Know-how and know-who: Effects of a randomized training on network changes among small urban entrepreneurs

Mattea Stein*

November 17, 2018

Latest version here.

Abstract

Micro-enterprise owners in developing country industrial clusters interact through multi-layered networks of horizontal business collaboration, information-sharing, and friendship links, despite the potential for close competition inherent in this setting. This paper explores whether such business links can be endogenous to a public policy intervention that provides training to some network members but not others. Using a randomized training for micro-entrepreneurs in Kampala, Uganda, together with novel panel network data, I find a positive effect on linking likelihoods, driven by untreated entrepreneurs to whom links with treated entrepreneurs become more desirable. As predicted by a bilateral network formation framework, it is the relatively lower-status treated who attract new connections with relatively higher-status untreated. Furthermore, links within clusters of treated enterprises are strengthened, which is not due to a strategic replacement of untreated with treated partners out of a competition motive but seems to be an effect of jointly attending the training. Together, my findings show that public policy interventions can cause networks to re-wire, with important implications both for research and policy.

Keywords: Network formation, network change, social networks, firms, micro-enterprises

*Paris School of Economics and EHESS; mattea.stein@psemail.eu. I thank Abhijit Banerjee, Luc Behaghel, Margherita Comola, Denis Cogneau, Esther Duflo, Alice Duhaut, Marcel Fafchamps, Markus Goldstein, Matthew Jackson, Florence Kondylis, Michael Kremer, Sylvie Lambert, Karen Macours, Norihiko Matsuda, Simon Quinn, Natalia Rigol, Daniel Rogger, Akiko Suwa-Eisenmann, Christopher Udry, Oliver Vanden Eynde, Christopher Woodruff, Liam Wren Lewis and seminar participants at the NEUDC 2018 conference, the CSAE 2018 conference and the Paris School of Economics for insightful discussions and feedback. I am indebted to the team at Innovations for Poverty Action (IPA) Uganda for their excellent field work and outstanding dedication. Generous funding from the EHESS and the World Bank's Africa Gender Innovation Lab and its Development Impact Evaluation Initiative (DIME) is gratefully acknowledged.

1 Introduction

Social interactions matter for economic decision making and outcomes, a fact that is increasingly recognized and studied in the economic literature. Networks help shape markets and many aspects of economic life (Granovetter (2005), Jackson (2008, 2011)). One of their fundamental functions is as a conduit of information between individuals and as a medium that can spread behaviors through social learning; examples have been observed widely, including in employment search, health behaviours, and migration.[1] In development economics, a focus has been information diffusion and social learning in agricultural technology adoption, beginning with Foster and Rosenzweig (1995), Bandiera and Rasul (2006), and Conley and Udry (2010),[2] and in the adoption of new financial and health products and services (Banerjee et al. (2013), Cai, de Janvry and Sadoulet (2015), Kremer and Miguel (2007), Oster and Thornton (2012) and Miller and Mobarak (2015)).

This line of inquiry has particular policy relevance as it points to the possibility of leveraging networks to boost take-up of specific programs or behaviours, maximizing policy reach and impact at a given budget (e.g. Kim et al. (2015)). Recent empirical work further contributes to this policy relevance by identifying optimal information injection points and network structures conducive to diffusion (Banerjee et al. (2013, 2018b), Beaman et al. (2018), Beaman and Dillon (2018), BenYishay and Mobarak (2018), Alatas et al. (2016)).

In practice, however, leveraging networks to increase a program's impact requires targeting it to a subset of individuals within an interconnected group. This has the potential to modify the structure of these connections – in other words, to change the social network. This may happen if two conditions are met. The first is that there is a strategic element in network formation, and the network is not fully pre-determined (e.g. by kinship). And the second is that the program changes the characteristics of those targeted (relative to those not targeted) enough to affect network formation incentives. For example, it may be valuable to be connected to those with more financial means or special skills, the well-informed or those with connections beyond the group (politicians, NGOs, banks etc.). If the program improves the economic outcomes, skills, information set, or outside-connections of those receiving it, they may become more popular within the group. Such network re-wiring would typically be an unintended consequence and merits research attention.

This paper sheds light on this issue. While much of the previous (and above-mentioned) work has focused on the rural setting, I investigate networks between urban enterprise owners, contributing to an emerging literature that has studied the role of networks in diffusion of business practices (Fafchamps and Quinn (2018), Hardy and McCasland (2018), Fafchamps and Soderbom (2014)),

[1] See Jackson and Yariv (2011) and Mobius and Rosenblat (2014) for overviews.

