+1,x\in X$ then $\{ij\in\mu:x_i=x,x_{i'}=x\}$ is a clique (i.e. any i,i' of type x are linked). This entails that $p_{ii'}(\mu\backslash\{ii'\})=2$ and thus $p_{ii'}(\mu\backslash\{ii'\})<\infty$. Suppose instead that $n_x>\kappa+1,x\in X$ then by Lemma 3 there exists some i,i', both of type

x such that $p_{ii'}(\mu\backslash\{ii'\}) < \infty$. Thus when evaluating losses at the threshold we can assume that when deleting some link ij that i, j are connected in $\mu\backslash\{ij\}$. Although the length of the shortest paths may increase, there will still be an indirect connection and therefore no loss of utility for anyone but the two agents who lose their link. Therefore we assume throughout that when evaluating thresholds if ii' is deleted in $M^{p-srt:conn} \cap M^{no-surpl.}$ then only agents i, i', who must be of same type, will each lose $(1-\delta)\cdot z(x,x)$ while no other agents of incur a loss.

Suppose two agents i, j of distinct types respectively $x, \tilde{x}$ deviate by forming a link and delete a link each from μ. The total loss for i and j for deleting a link each is:

$$(1-\delta)\cdot[z(x,x)+z(\tilde{x},\tilde{x})] = (1-\delta)\cdot(\hat{Z}_{x,\tilde{x}}+1)\cdot Z(x,\tilde{x}).$$

The benefits gained for agent i for establishing a link to j is $[1+(n_{\tilde{x}}-1)\cdot(1-\delta)]\cdot z(x,\tilde{x})$. Thus the total benefits gained for i and j from pairwise deviation can be bounded as follows.

$$\begin{aligned} & [1+(n_x-1)\cdot\delta]\cdot z(x,\tilde{x}) + [1+(n_{\tilde{x}}-1)\cdot\delta]\cdot z(\tilde{x},x), \\ = \quad & \langle 1+[\max(n_x,n_{\tilde{x}}) - |n_x - n_{\tilde{x}}|\cdot \hat{z}_{x,\tilde{x}} - 1]\cdot\delta\rangle \cdot Z(x,\tilde{x}). \end{aligned}$$

where $\hat{z}_{x,\tilde{x}} = \frac{z(\arg\min_{x,\tilde{x}} n_x,\, \arg\max_{x,\tilde{x}} n_x)}{Z(x,\tilde{x})}$.

We can derive the threshold for pairwise stability:

$$\begin{aligned} (1-\delta)\cdot(\hat{Z}_{x,\tilde{x}}+1)\cdot Z(x,\tilde{x}) &= \langle 1+[\max(n_x,n_{\tilde{x}}) - |n_x-n_{\tilde{x}}|\cdot\hat{z}_{x,\tilde{x}} - 1]\cdot\delta\rangle\cdot Z(x,\tilde{x}), \\ (1-\delta)\cdot(\hat{Z}_{x,\tilde{x}}+1) &= \langle 1+[\max(n_x,n_{\tilde{x}}) - |n_x-n_{\tilde{x}}|\cdot\hat{z}_{x,\tilde{x}} - 1]\cdot\delta\rangle, \\ \hat{Z}_{x,\tilde{x}} &= \left[\max(n_x,n_{\tilde{x}}) - |n_x-n_{\tilde{x}}|\cdot\hat{z}_{x,\tilde{x}} + \hat{Z}_{x,\tilde{x}}\right]\cdot\delta, \\ \delta &= \frac{\hat{Z}_{x,\tilde{x}}}{\max(n_x,n_{\tilde{x}}) - |n_x-n_{\tilde{x}}|\cdot\hat{z}_{x,\tilde{x}} + \hat{Z}_{x,\tilde{x}}}. \end{aligned} \tag{6}$$

Thus we can establish a lower bound for $\bar{\delta}$ (i.e. the upper bound in δ for pairwise stability of μ) by taking the minimum of right-hand-side in Equation 6; thus it follows that: $\bar{\delta} = \min_{x,\tilde{x}\in X}\left(\frac{\hat{Z}_{x,\tilde{x}}}{\max(n_x,n_{\tilde{x}})-|n_x-n_{\tilde{x}}|\cdot\hat{z}_{x,\tilde{x}}+\hat{Z}_{x,\tilde{x}}}\right)$. Thus we have established (i). ■

Proof of Theorem 1.ii. We move on to establishing property (ii). Suppose $\mu \in M^{p-srt:conn} \cap M^{no-surpl.}$ for which μ is inefficient. We're interested in showing there exist a threshold such that for all values of externalities above this then inefficiency of sorting holds. Note as we only compare against alternative networks that involve two types there may exist other networks which provide higher aggregate utility than the sorted networks for values below this threshold.

We follow a similar procedure for establishing property (ii) by finding $\underline{\delta}$ for when the aggregate gains of establishing links across types is zero. We will evaluate the bridged network $\tilde{\mu} = \mu \cup \{ij, i'j'\}\backslash\{ii', jj'\}$ where $x_{i'} = x_i$ and $x_{j'} = x_j$. The total loss $U(\mu) - U(\mu\backslash\{ii', jj'\})$ is equal to double that of i, j suffers, i.e. $2(1-\delta)\cdot\hat{Z}_{x,\tilde{x}}\cdot Z(x,\tilde{x})$.

The gains from connecting are two links of value $Z(x, \tilde{x})$ as well as $n_x \cdot n_x - 2$ indirect connections (between all agents of type x and $\tilde{x}$) of value $Z(x, \tilde{x}) \cdot \delta$. It follows that the total gains in benefits are $Z(x, \tilde{x}) \cdot [2 + (n_x \cdot n_{\tilde{x}} - 2) \cdot \delta]$. The threshold $\underline{\delta}$ can be found from finding when gains equal losses:

$$\begin{aligned} 2(1-\delta) \cdot (\hat{Z}_{x,\tilde{x}} + 1) \cdot Z(x, \tilde{x}) &= 2Z(x, \tilde{x}) + (n_x \cdot n_{\tilde{x}} - 2) \cdot Z(x, \tilde{x}) \cdot \delta, \\ (1-\delta) \cdot (\hat{Z}_{x,\tilde{x}} + 1) &= 1 + (\tfrac{1}{2} n_x \cdot n_{\tilde{x}} - 1) \cdot \delta, \\ \frac{\hat{Z}_{x,\tilde{x}}}{\hat{Z}_{x,\tilde{x}} + \frac{1}{2} n_x \cdot n_{\tilde{x}}} &= \delta. \end{aligned} \tag{7}$$

It follows that the threshold $\underline{\delta}$ (which is such that every δ above implies there exists an efficient deviation from μ) must be at least the minimum of left-hand-side in Equation 7 over possible types, i.e. it must hold that $\underline{\delta} = \min_{x,\tilde{x} \in X} \left(\frac{\hat{Z}_{x,\tilde{x}}}{\hat{Z}_{x,\tilde{x}} + \frac{1}{2} n_x \cdot n_x} \right)$. This terminates the proof for property (ii) of the theorem.

Proof of Theorem 1.iii. Property (iii) follows as Lemma 2 can be applied to the subset of agents associated with each type as $\forall x \in X : n_x > \kappa$ and $\kappa \cdot n_x \in 2\mathbb{N}$. ■

Proof of Theorem 2.i. We move on to property (iv.a) of the proof where we establish characterization of pairwise stable networks when there is type self-sufficiency and moderate externalities. We need to show that $\hat{M} = M^{p-stb}_{\delta \leq \bar{\delta}}$ if $|X| = 2$. From property (ii) $\hat{M} \subseteq M^{p-stb}_{\delta \leq \bar{\delta}}$ thus it remains to show: that $M^{p-stb}_{\delta \leq \bar{\delta}} \subseteq \hat{M}$. As $|X| = 2$ it holds that $X = \{\underline{x}, \overline{x}\}$. Define Υ:

$$\Upsilon = (1-\delta) \cdot [z(\bar{x}, \bar{x}) + z(\underline{x}, \underline{x})] - [1 + (n_{\underline{x}} - 1) \cdot \delta] \cdot z(\underline{x}, \bar{x}) - [1 + (n_{\bar{x}} - 1) \cdot \delta] \cdot z(\bar{x}, \underline{x}).$$

As $\delta < \underline{\delta}$ it follows from Equation 6 that:

$$\Upsilon > 0. \tag{8}$$

Suppose $\mu \notin M^{p-sort}$. Define a sequence of agent pairs, $i_0 j_0, i_1 j_1, \ldots$ as follows. Let agents $i_0, j_0 \in N$ be such that $x_i \neq x_j$ and $ij \in \mu$; such i_0, j_0 must exist as $\mu \notin M^{p-sort}$. Without loss of generality let $x_{i_0} = x$ and $x_{j_0} = \tilde{x}$ where $x, \tilde{x} \in X$, and assume that:

$$-\tau_{i_0 j_0} > (1-\delta) \cdot z(\tilde{x}, \tilde{x}) - [1 + (n_{\tilde{x}} - 1) \cdot \delta] \cdot z(\tilde{x}, x). \tag{9}$$

The above inequality must hold for either type x or $\tilde{x}$ as we substitute labels for i, j as well as $x, \tilde{x}$ due to $\Upsilon > 0$.

Let $l \in \mathbb{N}$. It is assumed that for any $q < l : x_{i_q} = x, x_{j_q} = \tilde{x}$. Also assume an associated set collection of links, $\mu_{l-1} \subset \mu$, and let the set be defined as $\mu_{l-1} = \cup_{q=0}^{l-1} \{i_q j_q\}$ such that for each $q < l$: $i_q i_{q-1} \notin \mu$ if q is odd and $j_q j_{q-1} \notin \mu$ if q is even. At step $q \in \mathbb{N}$ let $\iota_q = i_{q-1}$ if q is even else denote $\iota_q = j_{l-1}$. Also let $\eta_q \in \{i_{q-1}, j_{q-1}\} : \eta_q \neq \iota_q$. This entails that $\iota_1 = i_0$ and $\eta_1 = j_0$.

Define $x_q = x, \tilde{x}_q = \tilde{x}$ if q is even else vice versa. Also let $N_q = \{\iota \in N : x_\iota = x_q\}$.

Suppose that at every $q \in \mathbb{N} : q < l$ it holds that $\iota'_q \notin \nu_{\iota_q}(\mu)$, $x_{\iota'_l} = x_l$ and let $\eta'_l \in \nu_{\iota'_l}(\mu)$. Finally also define at every $q < l$ the move $\Delta\mu_q = \mu \cup \{\iota_q \iota'_q\} \backslash \{\iota_q \eta_q, \iota'_q \eta'_q\}$ and let:

$$\begin{aligned} \Delta U_q &= u_{\iota_q}(\Delta\mu_q) - u_{\iota_q}(\mu) + u_{\iota'_q}(\Delta\mu_q) - u_{\iota'_q}(\mu) && (10) \\ \Delta \hat{U}_q &= u_{i_q}(\Delta\mu_{q+\mathbf{1}_{q:even}}) - u_{i_q}(\mu) + u_{j_q}(\Delta\mu_{q+\mathbf{1}_{q:odd}}) - u_{j_q}(\mu) && (11) \end{aligned}$$

Note that $\Delta\mu_q = \{i_q i_{q-1}\} \cup \Delta\tilde{\mu}_q$ if q is even and $\Delta\mu_q = \{j_q j_{q-1}\} \cup \Delta\tilde{\mu}_q$ if q is even; $\Delta\tilde{\mu}_q = \mu \cup \backslash \{i_{q-1} j_{q-1}, i_q j_q\}$. By inserting i, j for ι, η we yield the following expression:

$$\sum_{q=l'}^{l-1} \Delta U_q = \sum_{q=l'}^{l-2} \Delta \hat{U}_q + u_{\iota'_{l-1}}(\Delta\mu_{l-1}) - u_{\iota'_{l-1}}(\mu) + u_{\iota_{l'}}(\Delta\mu_{l'}) - u_{\iota_{l'}}(\mu) \qquad (12)$$

Assume that for every $q \in \mathbb{N}$ where $q < l$:

$$\begin{aligned} |\{\iota \in N : x_\iota = x_q \wedge p_{\iota\iota_q}(\mu) < \infty \wedge p_{\iota\iota_q}(\mu \cup \{\iota_q \iota'_q\} \backslash \{\eta_q \iota_q\}) = \infty\}| &= 0 && (13) \\ |\{\iota \in N : x_\iota = x_q \wedge p_{\iota\iota'_q}(\mu) < \infty \wedge p_{\iota\iota'_q}(\Delta\mu_q) = \infty\}| &= 0 && (14) \end{aligned}$$

Suppose Equation 13 is satisfied. It follows that net gains of benefits for ι_q from deleting the link with η_q while forming a link together with ι'_q can be bounded: the upper bound on losses is when a connection is lost to all agents of type $\tilde{x}_q$: $[1 + (n_q - 1) \cdot \delta] \cdot z(x_q, x_q)$; the lower bound on gains is $(1 - \delta) \cdot z(x_q, \tilde{x}_q)$ as the distance between $\iota_q \iota'_q$ is shortened to 1.

$$u_{\iota_q}(\mu \cup \{\iota_q \iota'_q\} \backslash \{\iota_q \eta_q\}) - u_{\iota_q}(\mu) \quad \geq \quad (1 - \delta) \cdot z(x_q, x_q) - [1 + (n_q - 1) \cdot \delta] \cdot z(x_q, \tilde{x}_q) \qquad (15)$$

Suppose Equations 13 and 14 hold - we can demonstrate that Equation 14 also where we replace ι'_q with ι_q. If $p_{\iota_q \iota'_q}(\mu) < \infty$ then as it also holds that $p_{\iota_q \iota'_q}(\Delta\mu_q) < \infty$ it follows that $p_{\iota'_q \iota''_q}(\mu) < \infty$ and $p_{\iota'_q \iota''_q}(\Delta\mu_q) = \infty$ which violates Equation 14. Thus it must be that $p_{\iota_q \iota'_q}(\mu) = \infty$. Suppose instead $p_{\iota_q \iota'_q}(\mu) = \infty$. If $\exists \iota''_q \in N : p_{\iota_q \iota''_q}(\mu \cup \{\iota_q \iota'_q\} \backslash \{\iota_q \eta_q\}) < \infty \wedge p_{\iota_q \iota''_q}(\Delta\mu_q) = \infty$ then it must be that $p_{\eta'_q \iota''_q}(\mu) < \infty$ and thus $p_{\iota'_q \iota''_q}(\mu) < \infty$ which implies that $p_{\iota_q \iota''_q}(\mu) = \infty$. However, this is a violation of $p_{\iota_q \iota''_q}(\mu) < \infty$.

$$\{\iota \in N : x_\iota = x_q \wedge p_{\iota\iota_q}(\mu) < \infty \wedge p_{\iota\iota_q}(\Delta\mu_q) = \infty\}| = 0 \qquad (16)$$

Analogue to the derivation of Inequality 15 the net gains are bounded when Equations 13 and 14 are satisfied:

$$\min_{\iota \in \{\iota_q, \iota'_q\}} [u_\iota(\Delta\mu_q) - u_\iota(\mu)] \geq (1 - \delta) \cdot z(x_q, x_q) - [1 + (n_q - 1) \cdot \delta] \cdot z(x_q, \tilde{x}_q), \qquad (17)$$

One implication of Inequalities 9 and 17 if $l > 1$:

$$\begin{aligned} u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) - \tau_{\iota_1\eta_1} &\geq (1-\delta)\cdot z(x,x) - [1 + (n_x - 1)\cdot\delta]\cdot z(x,\tilde{x}) - \tau_{i_0 j_0} \\ u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) - \tau_{\iota_1\eta_1} &\geq \Upsilon \end{aligned} \tag{18}$$

Another implication of Inequality 17 is that:

$$u_{i_q}(\Delta\mu_{q+\mathbf{1}_{odd}(q)}) - u_{i_q}(\mu) + u_{j_q}(\Delta\mu_{q+\mathbf{1}_{even}(q)}) - u_{j_q}(\mu) \geq \Upsilon, \qquad \forall q \in [[1, l-1]]$$

In order for $\Delta\mu_q$ not to be a profitable pairwise deviation it must hold that:

$$\begin{aligned} u_{\iota_q}(\mu) + u_{\iota'_q}(\mu) + \tau_{\iota'_q\eta'_q} + \tau_{\iota_q\eta_q} &\geq u_{\iota_q}(\Delta\mu_q) + u_{\iota'_q}(\Delta\mu_q) \\ \tau_{\iota'_q\eta'_q} &\geq \Delta U_q + \tau_{\eta_q\iota_q} \end{aligned}$$

We can rewrite the above inequality using that $\iota'_{q-1} = \eta_q, \eta'_{q-1} = \iota_q$ and thus $\tau_{\iota'_{q-1}\eta'_{q-1}} = \tau_{\eta_q\iota_q}$. We also substitute in Equation 10 and assume the above inequality holds for any $q < l$:

$$\begin{aligned} \tau_{\iota'_{l-1}\eta'_{l-1}} &\geq \Delta U_{l-1} + \tau_{\iota'_{l-2}\eta'_{l-2}} \\ \tau_{\iota'_{l-1}\eta'_{l-1}} &\geq \sum_{q=l'}^{l-1} \Delta U_q + \tau_{\iota'_{l'-1}\eta'_{l'-1}} \end{aligned} \tag{19}$$

As $\tau_{\eta_l\iota_l} = \tau_{\iota'_{l-1}\eta'_{l-1}}$ and $-\tau_{\iota_l\eta_l} = \tau_{\eta_l\iota_l}$ it follows that using Equation 12:

$$\begin{aligned} -\tau_{\iota_l\eta_l} &\geq \sum_{q=l'}^{l-1} \Delta U_q + \tau_{\iota'_{l'-1}\eta'_{l'-1}} \\ &= \sum_{q=1}^{l-2} \Delta\hat{U}_q + u_{\iota'_{l-1}}(\Delta\mu_{l-1}) - u_{\iota'_{l-1}}(\mu) + u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) + \tau_{\iota'_0\eta'_0} \\ &= \sum_{q=1}^{l-2} \Delta\hat{U}_q + u_{\eta_l}(\Delta\mu_{l-1}) - u_{\eta_l}(\mu) + u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) - \tau_{\iota_1\eta_1} \end{aligned} \tag{20}$$

Define the set of partners for ι_l:

$$\begin{aligned} \hat{N}^*_l(\iota_l, \mu_{l-1}) &= \{\iota \in N\backslash\{\iota_l\} : x_\iota = x_{\iota_l},\, p_{\iota\iota_l}(\mu) < \infty,\, p_{\iota\iota_l}(\mu\backslash\{\iota_l\eta_l\}) = \infty\} \\ \hat{N}^{**}_l(\iota_l, \mu_{l-1}) &= \{\iota \in N\backslash\{\iota_l\} : x_\iota = x_{\iota_l},\, \iota\iota_l \notin \mu\} \\ \hat{N}_l(\iota_l, \mu_{l-1}) &= \begin{cases} \hat{N}^*_l(\iota_l, \mu_{l-1}) \text{ if } \hat{N}^*_l(\iota_l, \mu_{l-1}) \neq \emptyset, \\ N^{**}_l(\iota_l, \mu_{l-1}) \text{ else.} \end{cases} \end{aligned} \tag{21}$$

A property of $\hat{N}_l$ is that $\hat{N}_l \neq \emptyset$; this follows as $\min_{\hat{x}\in X} n_{\hat{x}} \geq \kappa + 1$. Let $\iota'_l \in \hat{N}_l$ which implies that Equation 13 holds.

Suppose that $k_{\iota_l'}(\mu) < \kappa$. As Equation 13 holds it follows that

$$u_{\iota_q'}(\mu \cup \{\iota_q \iota_q'\} \backslash \{\iota_q \eta_q\}) - u_{\iota_q'}(\mu) \geq (1-\delta) \cdot z(x_q, x_q),$$

and thus $u_{\iota_q'}(\mu \cup \{\iota_q \iota_q'\} \backslash \{\iota_q \eta_q\}) > 0$.

We can also derive utility bounds using Inequality 20 along with Inequalities 8, 18:

$$\begin{aligned}
& u_{\iota_l}(\mu \cup \{\iota_l \iota_l'\} \backslash \{\iota_l \eta_l\}) - u_{\iota_l}(\mu) - \tau_{\iota_l \eta_l}, \\
\geq\ & u_{\iota_l}(\mu \cup \{\iota_l \iota_l'\} \backslash \{\iota_l \eta_l\}) - u_{\iota_l}(\mu) + u_{\eta_l}(\Delta\mu_{l-1}) - u_{\eta_l}(\mu) + \sum_{q=1}^{l-2} \Delta\hat{U}_q + u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) - \tau_{\iota_1 \eta_1}, \\
\geq\ & l \cdot \Upsilon, \\
>\ & 0,
\end{aligned}$$

thus ι_l, ι_l' can profitably from deviate pairwise. Thus it must be that $k_{\iota_l'}(\mu) = \kappa$.

Suppose there exists $\iota_l' \in N_l \backslash \nu_i(\mu), \iota_l'' \in N_l \backslash \{\iota_l, \iota_l'\}$ such that $\iota_l' \iota_l'' \in \mu$, $p_{\iota_l' \iota_l''}(\mu \backslash \{\iota_l' \iota_l''\}) < \infty$ and $\tau_{\iota_l' \iota_l''} \leq 0$. This entails that $u_{\iota_l'}(\Delta\hat{\mu}_l) - u_{\iota_l'}(\mu) \geq 0$ where $\Delta\mu_l = \mu \cup \{\iota_l \iota_l'\} \backslash \{\iota_l \eta_l, \iota_l' \iota_l''\}$. This follows from $u_{\iota_l'}(\Delta\mu_l) - u_{\iota_l'}(\mu) = u_{\iota_l'}(\Delta\mu_l) - u_{\iota_l'}(\mu \cap \Delta\mu_l) - [u_{\iota_l'}(\mu \cap \Delta\mu_l) - u_{\iota_l'}(\mu)]$ and $u_{\iota_l'}(\Delta\mu_l) - u_{\iota_l'}(\mu \cap \Delta\mu_l) \geq 1 - z(x,x)$ and $u_{\iota_l'}(\mu \cap \Delta\mu_l) - u_{\iota_l'}(\mu) = 1 - z(x,x)$. As $\tau_{\iota_l' \iota_l''} \leq 0$ it follows that that utility for ι_l' is:

$$u_{\iota_l'}(\Delta\hat{\mu}_l) - u_{\iota_l'}(\mu) - \tau_{\iota_l' \iota_l''} \geq 0.$$

And utility for ι_l can bounded be as follows using Inequality 15 for $u_{\iota_l}(\Delta\hat{\mu}_l) - u_{\iota_l}(\mu)$ as Equation 13 holds :

$$\begin{aligned}
& u_{\iota_l}(\Delta\hat{\mu}_l) - u_{\iota_l}(\mu) - \tau_{\iota_l \eta_l} \\
=\ & u_{\iota_l}(\Delta\hat{\mu}_l) - u_{\iota_l}(\mu) + \tau_{\eta_l \iota_l} \\
\geq\ & \sum_{q=1}^{l-1} \Delta U_q + u_{\iota_l}(\Delta\hat{\mu}_l) - u_{\iota_l}(\mu) + \tau_{j_0 i_0} \\
=\ & \sum_{q=1}^{l-2} \Delta\hat{U}_q + u_{\iota_l}(\Delta\hat{\mu}_l) - u_{\iota_l}(\mu) + u_{\eta_l}(\Delta\mu_{l-1}) - u_{\eta_l}(\mu) + u_{i_0}(\Delta\mu_1) - u_{i_0}(\mu) - \tau_{i_0 j_0} \\
\geq\ & l \cdot \Upsilon \\
>\ & 0
\end{aligned} \tag{22}$$

The above inequalities entails that ι_l, ι_l' can deviate profitably pairwise; this is a violation of pairwise stability and thus cannot be true. Thus there exists no $\iota_l' \iota_l'' \in \mu$ such that $\iota_l' \in N_l \backslash \nu_i(\mu), \iota_l'' \in N_l \backslash \{\iota_l, \iota_l'\}$ as well as $p_{\iota_l' \iota_l''}(\mu \backslash \{\iota_l' \iota_l''\}) < \infty$ and $\tau_{\iota_l' \iota_l''} \leq 0$.

Suppose that $\forall \iota_l' \in N_l : \nexists \eta_l' \in \nu_{\iota_l'}(\mu \backslash \mu_{l-1}) : x_{\eta_l'} \neq x_l$. This entails that $\forall \iota_l' \in N_l : \nexists \eta_l' \in \nu_{\iota_l'}(\mu) : x_{\eta_l'} \neq x_l$ as $k_{\iota_l'}(\mu \backslash \mu_{l-1}) = k_{\iota_l'}(\mu)$. By Lemma 3 it follows there exists $\iota_l', \iota_l'' \in N_l \backslash \nu_i(\mu)$ such that $p_{\iota_l'\iota_l''}(\mu\backslash\{\iota_l'\iota_l''\}) < \infty$, $\iota_l'\iota_l'' \in \mu$ and $\tau_{\iota_l'\iota_l''} \leq 0$ which by the arguments above cannot be true. Therefore there has to exist some $\iota_l' \in N_l$ for which there is an agent $\eta_l' \in \nu_{\iota_l'}(\mu\backslash\mu_{l-1})$ where it holds that $x_{\eta_l'} \neq x_l$.

A duplicate occurs if $i_{l-1}, j_{l-1} \in \mu_{l-2}$. That is for some $l' < l$ it holds that either $\iota_l, \eta_l = \iota_{l'}, \eta_{l'}$ if $l - l'$ is even or $\iota_l, \eta_l = \eta_{l'}, \iota_{l'}$ if $l - l'$ is odd.

If $l - l'$ is odd, then $\tau_{\iota_{l'-1}'\eta_{l'-1}'} = -\tau_{\iota_{l-1}'\eta_{l-1}'}$ and therefore we can reduce the Inequality 19:

$$
\begin{aligned}
0 &\geq \sum_{q=l'}^{l-1}[u_{\iota_q}(\Delta\mu_q) - u_{\iota_q}(\mu) + u_{\iota_q'}(\Delta\mu_q) - u_{\iota_q'}(\mu)] + 2\tau_{\iota_{l'-1}'\eta_{l'-1}'} \\
&= \sum_{q=l'}^{l-2} \Delta\hat{U}_q + u_{\iota_{l-1}'}(\Delta\mu_{l-1}) - u_{\iota_{l-1}'}(\mu) + u_{\iota_{l'}}(\Delta\mu_{l'}) - u_{\iota_{l'}}(\mu) + 2\tau_{\iota_{l'-1}'\eta_{l'-1}'} \\
&= \sum_{q=l'}^{l-2} \Delta\hat{U}_q + 2 \cdot \left\langle u_{\eta_{l'-1}'}(\Delta\mu_{l'}) - u_{\eta_{l'-1}'}(\mu) + \tau_{\iota_{l'-1}'\eta_{l'-1}'} \right\rangle \\
&= \sum_{q=l'}^{l-2} \Delta\hat{U}_q + 2 \cdot \left\langle u_{\eta_{l'-1}'}(\Delta\mu_{l'}) - u_{\eta_{l'-1}'}(\mu) + \sum_{q=1}^{l'-1} \Delta U_q + \tau_{\iota_0'\eta_0'} \right\rangle \\
&= \sum_{q=l'}^{l-2} \Delta\hat{U}_q + 2 \cdot \sum_{q=1}^{l'-1} \Delta\hat{U}_q + 2 \cdot [u_{\iota_1}(\Delta\mu_1) - u_{\iota_1}(\mu) - \tau_{\iota_1\eta_1}] \\
&\geq (l + l') \cdot \Upsilon \\
&> 0,
\end{aligned}
$$

thus there must be a feasible pairwise deviation for ι_q, ι_q' where $q \in [[1, l-1]]$.

If $l - l'$ is even then $\tau_{\iota_{l-1}'\eta_{l-1}'} = \tau_{\iota_{l'-1}'\eta_{l'-1}}$; thus Inequality 19 for no pairwise deviation becomes: $0 \geq \sum_{q=l'}^{l-1} \Delta U_q$. This can in turn be rewritten as follows:

$$
0 \geq \sum_{q=l'}^{l-2} \Delta\hat{U}_q + u_{\iota_{l-1}'}(\Delta\mu_{l-1}) - u_{\iota_{l-1}'}(\mu) + u_{\iota_{l'}}(\Delta\mu_{l'}) - u_{\iota_{l'}}(\mu)
$$

Using that $\iota_q = \eta_{q-1}'$ and $\eta_{l'-1}' = \eta_{l-1}'$ we get: $0 \geq \sum_{q=l'}^{l-1} \Delta\hat{U}_q$. Recall that for all $q \in \mathbb{N} : q < l$ it holds that $\Delta\hat{U}_q \geq \Upsilon$ where $\Upsilon > 0$. Thus there must be a feasible pairwise deviation.

Due to $\iota_l', \eta_l' = \eta_{l+1}, \iota_{l+1}$ it follows that it cannot be that $\iota_l'\eta_l' \in \mu_{l-1}$ as otherwise $i_l j_l \in \mu_{l-1}$. This entails $\nexists \iota_l' \in N_l : \exists \eta_l' \in \nu_{\iota_l'}(\mu_{l-1})$. Therefore we can assume $\forall \iota_l' \in N_l : \nexists \eta_l' \in \nu_{\iota_l'}(\mu_{l-1})$ and thus $\forall \iota_l' \in N_l : k_{\iota_l'}(\mu\backslash\mu_{l-1}) = k_{\iota_l'}(\mu)$.

Suppose that Equation 14 is violated for any $\iota_l' \in \hat{N}_l$. This is equivalent to it holds for any $\iota_l' \in \hat{N}_l$ where $\eta_l' \in \nu_{\iota'}(\mu)$ that there is some other $\iota_l'' \in \hat{N}_l$ such that $p_{\iota_l'\iota_l''}(\Delta\mu_l) = \infty$. Let $\iota_l^{(1)} = \iota_l'$. As Equation 14 must hold for any $\iota_l' \in \hat{N}_l$ we can reproduce the argument iteratively and thus for $\iota_l^{(q)} \in \hat{N}_l, q \in \mathbb{N}$ there is some $\eta_l^{(q)} \in \nu_{\iota_l^{(q)}}(\mu)$ such that for some $\iota_l^{(q+1)} \in \hat{N}_l\backslash\{\iota_l^{(1)}, .., \iota_l^{(q)}\}$ it holds that $p_{\iota_l^{(1)}\iota_l^{(q+1)}}(\Delta\mu_l) = \infty$. However, as $n < \infty$ it follows that there for some $q \in \mathbb{N}$ that $N_l \backslash \{\iota_l^{(1)}, .., \iota_l^{(q)}\} = \emptyset$. Thus let instead $\iota_l' = \iota_l^{(q)}$; for any $\eta_l' \in \nu_{\iota_l'}(\mu)$ there is no $\iota_l'' \in N_l$ such that $p_{\iota_l'\iota_l''}(\mu) = \infty$. This contradicts that Equation 14 is violated for agent $\iota_l' = \iota_l^{(q)}$.

Suppose $\mu \notin M^{no-surpl.}$. This would entail that $\exists i \in N : k_i(\mu) < \kappa$. As $n_x > \kappa$ it must be that $\exists i' \in N : x_{i'} = x_i, ii' \notin \mu$. Suppose that $k_{i'} < \kappa$ then $\sum_{\iota \in \{i,i'\}}[u_\iota(\mu \cup \{ii'\}) - u_\iota(\mu)] > 0$ and thus ii' can be formed profitably pairwise. Moreover, as $k_{i'}(\mu) = \kappa$ it follows that $\exists i'' \in \nu_{i'}$: $ii'' \notin \mu$, $x_{i''} = x_i$. By Lemma 3 it follows there exists $\iota, \iota' \in \tilde{N} \backslash \nu_i(\mu)$ such that $p_{\iota\iota'}(\mu \backslash \{\iota\iota'\}) < \infty$, $\iota\iota' \in \mu$ and $\tau_{\iota\iota'} \leq 0$. This entails that $u_\iota(\mu) - u_\iota(\mu \backslash \{\iota\iota'\}) + \tau_{\iota\iota'} \leq (1-\delta)z(x,x)$. Moreover, as $\sum_{j \in \{i,\iota\}}[u_j(\mu \cup \{i\iota\} \backslash \{\iota\iota'\}) - u_j(\mu \backslash \{\iota\iota'\})] \geq (1-\delta) \cdot Z(x,x)$ it holds that:

$$\sum_{j \in \{i,\iota\}} [u_j(\mu \cup \{i\iota\} \backslash \{\iota\iota'\}) - u_j(\mu)] - \tau_{\iota\iota'} \geq (1-\delta) \cdot z(x,x)$$

.

Thus i, ι can deviate profitably pairwise which contradicts pairwise Nash stability. Therefore it must be that $\mu \in M^{no-surpl.}$

Suppose $\mu \notin M^{p-sort+conn}$. As $\mu \in M^{p-sort} \cap M^{no-surpl.}$ there exist $i, i', j, j' \in N : x_i = x_{i'} = x_j = x_{j'}$ and $ij, i'j' \in \mu$ and $p_{ii'}(\mu) = \infty$. Without loss of generality we assume that $\tau_{ij}, \tau_{i'j'} \leq 0$ (otherwise we could simply switch identities some i's and j's). This entails:

$$\min_{\iota \in \{i,i'\}} [u_\iota(\mu \backslash \{ij, i'j'\}) - u_\iota(\mu)] + \tau_{ij} + \tau_{i'j'} \leq 2(1-\delta) \cdot z(x,x)$$

Also we have that:

$$\min_{\iota \in \{i,i'\}} [u_\iota(\mu \cup \{ii'\} \backslash \{ij, i'j'\}) - u_\iota(\mu \backslash \{ij, i'j'\})] \geq (\kappa+1) \cdot (1-\delta) \cdot z(x,x)$$

This entails that $\sum_{\iota \in \{i,i'\}}[u_\iota(\mu \cup \{ii'\} \backslash \{ij, i'j'\}) - u_\iota(\mu)] - \tau_{ij} - \tau_{i'j'} \geq \kappa \cdot (1-\delta) \cdot Z(x,x)$; thus i, i' can deviate profitably. Thus we have shown that $M^{p-stb}_{\delta \leq \bar{\delta}} \subseteq \hat{M}$ which terminates the proof of property (iv). ■

Proof of Theorem 2.ii and 2.iii In order to prove property (iv.b) and (iv.c) we begin by noting that utility under hyperbolic decay (from Equation 3) can be expressed as:

$$w_{ij}(\mu) = (1-\delta)\mathbf{1}_{=1}(p_{ij}(\mu)) + \delta \cdot \mathbf{1}_{\in[1,\infty)}(p_{ij}(\mu)). \tag{23}$$

Thus total utility from the network has the following form:

$$U(\mu) = \sum_{i \in N} \sum_{j \in N, j \neq i} \left[(1-\delta) \cdot \mathbf{1}_{=1}(p_{ij}(\mu)) + \delta \cdot \mathbf{1}_{\in[1,\infty)}(p_{ij}(\mu))\right] \cdot z(x_i, x_j). \tag{24}$$

The form for aggregate utility in Equation 24 has the advantage that it is easier to perform optimization on. From inspection we see that if a network is connected then indirect term in the weights, $\delta \cdot \mathbf{1}_{\in[1,\infty)}(p_{ij}(\mu))$, is one for all edges, and as a consequence the aggregate utility attains its maximal value.

We begin with restricting ourselves to look at perfect sorted networks. If it holds that each subnetwork $\mu_x \subseteq \mu$ consisting of all links within a given type is connected then the argument made above, that the aggregate utility from indirect links (i.e. stemming from $\delta \cdot \mathbf{1}_{\in[1,\infty)}(p_{ij}(\mu_x)) = 1$ for $x_i = x_j, i \neq j$ in Equation 24), is maximized (conditional on perfect sorting). Finally, it must be that each subnetwork has no link surplus. This follows as there exists a subnetwork $\tilde{\mu}_x$ with no link surplus which is connected from Lemma 2. Thus a violation of link surplus would imply inefficiency of μ_x as it would hold that the number of links between type x would be lower than the possible, i.e. $\sum_{ij\in\tilde{\mu}_x} \mathbf{1}_{=1}(p_{ij}(\tilde{\mu}_x)) > \sum_{ij\in\mu_x} \mathbf{1}_{=1}(p_{ij}(\mu_x))$, and thus provide lower welfare by Equation 24. As any network which consists of perfectly sorted subnetworks where each subnetwork is connected and has no link surplus gains exactly the same utility we know that $\hat{M}$ constitutes the set of efficient networks among networks with perfect sorting. We know from Theorem 1.iv that $\hat{M}$ set is non-empty and by Theorem 1.ii it follows that $\hat{M}$ are inefficient when $\delta < \underline{\delta}$.

We proceeed with restricting our search for efficient networks among those that do not have perfect sorting. In the case where there is not perfect sorting we can follow the same procedure by the same arguments that we used for the perfect sorting case. Again, if a network μ is connected then the total utility from indirect links is maximized as $\delta \cdot \mathbf{1}_{\in[1,\infty)}(p_{ij}(\mu)) = 1$ for every $i \neq j$. The utility accruing from (direct) links stems from the term $\mathbf{1}_{=1}(p_{ij}(\mu))$ in Equation 24. Due to supermodularity this utility from (direct) linking will be maximized if there is perfect sorting, however, this is not feasible as we require some links across type.

The minimal required links across types are two. This follows as at least one link across types is required and thus there the number of same type links must be at least one lower. With one link less among same types for each type it follows that two links can be established across types as each link less for type x means that $\sum_{x_i=x} k_i$ is two lower.

We also know that as long as μ is connected we maximize indirect benefits. To maximize utility from direct links we must maximize the number of same type links. We have argued that the highest attainable number of links within same type is $\frac{n_x\kappa}{2} - 1$ with two links across. It follows that a network which is connected and has $\frac{n_x\kappa}{2} - 1$ of same type links for each type with two links across must achieve the highest aggregate utility; the set of networks which fulfill this property is in fact $\bar{M}$.

It remains to show that $\bar{M}$ is non-empty; we can construct a cross-type bridged network $\bar{\mu}$ from another $\hat{\mu} \in \hat{M}$ such that $\bar{\mu} = \hat{\mu} \cup \{ij, i'j'\} \backslash \{ii', jj'\}$ where $x_i = x_{i'}, x_j = x_{j'}, x_i \neq x_j$ which by construction has the feature that $\bar{\mu}$ is connected (as the subnetworks for each type are connected if we choose each subnetwork using Lemma 2) and there are exactly $\frac{n_x\kappa}{2} - 1$ links of each type. Thus we have established that $\bar{M}$ constitutes the set of efficient networks among the networks that are not sorted.

From Theorem 1.ii we know there exist a threshold $\underline{\delta}$ such that if $\delta > \underline{\delta}$ then networks in $\hat{M}$ are inefficient. as such it must be that $\bar{M}$ are efficient for $\delta > \underline{\delta}$ as they are efficient among non-sorted networks. ■

Proof Proposition 3. Let $\mu \in \hat{M}$ and $\delta \in (\underline{\delta}, \bar{\delta})$. By construction there exists a network $\tilde{\mu}$ which has higher aggregate utility. Let the two pairs of agents ii', jj' be agents such that $\tilde{\mu} = \mu \cup \{ij, i'j'\} \backslash \{ii', jj'\}$ and $x_i = x_{i'} = x$ and $x_j = x_{j'} = x$. Specify a link-contingent contract to i, j where $\hat{\mu} = \mu \cup \{ij\} \backslash \{ii', jj'\}$ such that:

$$\forall \iota\iota' \in \{ij, i'j'\}: \qquad \mathcal{C}_{\iota\iota'} + \mathcal{C}_{\iota'\iota} \in \left(\tfrac{1}{2}[Z(x,x) + Z(\tilde{x},\tilde{x}) - 2Z(x,\tilde{x})],\ \tfrac{1}{2}[U(\tilde{\mu}) - U(\mu)]\right), \tag{25}$$

$$\forall \iota\iota' \notin \{ij, ji, i'j', j'i'\}: \qquad \mathcal{C}_{\iota\iota'} = 0. \tag{26}$$

By Theorem 1 we know that μ is pairwise stable. Pairwise stability implies that $\frac{1}{2}[Z(x,x) + Z(\tilde{x},\tilde{x}) - 2Z(x,\tilde{x})] > b_i(\mu) - b_i(\hat{\mu}) + b_j(\mu) - b_j(\hat{\mu})$ as deviation is not profitable. Using this fact together with Inequality 25 it follows that:

$$\mathcal{C}_{ij} + \mathcal{C}_{ji} > b_i(\mu) - b_i(\hat{\mu}) + b_j(\mu) - b_j(\hat{\mu}).$$

The above inequality entails agents i, j are a blocking coalition that can gain by deviating to $\hat{\mu}$; this blocking move is also the only profitable move for i, j due to pairwise stability of μ and Equation 26.

In network $\hat{\mu}$ agents i', j' have an incentive to form a link with one another as both have surplus link capacity (i.e. degree below the quota) and forming a link is profitable from Inequality 25. Moreover, we show in the following that this move is the one that ensures the highest aggregate net benefits to i', j'.

We begin with showing that linking across types to other agents of type $x, \tilde{x}$ is not profitable. Suppose i' links across types to another agent $j'' \in \{\iota \neq j' : x_\iota = x_{j'}\}$. First, note the pairwise deviation from μ to form $i'j''$ is unprofitable (due to pairwise stability), thus it less profitable than forming $i'j'$ from μ (which is profitable by Inequality 25). Second, the net-increase in value of the pairwise deviation to form $i'j'$ over $i'j''$ increases from μ to $\hat{\mu}$ - this is true as j' loses the link with i from μ while j'' has an unchanged number - thus j' will have a weakly lower opportunity cost of deleting links in $\hat{\mu}$. The same argument can be applied to j' for $i'' \in \{\iota \neq i' : x_\iota = x_{i'}\}$.

We turn to showing that linking to other agents of same type (staying sorted) is not more profitable as well. Suppose i' and j' link to same types as themselves respectively, i.e. $i'' \in \{\iota \neq i' : x_\iota = x_{i'}\}$ and $j'' \in \{\iota \neq j' : x_\iota = x_{j'}\}$. Suppose $ii'' \in \mu$ then no feasible pairwise moves to same type can exist in $\hat{\mu}$ as the move can only involve deleting links; same is true if $jj'' \in \mu$. Thus instead we use $ii'', jj'' \notin \mu$. It must be that any pairwise deviation forming either ii'' or jj'' from μ is unprofitable (as μ is pairwise stable); this implies that for any $\iota \in \nu_{i''}(\hat{\mu})$ and $\iota' \in \nu_{j''}(\hat{\mu})$ it holds that:

$$b_{i'}(\hat{\mu} \cup \{i'i''\} \backslash \{i''\iota\}) - b_{i'}(\hat{\mu}) + b_{i''}(\hat{\mu} \cup \{i'i''\} \backslash \{i''\iota\}) - b_{i''}(\hat{\mu}) - \tau_{i''\iota} \quad \leq \quad z(x,x), \quad (27)$$

$$b_{j'}(\hat{\mu} \cup \{j'j''\} \backslash \{j''\iota'\}) - b_{j'}(\hat{\mu}) + b_{j''}(\hat{\mu} \cup \{j'j''\} \backslash \{j''\iota'\}) - b_{j''}(\hat{\mu}) - \tau_{j''\iota'} \quad \leq \quad z(\tilde{x},\tilde{x}). \quad (28)$$

As $b_{i'}(\tilde{\mu}) - b_{i'}(\hat{\mu}) + b_{j'}(\tilde{\mu}) - b_{j'}(\hat{\mu}) = z(x,\tilde{x}) + z(\tilde{x},x)$ it follows that

$$b_{i'}(\tilde{\mu}) - b_{i'}(\hat{\mu}) + b_{j'}(\tilde{\mu}) - b_{j'}(\hat{\mu}) + \mathcal{C}_{i'j'} + \mathcal{C}_{j'i'} > z(x,x) + z(\tilde{x},\tilde{x}).$$

The above inequality implies together with Inequalities 27 and 28 that the total gains for i' and j' exceeds the total value that could be generated from alternative deviations. Thus there are two pairwise moves from μ to $\hat{\mu}$ and from $\hat{\mu}$ to $\tilde{\mu}$ which both provide strictly higher utility to the deviating agents.

Pairwise stability follows from three arguments. First, all deviations among agents where only links in $\tilde{\mu} \cap \mu$ are deleted will provide at most the same value in $\tilde{\mu}$ that the deviations did in μ - this follows as these agents all have the same links and in $\tilde{\mu}$ all agents are connected in $\tilde{\mu}$ and thus only direct links matter. This upper limit too gains from deviations implies none of these moves can be profitable as they were unprofitable form μ. Second, deviations that involve deletion of links in $\tilde{\mu} \backslash \mu$ are shown above to provide strictly higher value than any other deviations - thus deviating from $\tilde{\mu}$ must also provide strictly lower value. ■

Proof of Proposition 4. Under asymptotic independence it follows that average per agent utility for type x under asymptotic perfect sorting converges to (using a geometric series):

$$\frac{(\kappa-1)\,\delta}{1-(\kappa-1)\,\delta} z(x,x)$$

Let $\omega_{x\tilde{x}} = \kappa \cdot \mathbb{E}[\delta^{p_{ij}} | x_i = x, x_j = \tilde{x}]$. Suppose that for two types, $x, \tilde{x}$ there is not perfect sorting, and in particular there is some mixing between them, i.e. $\omega_{x\tilde{x}} > 0$; the average per agent utility is:

$$\left[\frac{(\kappa-1)\,\delta}{1-(\kappa-1)\,\delta} - \omega_x\right] \cdot z(x,x) + \omega_x \cdot z(x,\tilde{x}).$$

Each agent will almost surely have κ links as it is assumed that each link adds positive value and there are asymptotic infinite agents (only a finite number can then not fulfill the degree quota).

As we have a finite set of types we can assume then for large populations there is a subset of types, $\hat{X} \subseteq X$, where for every type $x \in \hat{X}$ it holds that there is an asymptotic strictly positive share of the total number of agents of that type, i.e., $\lim_{n\to\infty}(|\{i \in N_n\}_{x_i=x}|/n) > 0$. If there is only one such type, i.e. $|\hat{X}| = 1$, then asymptotic perfect sorting follows by assumption as the asymptotic number of links is κ.

For any two types $x, \tilde{x} \in \hat{X}$ which are mixing their average utility is:

$$\frac{(\kappa-1)\,\delta}{1-(\kappa-1)\,\delta}\left[\frac{n_x \cdot z(x,x) + n_{\tilde{x}} \cdot z(\tilde{x},\tilde{x})}{n_x + n_{\tilde{x}}}\right] - \frac{1}{2} \cdot \left[\frac{n_x \cdot \omega_{x\tilde{x}}}{n_x + n_{\tilde{x}}}\right] \cdot [Z(x,x) + Z(\tilde{x},\tilde{x}) - 2Z(x,\tilde{x})].$$

As there is supermodularity it follows that $Z(x,x) + Z(\tilde{x},\tilde{x}) - 2Z(x,\tilde{x}) > 0$ and thus mixing must decrease utility. The same argument can be applied by mixing between multiple types. ■

B Supplementary appendix: Externalities with finite poulation and constant decay

This appendix shows how suboptimally sorted networks are also prevalent under constant decay. It is split into two sub-appendices: sub-appendix B.1 which deals with demonstrating the results and sub-appendix B.2 which only contains auxiliary results.

B.1 Suboptimal sorting in local trees

We show sorting may be pairwise stable but suboptimal under constant decay for a subclass of networks. We begin by describing this subclass. Informally put, the relevant subclass of perfectly sorted networks where each subnetwork for a given type has a certain structure. The structure of each subnetwork is such that from the perspective of every agent (i.e. the ego-network) each subnetwork appears as a tree when disregarding the links of the agents furthest away. Note that a *tree* is network where every pair of agents are connected by a unique path. Thus these subnetworks are called local trees as they are not trees in a global sense but only when disregarding most distant agents.

The formal definition is as described below. The definition employs the network *diameter* which is the maximum distance between any two agents, i.e. $m(\mu) = \sup_{i,j \in N} p_{ij}(\mu)$.

Definition 9. *A network μ is a **local tree** when each agent i has κ links where:*

- *for each other agent $j \neq i$ at distance $p_{ij}(\mu) \leq m_{n,\kappa} - 2$ there are $\kappa - 1$ links between agent j and j' such that j' is one step further away, i.e. $p_{ij}(\mu) = p_{ij'}(\mu) - 1$;*
- *the network diameter $m(\mu) = m_{n,\kappa}$,*

$$m_{n,\kappa} \quad = \quad \arg\min_{m} \{m : \Sigma_{l=1}^{m}(\kappa(\kappa-1)^{l-1}) + 1 \geq n\}. \tag{29}$$

The structure of local trees entails that each agent has $\kappa \cdot (\kappa-1)^{p-1}$ agents at distance $p < m$, where $m = m_{n,\kappa}$. At distance $p = m$ there are $n - \sum_{l=1}^{m-1} \kappa \cdot (\kappa-1)^{l-1}$ (all remaining agents). This structure implies that every agent's utility is maximized subject to the constraint of all agents having at most κ links;[18] a side effect is that utility before transfers is symmetric.

A necessary condition for local trees to exist is that there is no link surplus, i.e. degree quota is binding ($\forall i \in N : k_i = \kappa$). Note this binding condition is only possible when $n \cdot \kappa$ is even.

When a local tree network fulfills $n = \sum_{l=1}^{m} \kappa \cdot (\kappa - 1)^{l-1}$ then it is an *exact local tree*. See the next sub-appendix for an elaborate treatment of structure of exactly local trees. Two subclasses of exact local trees which are worth mentioning. The first is a network known as a cycle or a ring. The cycle is characterized by having a minimal possible degree quota ($\kappa = 2$) among local trees and a maximal diameter ($m = \lceil \frac{n-1}{2} \rceil$). The second is a *clique* where all agents are linked, i.e. the complete network. Cliques have maximal degree quotas ($\kappa = n - 1$) and minimal diameters ($m = 1$). Both subclasses has a network which exists for any n. Note that in Example 1 each of the two components is both a cycle and a clique. Note that there exist non-trivial networks beyond the cycle and the clique.[19]

In order to derive our results it is necessary to restrict ourselves to a subset of local trees. The subset are those local trees where the deletion of links leads to equal losses to both of agents whose link is deleted; thus we refer to these local trees as having symmetric losses:

Definition 10. *A local tree μ has* ***symmetric losses*** *when at every distance $p = 1, .., m$ it holds that $|\{i \in N : p_{\iota i}(\mu \backslash \{\iota\iota'\}) = p\}| = |\{i \in N : p_{\iota' i}(\mu \backslash \{\iota\iota'\}) = p\}|$.*

Denote the set of perfectly sorted networks where the subnetwork for each type is a local tree with symmetric losses as $M^{p-srt:symm.\,loc-tree}$.

Whether or not symmetric losses is a generic property for all local trees is an open question. However, in simulations that we perform it holds all network configurations which are local trees up to size $n = 10$ have symmetric losses (see result below and proof for exhibition of examples). Moreover for size up to $n = 16$ it has been shown to hold for any networks examined in the simulation.

A generalization of stable but suboptimal sorting under constant decay is expressed below. While allowing for constant decay rather than hyperbolic it the set of networks are further restricted.

Theorem 3. *Suppose there is supermodularity, a degree quota κ and each type has equal number of agents then*

(i) $\hat{M} \subseteq M^{p-stb}_{\delta \leq \bar{\delta}}$;

(ii) $\hat{M} \cap M^{\max U}_{\delta > \underline{\delta}} = \emptyset$;

where $\hat{M} = M^{p-srt:symm.\,loc-tree}$ *and thresholds* $\underline{\delta}, \bar{\delta} \in (0, 1)$ *where* $\underline{\delta} < \bar{\delta}$

[18] The maximization of utility follows from the observation that each agent has at most κ links, so at distance p there can be at most $\kappa \cdot (\kappa - 1)^{p-1}$ agents.

[19] An example is $\{i_1i_2, i_1i_3, i_1i_4, i_2i_5, i_2i_6, i_3i_7, i_3i_8, i_4i_9, i_4i_{10}, i_5i_7, i_5i_9, i_6i_8, i_6i_{10}, i_7i_9, i_8i_{10}\}$ when $n = 10, \kappa = 3$ and $N = \{i_1, i_2, .., i_{10}\}$.

Proof. We show properties (i) and (ii) together. Let μ be a network which is segregated into $|X|$ components where each component is a local tree with $n/|X|$ agents. Let there be no transfers between any agents.

As each subnetwork for a given type is a local tree it is stable against deviations by agents of the same type - this follows as local trees provides maximal possible benefits among feasible structures of the subnetwork for all agents in the subnetwork. Thus only two agents of different types may have a profitable deviation which is feasible.

Let ι, j be agents of respectively types x and $\tilde{x}$. These two agents can deviate by each deleting a link to ι' and j' respectively while jointly forming a link. The new network resulting from deletion is denoted $\hat{\mu} = \mu \backslash \{\iota\iota', jj'\}$. The move resulting from deletion and forming a link is denoted $\breve{\mu} = \hat{\mu} \cup \{\iota j\}$. An alternative network is $\tilde{\mu}$, the type-bridged network of μ, where the links $\iota\iota', jj'$ are removed while the links $\iota j, \iota', j'$ have been formed; thus $\tilde{\mu} = \hat{\mu} \cup \{\iota j, \iota' j'\}$.

Define the gross loss of benefits for i as $u_i(\hat{\mu}) - u_i(\mu)$ while the gross gains are $u_i(\tilde{\mu}) - u_i(\hat{\mu})$. There must exist a threshold of externalities $\bar{\delta} \in (0,1)$ where μ is no longer pairwise stable as cost of deviation monotonically decreases and approaches zero as $\delta \to 1$ while gains are monotonically increasing. The monotonicity of losses is a consequence of the fact that gross loss consists of shortest paths from μ, where $\iota\iota'$ is included in the shortest path, which have longer length in $\hat{\mu}$ and thus are discounted more. Therefore the gross loss is mitigated by a higher δ as the longer shortest paths are punished less. The monotonicity of gains follows as the gains consist of new shortest paths to agents of type $\tilde{x}$ through ιj and $j'\iota'$ the value of these increases for higher δ.

Exploiting the that Fact 1 and 2 from Appendix B.2 hold for local trees it follows that for any other agent i of type x (i.e. i is in $N\backslash\{\iota, \iota'\}$ and $x_i = x$):

$$u_i(\tilde{\mu}) - u_i(\mu) > \delta^{\min(p_{i\iota}(\tilde{\mu}), p_{i\iota'}(\tilde{\mu}))}[u_\iota(\breve{\mu}) - u_\iota(\mu)].$$

Aggregating for all agents this implies:

$$U(\tilde{\mu}) - U(\mu) \quad > \quad [u_\iota(\breve{\mu}) - u_\iota(\mu)] \cdot \sum_{x_i = x} \delta^{\min(p_{i\iota}(\tilde{\mu}), p_{i\iota'}(\tilde{\mu}))} + [u_j(\breve{\mu}) - u_j(\mu)] \cdot \sum_{x_i' = \tilde{x}} \delta^{\min(p_{ij}(\mu), p_{ij'}(\mu))}.$$

where $m = m_{n,\kappa}$. The inequality above implies the following: if $U(\tilde{\mu}) - U(\mu) = 0$ then $u_\iota(\breve{\mu}) - u_\iota(\mu) + u_j(\breve{\mu}) - u_j(\mu) < 0$; $U(\tilde{\mu}) - U(\mu) > 0$ when $u_\iota(\breve{\mu}) - u_\iota(\mu) + u_j(\breve{\mu}) - u_j(\mu) = 0$. It can also be argued that there must exist a threshold, $\underline{\delta}$, such that when $\delta = \underline{\delta}$ then $U(\tilde{\mu}) - U(\mu) = 0$ and that $\underline{\delta} < \bar{\delta}$. This follows as $U(\tilde{\mu}) - U(\mu) < 0$ for $\delta = 0$ and $U(\tilde{\mu}) - U(\mu) > 0$ when $u_\iota(\tilde{\mu}) - u_\iota(\mu) + u_j(\tilde{\mu}) - u_j(\mu) = 0$ as well as continuity of $U(\tilde{\mu}) - U(\mu)$ in δ.

This entails that for $\delta > \underline{\delta}$ then $\tilde{\mu}$ provide higher aggregate payoff. Moreover we showed previously that for $\delta < \bar{\delta}$ then μ is pairwise (Nash) stable. Thus we have proven properties (i) and (ii).

□

For constant decay the thresholds governing when sorting is respectively suboptimal and stable, i.e. $\underline{\delta}, \bar{\delta}$, can be determined explicitly by solving polynomial equations for every deviation. Moreover, for exact local trees there is a unique solution. In Figure 3 the two thresholds from Theorem 3, $\underline{\delta}(\hat{Z}), \bar{\delta}(\hat{Z})$.

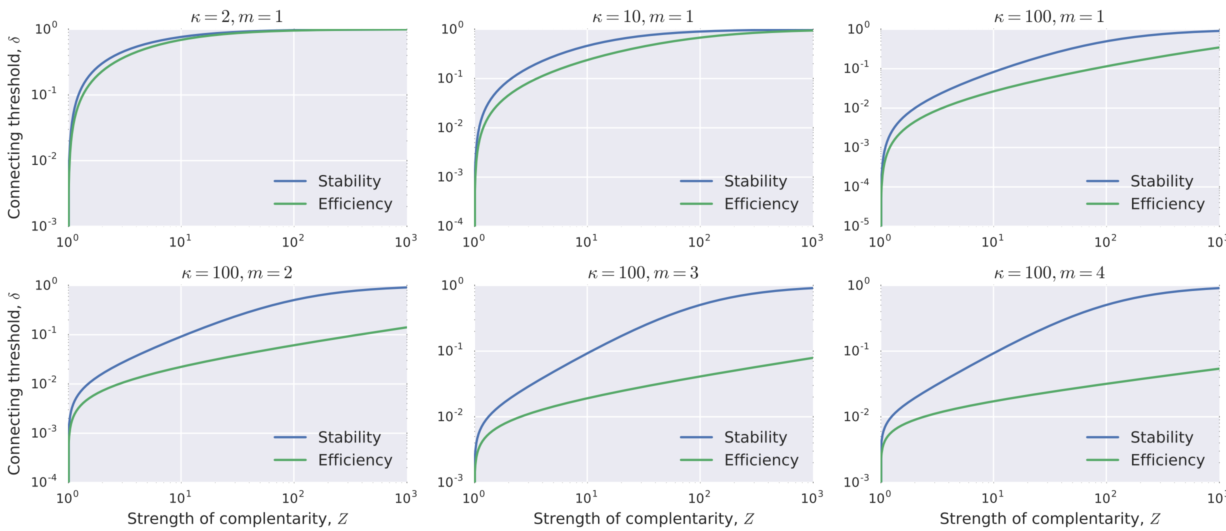


Figure 3: Visualization of thresholds for connecting from Theorem 3. The upper diagrams correspond to cliques and the lower ones to exact local trees (where thresholds stem from Equations 32, 33, 38, 39).

The plots in Figure 3 are made for variations of exact local trees. The upper plots corresponds to cliques with various sizes. The lower plot have fixed degree quota (κ=100) and the threshold is simulated using pattern in utility that is demonstrated in Appendix B.2. The plots show the scope for inefficiency, i.e. the gap between $\underline{\delta}(\hat{Z}), \bar{\delta}(\hat{Z})$, increases with the number of agents involved. This makes sense intuitively as the two agents forming the link will fail to account for an increasing number of indirect connections between the two groups. As the number of indirect connections increases at with the squared with total number of agents then larger populations will lead to larger gaps of inefficiency.

B.2 Local trees

This sub-appendix provides auxiliary results for deriving the generalization of suboptimal sorting. We begin our focus on exact local trees and subsequently more generally in local tree networks, see Definition 9 in the previous sub-appendix.

We will examine a generic network μ which is perfectly sorted and assume that the subset of links for each type is a component that can be classified as either a local tree or an exact local tree. Let networks μ_x and $\mu_{\tilde{x}}$ be the components associated with respectively types $x, \tilde{x} \in X$. We will focus on three particular moves:

- *Pairwise deletion of a link*: Suppose two links $\iota\iota', jj' \in \mu$ are deleted and agents ι and j

have respectively type x and $\tilde{x}$; thus the two links are not from the same component. Let the new network that results from removal of the links be denoted $\hat{\mu} = \mu \backslash \{\iota\iota', jj'\}$.

- *Pairwise formation of a link across types*: This move presumes that both agents are also deleting a link. We denote this as a move where agents ι and j form a link: $\breve{\mu} = \hat{\mu} \cup \{\iota j\}$.
- *Double pairwise formation of a link across types*: When two links are formed across types in μ this corresponds to a non-pairwise deviation as it requires four coalition members. We denote this as a move where both agents ι and j as well as ι' and j' form a link: $\tilde{\mu} = \hat{\mu} \cup \{\iota j, \iota' j'\}$.

Finally let i denote a generic agent of type x. Let the shortest path in μ from i to either ι or ι' be denoted $\hat{p}_i$ where $\hat{p}_i = \min(p_{i\iota}(\hat{\mu}), p_{i\iota'}(\hat{\mu}))$. When $\hat{p}_i = 0$ then either $i = \iota$ or $i = \iota'$.

Basic properties We exploit that μ is a local tree (see Definition 9). Throughout the remainder of the paper let $m = m_{n,\kappa}$ (see Equation 29). We express each agent's number of paths of length p as a function of the number of agents and the degree quota:

$$\#_i^p(\mu) = \kappa(\kappa-1)^{p-1} - \mathbf{1}_{=m}(p) \cdot \Delta\#(n,\kappa), \qquad \Delta\#(n,\kappa) = \sum_{l=1}^{m} (\kappa \cdot (\kappa-1)^{l-1}) - n, \tag{30}$$

where $\mathbf{1}_{=m}(p)$ is the Dirac measure of whether $p = m$. Using the local tree structure we can express utility without transfers of each agent:

$$u_i(\mu) = \sum_{l=1}^{m} \#_i^l(\mu) \cdot \delta^l \cdot z(x,x).$$

Exact local trees

Recall exact local trees are local trees where $\Delta\#(n,\kappa) = 0$. We will argue that this entails that exact local trees have the essential property that for every pair of agents there is a unique shortest path of at most length m and the number of paths for every agent is prescribed by Equation 30. This can be deducted as follows.

Note first that the fact that the number of walks with at most length m starting in a given agent i cannot exceed $\sum_{p=1}^{m} \#_i^p(\cdot)$. Recall also that local trees has the property that all agents are reached within distance m. Moreover exact local trees has the property that for any agent i it holds that $n - 1 = \sum_{p=1}^{m} \#_i^p(\mu)$; thus every shortest path with distances less than or equal to m must be a unique path between the two particular agents.

The uniqueness and countability of paths can be used to infer the losses when links are either removed or added to an exact local tree.

Exact local trees - loss from deletion In order to examine the impact of deletion of a link it is sufficient to analyze what happens to one component of types. This is sufficient as other components as the conclusions are valid for all.

The deletion of link $\iota\iota'$ implies that any pair of agents i, i' whose (unique) shortest path in μ includes the link $\iota\iota'$ will have a new shortest routing path. For exact local trees we can exactly determine the length of the new path. Let i be the agent whose distance to ι is least and let i' be the agent whose distance to ι' is least, i.e. $p_{i\iota}(\mu) < p_{i\iota'}(\mu)$ and $p_{i'\iota'}(\mu) < p_{i'\iota}(\mu)$.

First when link $\iota\iota'$ is deleted we can show there is no shortest path between i and i' in $\hat{\mu}$ with length below $2m - \hat{p}_i - \hat{p}_{i'}$; that is there is no ii' whose shortest path in μ includes $\iota\iota'$ such that $p_{ii'}(\hat{\mu}) < 2m - \hat{p}_i - \hat{p}_{i'}$. Suppose this was not true. Then there would exist an agent j who (1) is on the new shortest path between i and i' in $\hat{\mu}$ and (2) whose shortest path to agents ι and ι' does not include the link $\iota\iota'$ and (3) such that

$$\begin{aligned}
p_{ji}(\hat{\mu}) + p_{ji'}(\hat{\mu}) &< 2m - \hat{p}_i - \hat{p}_{i'}, \\
p_{ji}(\hat{\mu}) + p_{ji'}(\hat{\mu}) &< 2m - \min(p_{i\iota}(\mu), p_{i\iota'}(\mu)) - \min(p_{i'\iota}(\mu), p_{i'\iota'}(\mu)).
\end{aligned}$$

As by construction $p_{i\iota}(\mu) < p_{i\iota'}(\mu)$ and $p_{i'\iota'}(\mu) < p_{i'\iota}(\mu)$ then the expression above is equivalent to: $p_{ji}(\hat{\mu}) + p_{ji'}(\hat{\mu}) < 2m - p_{i\iota}(\mu) - p_{i'\iota'}(\mu)$. As the shortest path between i and ι as well as between i' and ι' are unchanged from μ to $\hat{\mu}$ it follows that we can further rewrite into:

$$p_{ji}(\hat{\mu}) + p_{ji'}(\hat{\mu}) < 2m - p_{i\iota}(\hat{\mu}) - p_{i'\iota'}(\hat{\mu})$$

However, the above statement implies that in network μ that either ι or ι' has two paths with lengths of at most m but this violates the definition of exact local trees.

We can now show that when link $\iota\iota'$ is deleted the new shortest path between i and i' in $\hat{\mu}$ has a length of exactly $2m - \hat{p}_i - \hat{p}_{i'}$. This is shown by demonstrating there is an agent j such that $p_{ji}(\hat{\mu}) = m - \hat{p}_i$ and $p_{ji'}(\hat{\mu}) = m - \hat{p}_{i'}$. This can be shown follows. Suppose that $p_{ji}(\hat{\mu}) = m - \hat{p}_i$. We will demonstrate that $p_{ji'}(\hat{\mu}) = m - \hat{p}_{i'}$. As $p_{ji}(\hat{\mu}) = m - \hat{p}_i$ it follows that $p_{j\iota}(\hat{\mu}) = m$. From the definition of exact local trees there must exist a path of length less than m between j and ι' in network μ. As argued in the paragraph above neither of these paths can be strictly shorter than m and consequently they must both be exactly m.

The number of shortest paths of length p which become altered for agent i is $(\kappa - 1)^{p - \hat{p}_i - 1}$ for $p = \hat{p}_i, .., m-2, m-1$. This can be demonstrated as follows. If agent $p_{i\iota}(\mu) = m$ and $p_{i\iota'}(\mu) = m$ then no shortest paths are altered; this is clear as agent i as none of the unique shortest paths includes $\iota\iota'$ as they have at most length m. If instead $p_{i\iota}(\mu) = m - 1$ then the unique shortest path from i to ι' includes $\iota\iota'$ is the last link; this implies a new shortest path if $\iota\iota'$ is deleted. Thus if $p_{i\iota}(\mu) = m - 1$ then one shortest path of length m is lost. When $p_{i\iota}(\mu) = m - 2$ then one path of length $m - 1$ is lost by the same argument; moreover $\kappa - 1$ paths that has $\iota\iota'$ as the second last link. By induction this can be done at higher order and thus for shorter distances. Using the number of rerouted paths shown above we can establish the total number of shortest paths in network $\hat{\mu}$ for agent i that has a length of p:

$$\#_i^p(\hat{\mu}) = \begin{cases} \kappa(\kappa-1)^{p-1} - \mathbf{1}_{>\hat{p}_i}(p) \cdot (\kappa-1)^{p-\hat{p}_i-1}, & p \leq m \\ (\kappa-1)^{2m-\hat{p}_i-p}, & p \in (m, 2m-\hat{p}_i]. \end{cases} \tag{31}$$

By combining the count of shortest paths rerouted with their new length we can generalize the loss for any agent from the deletion of link $\iota\iota'$ when all agents are homogeneous of type x:

$$u_i(\mu) - u_i(\hat{\mu}) = \sum_{l=1}^{m-\hat{p}_i} \left[(\kappa-1)^{l-1} \cdot \left(\delta^{l-1+\hat{p}_i} - \delta^{2m-(l-1)-\hat{p}_i}\right)\right] \cdot z(x,x). \tag{32}$$

We can aggregate the losses across homogeneous agents of type x and we arrive at the following expression:

$$U(\mu) - U(\hat{\mu}) = \sum_{l=1}^{m} \left[2l \cdot (\kappa-1)^{l-1} \cdot \left(\delta^{l-1} - \delta^{2m-(l-1)}\right)\right] \cdot z(x,x). \tag{33}$$

Exact local trees - gains from linking across types We move on to establishing the gains of establishing a link in a perfectly sorted network where each component is an exact local tree.

The gains to agents ι and j of forming a link ιj are direct benefits and the new indirect connections that are accessed through the link ιj. For agent ι the benefits from forming a link with j can be computed with Equation 31 where the input length is added one (as ιj is added to the shortest path). Recall $\breve{\mu} = \mu \cup \{\iota j\} \backslash \{\iota\iota', jj'\}$.

$$u_\iota(\breve{\mu}) - u_\iota(\mu) = \left[\sum_{l=0}^{m} (\kappa-1)^l \cdot \delta^l + \sum_{l=0}^{m-1} (\kappa-1)^l \cdot \delta^{2m-l}\right] \cdot z(x,\tilde{x}). \tag{34}$$

The above expression is relevant for evaluating the pairwise gains as it captures individual benefits for a pairwise formation of a link by ι and j. However, we are also interested in the sub-connected network as it allows to assess the efficiency. Suppose instead now that ι' and j' also form a link; thus $\iota j, \iota' j'$ are formed while $\iota\iota', jj'$ are deleted. Let $\tilde{\mu} = \mu \cup \{\iota j, \iota' j'\} \backslash \{\iota\iota', jj'\}$.

Let i be an agent of type x and let $\hat{p}_i$ still denote the least distance to either ι or ι'. We can calculate the benefits for i when $\iota j, \iota' j'$ are formed. The benefits are the indirect connections to agents of type $\tilde{x}$ with whom agent i has no connections in μ. The aim is to count the number of paths of a given length.

For a given agent i' of the other type $\tilde{x}$ it must hold that the shortest path in $\tilde{\mu}$ between i, i' either contains the link ιj or the link $\iota' j'$, and thus the distance can be computed as follows:

$$p_{ii'}(\tilde{\mu}) \quad = \quad \min[p_{ij}(\tilde{\mu}) + p_{i'j}(\tilde{\mu}),\ p_{ij'}(\tilde{\mu}) + p_{i'j'}(\tilde{\mu})] \tag{35}$$

We further restrict the above expression. We can use that i and i' of type $\tilde{x}$ can be at most $2m+1$ away from each other. This follows from the fact that $p_{i\iota}(\tilde{\mu}) + p_{i\iota'}(\tilde{\mu}) = 2m$ and $p_{i'j}(\tilde{\mu}) + p_{i'j'}(\tilde{\mu}) = 2m$. As $p_{i\iota}(\tilde{\mu}) + p_{i\iota'}(\tilde{\mu}) = 2m$ and $\iota j, \iota' j' \in \tilde{\mu}$ then it must be that $p_{ij} + p_{ij'} = 2m+2$. These facts together entail we can rewrite Equation 35:

$$\begin{aligned} p_{ii'}(\tilde{\mu}) &= \min[p_{ij}(\tilde{\mu}) + p_{i'j}(\tilde{\mu}),\ p_{ij'}(\tilde{\mu}) + p_{i'j'}(\tilde{\mu})] \\ &= \min[p_{ij}(\tilde{\mu}) + p_{i'j}(\tilde{\mu}),\ 4m+2 - p_{ij}(\tilde{\mu}) - p_{i'j}(\tilde{\mu})]. \end{aligned} \tag{36}$$

From the above expression it follows that $p_{ii'} \leq 2m+1$ as the expression is maximized for $p_{ij} + p_{i'j} = 2m+1$.

The number of shortest paths from i through ιj to agents of the other type $\tilde{x}$ can be found using Equation 31 for agent ι adding extra distance $1+\hat{p}_i$:[20]

- for distance $p \in \{1+\hat{p}_i, ..., m+1+\hat{p}_i\}$ there are $(\kappa-1)^{p-1-\hat{p}_i}$ agents;
- for distance $p \in \{m+2+\hat{p}_i, ..., 2m+1\}$ there are $(\kappa-1)^{2m+1-(p-1-\hat{p}_i)}$.

The shortest paths from i not routed through ι but instead through ι' are those where $p+1+\hat{p}_i > 2m+1$; from Equation 36 we know the new shortest path length is $4m+2-p-1-\hat{p}_i$. The number of shortest paths through ι' in network $\tilde{\mu}$ will be $(\kappa-1)^{2m+1-(p-1-\hat{p}_i)}$ and the new length $4m+2-p-1-\hat{p}_i$. These facts together imply:

$$\#_i^p(\tilde{\mu}) - \#_i^p(\hat{\mu}) = \begin{cases} (\kappa-1)^{p-1-\hat{p}_i}, & p \in \{\hat{p}_i+1, .., m+1+\hat{p}_i\}, \\ (\kappa-1)^{2m+1-p-\hat{p}_i}, & p \in \{m+\hat{p}_i+2, .., 2m+1\}, \\ (\kappa-1)^{p+\hat{p}_i-2m-1}, & p \in \{2m+1-\hat{p}_i, .., 2m\}. \end{cases} \tag{37}$$

From the number of paths above we can derive the change in utility from when $\iota j, \iota' j'$ are added to the network for a given agent i of type x.

$$u_i(\tilde{\mu}) - u_i(\hat{\mu}) = \begin{bmatrix} & \sum_{l=0}^{m}(\kappa-1)^l \cdot \delta^{l+\hat{p}_i} \\ + & \sum_{l=\hat{p}_i}^{m-1}(\kappa-1)^l \cdot \delta^{2m-l+\hat{p}_i} \\ + & \sum_{l=0}^{\hat{p}_i-1}(\kappa-1)^l \cdot \delta^{2m+l-\hat{p}_i} \end{bmatrix} \cdot z(x,\tilde{x}). \tag{38}$$

By aggregating over all agents of type the gain in benefits by forming $\iota j, \iota' j'$ is as follows:

$$U(\tilde{\mu}) - U(\hat{\mu}) = \sum_{p=0}^{m} \left(\begin{bmatrix} \mathbf{1}_{<m}(p) \cdot 2 \cdot (\kappa-1)^p + \\ \mathbf{1}_{=m}(p) \cdot (n - 2 \cdot \sum_{l=1}^{m-1}(\kappa-1)^l) \end{bmatrix} \cdot \begin{bmatrix} & \sum_{l=0}^{m}(\kappa-1)^l \cdot \delta^{l+p} \\ + & \sum_{l=p}^{m-1}(\kappa-1)^l \cdot \delta^{2m-l+p} \\ + & \sum_{l=0}^{p-1}(\kappa-1)^l \cdot \delta^{2m+l-p} \end{bmatrix} \right) \cdot Z(x,\tilde{x}). \tag{39}$$

[20]Shortest paths from i must contain both ιj and every link in the shortest path from i to j.

Local trees

We can use the analysis above on exact local trees to bound the gains and losses for (non-exact) local trees. Recall that exact local trees has the property that $\Delta\#(n, \kappa) = 0$ and for non-exact local trees $\Delta\#(n, \kappa) > 0$. Thus the difference between exact and non-exact local trees is that for a given agent the number of connected other agents at exactly distance m is lower for non-exact local trees.

Using the analysis of exact local trees we can compute the bounds on loss of utility for a given agent in the local when a link is deleted - this is done by reusing Equation 31 as follows.

We can discount the number of agents initially at distance m by $\Delta\#(n, \kappa)$. Moreover, the new distance between agents i and i' after deletion of the link $\iota\iota'$ is at least $\min(p_{ii'}, 2m - 2 - \hat{p}_i - \hat{p}_{i'})$ at most $2m - \hat{p}_i - \hat{p}_{i'}$.[21] From these two facts we can derive the bound on loss of utility when $\iota\iota'$ is deleted. The upper bound on loss (in terms of magnitude) is when new shortest paths have most distance, i.e. $2m - \hat{p}_i - \hat{p}_{i'}$; the lower bound is found when new distance is least, i.e. $\min(p_{ii'}, 2m - 2 - \hat{p}_i - \hat{p}_{i'})$:

$$u_i(\mu) - u_i(\hat{\mu}) \quad \leq \quad \sum_{l=1}^{m-\hat{p}_i} \left[\max(0, (\kappa-1)^{l-1} - \mathbf{1}_{=m}(l) \cdot \Delta\#(n, \kappa)) \left(\delta^{l-1+\hat{p}_i} - \delta^{2m-(l-1)-\hat{p}_i}\right)\right] \cdot z(x, x), \quad (40)$$

$$u_i(\mu) - u_i(\hat{\mu}) \quad \geq \quad \sum_{l=1}^{\tilde{m}} \left[(\kappa-1)^{l-1} \cdot \left(\delta^{l-1+\hat{p}_i} - \delta^{2m-(l+1)-\hat{p}_i}\right)\right] \cdot z(x, x), \quad \tilde{m} = \min(m-1, m-\hat{p}_i). \quad (41)$$

Fact 1. *If μ is perfectly sorted and consists of $|X|$ components that each constitute a local tree with $n/|X|$ agents, then for any agent i of type x where $\hat{p}_i > 0$:*

$$u_i(\hat{\mu}) - u_i(\mu) > \delta^{\hat{p}_i} \cdot [u_\iota(\hat{\mu}) - u_\iota(\mu)], \quad \hat{p}_i = \min(p_{i\iota}(\hat{\mu}), p_{i\iota'}(\hat{\mu})). \quad (42)$$

Proof. Inequality 42 can be rewritten into: $\delta^{\hat{p}_i} \cdot [u_\iota(\mu) - u_\iota(\hat{\mu})] - [u_i(\mu) - u_i(\hat{\mu})] > 0$. This inequality is equivalent to the expression below (derived by substituting in Inequality 41 for agent ι and Inequality 40 for agent i):

$$\delta^{\hat{p}_i} \cdot \sum_{l=1}^{m-1} \left[(\kappa-1)^{l-1} \cdot \left(\delta^{l-1} - \delta^{2m-(l+1)}\right)\right] - \sum_{l=1}^{m-\hat{p}_i} \left[(\kappa-1)^{l-1} \left(\delta^{l-1+\hat{p}_i} - \delta^{2m-(l-1)-\hat{p}_i}\right)\right] \quad > \quad 0,$$

$$\sum_{l=1}^{m-\hat{p}_i} \left[(\kappa-1)^{l-1} \cdot \left(\delta^{2m-(l+1)-\hat{p}_i} - \delta^{2m-(l+1)+\hat{p}_i}\right)\right] + \sum_{l=m-\hat{p}_i+1}^{m-1} \left[(\kappa-1)^{l-1} \left(\delta^{l-1+\hat{p}_i} - \delta^{2m-(l-1)-\hat{p}_i}\right)\right] \quad > \quad 0.$$

As it holds that $2m-(l+1)-\hat{p}_i < 2m-(l+1)+\hat{p}_i$ and it holds that $l-1+\hat{p}_i < 2m-(l-1)-\hat{p}_i$ (equivalent to $l < m+1-\hat{p}_i$) the above inequality is satisfied. □

[21] The upper bound follows from the fact that for any two agents i and i' in the local tree there is still always an agent j at distances $p_{ij} = m - \hat{p}_i$ and $p_{i'j} = m - \hat{p}_{i'}$. The lower bound can be established by repeating an argument used for exact local trees. If the new distance between two agents i and i' after deletion of $\iota\iota'$ had been less than $\min(p_{ii'}(\mu), 2m - 2 - \hat{p}_i - \hat{p}_{i'})$ then the following would be true. There would be multiple shortest paths of length less than or equal to $m - 1$ between either (ι and j) or (ι' and j). This would violate the property of local trees that all shortest paths of length $\leq m - 1$ are unique.

We can also derive bounds on the gains from connecting across types for local trees. We will not do this explicitly but instead use Definition 10 on symmetric losses in local trees. This allows to express our next result:

Fact 2. *For the perfectly sorted network μ which consists of $|X|$ network components which each constitute a local tree of $n/|X|$ agents that has symmetric losses then it holds that for agents i, ι of type x and $\hat{p}_i > 0$*

$$u_i(\tilde{\mu}) - u_i(\hat{\mu}) \geq \delta^{\hat{p}_i} \cdot [u_\iota(\breve{\mu}) - u_\iota(\hat{\mu})], \quad \hat{p}_i = \min(p_{i\iota}(\hat{\mu}), p_{i\iota'}(\hat{\mu})). \tag{43}$$

Proof. It holds that $u_\iota(\tilde{\mu}) - u_\iota(\hat{\mu}) \geq u_\iota(\breve{\mu}) - u_\iota(\hat{\mu})$ as $\tilde{\mu} \subseteq \breve{\mu}$ (thus all shortest paths in $\tilde{\mu}$ cannot have a length that exceeds that in $\breve{\mu}$). Therefore it suffices to show:

$$u_i(\tilde{\mu}) - u_i(\hat{\mu}) \geq \delta^{\hat{p}_i} \cdot [u_\iota(\tilde{\mu}) - u_\iota(\hat{\mu})]. \tag{44}$$

As the local tree has symmetric losses it follows that $u_\iota(\tilde{\mu}) - u_\iota(\hat{\mu}) = u_{\iota'}(\tilde{\mu}) - u_{\iota'}(\hat{\mu})$; this follows from the fact that they both gain an equal number of new shortest paths through j, j', this follows as as j, j' have same number of paths after deletion of jj' due to symmetric losses. This entails that without loss of generality we can assume that $p_{i\iota} = \hat{p}_i$ as otherwise we could substitute ι with ι' and conduct the analysis again.

For ι and some agent i' of type $\tilde{x}$ it holds that $p_{ii'}(\tilde{\mu}) \leq p_{\iota i'}(\tilde{\mu}) + \hat{p}_i$. This follows as there exists a path between i, ι and ι, i' with respectively lengths $p_{\iota i'}(\tilde{\mu})$ and $\hat{p}_i$; thus $p_{ii'}(\tilde{\mu}) \leq p_{\iota i'}(\tilde{\mu}) + \hat{p}_i$. This implies the following inequality must hold:

$$\sum_{x_{i'} = \tilde{x}} \delta^{p_{ii'}(\tilde{\mu})} \geq \delta^{p_{\iota i}(\tilde{\mu})} \cdot \sum_{x_{i'} = \tilde{x}} \delta^{p_{\iota i'}(\tilde{\mu})}.$$

As $u_\iota(\tilde{\mu}) - u_\iota(\hat{\mu}) = \sum_{x_i = \tilde{x}} \prod_{l=1}^{p_{\iota i'}(\tilde{\mu})} \delta^{r_l} \cdot z(x, \tilde{x})$ and $u_i(\tilde{\mu}) - u_i(\hat{\mu}) = \sum_{x_i = \tilde{x}} \prod_{l=1}^{p_{ii'}(\tilde{\mu})} \delta^{r_l} \cdot z(x, \tilde{x})$ it follows that Inequality 44 holds which proves our fact.

□