[2] Recent methodological innovations include exogenous seeding of information through randomized experiments, combined with individual-level network information instead of general reference groups. The latter was already the case for Conley and Udry (2010). Examples are Carter, Laajaj and Yang (2014) and Magnan et al. (2015).

and their effect on firm performance (Cai and Szeidl (2018)). Horizontal networks between urban micro-enterprises are particularly interesting due to a specific tension inherent in the frequently observed geographic clustering of very similar firms selling near identical products. On the one hand, business interactions in this setting can be deeper and embedded in social interactions – as geographic, social, and sector proximity facilitates business collaboration and friendships. On the other hand, it may place firms in direct competition. Thus, the costs and benefits of linking may be different, and findings from the village setting not applicable.

In this paper, I use novel panel network data to investigate how the business collaboration and information sharing networks between the owners of small urban enterprises changed in response to a randomized training intervention. A randomly chosen subset of the members of a local small industry association in Kampala, Uganda, received a business and technical skills training that intended to advance their skills and practices and ultimately improve business performance. Networking ties in this group are tight: enterprise owners frequently interact with their linking partners, discussing techniques and market conditions (64% of pairs discuss business issues daily or weekly), and sharing employees or equipment and referring customers to each other (45%, 61% and 70%, respectively, have done so in the past year).

An exploration of link reporting at baseline, prompted by observed patterns of non-reciprocal nominations, suggests a systematic relationship with "status" differentials. Links with partners who are more knowledgeable and experienced than oneself, and have a larger and better equipped business, seem to be more beneficial. This descriptive finding feeds into a framework of bilateral link formation that conceptualizes business link benefits as a function of the relative status differential, while allowing for a competition cost from sharing exclusive knowledge with potential competitors. The training intervention can be interpreted as an exogenous manipulation of the pair-level status differential while also potentially increasing the competition cost to the treated.

Crucially for the analysis, I use panel network data, that is, I elicited network connections again after the intervention took place, in addition to the baseline connections that are typically available to the researcher. This panel network data allows me to trace link evolution at the enterprise-pair level. To identify causal effects on linking likelihoods, my empirical strategy relies on the randomness of the treatment assignments of the two sides in a potential link which is implied by the experiment, and the fact that (for a subset of links) these are orthogonal to each other. I thus conduct dyadic (pair-level) analysis.[3]

I find that the training changes linking patterns in this business network. It increases the popularity among untreated entrepreneurs of links with treated entrepreneurs, leading to a positive effect on their likelihood to report a link at endline (+32%). Given that link formation is conceptualized

[3] Dyadic analysis is indispensable, as finding an individual-level difference in the popularity ("in-degree") of treated and untreated could result from increased popularity of the treated or reduced popularity of the untreated.

as bilateral, which requires that both parties weakly benefit from link existence, why would the treated accept such links? The conceptual framework suggests that the effect has to be driven by a specific type of enterprise-owner pair: those in which the treated would have wanted a link even in the absence of the intervention, while the untreated would not have. These are pairs with an untreated owner of higher baseline status relative to the treated owner. I take this prediction to the data. I find that the effect is entirely driven by this type of pairs; relatively lower-status treated attract new connections with relatively higher-status untreated.

Furthermore, I find that baseline links within clusters of treated enterprises are strengthened, which is not due to a strategic replacement of untreated with treated partners but seems to be an effect of jointly attending the training. This may be through increased trust in each other's business practices and standards, or through a tightening of social ties. In addition, treated entrepreneurs do not seem to bear a competition cost from links with untreated entrepreneurs in this setting, as there is no evidence of strategic avoidance of such links.

In summary, I show that business networks can be endogenous to a public policy intervention. I find a positive effect on linking likelihoods, driven by untreated entrepreneurs to whom links with treated entrepreneurs become more desirable. Observed changes coincide with predictions from a bilateral link formation framework, and the main channel through which the intervention affects the network is by modifying the relative status differentials between entrepreneurs. Jointly attending the training strengthens ties, while potential competition costs play a minor role.

This paper contributes to, and expands on several strands of the literature. First, it is one of the very first studies to use survey panel network data to investigate network change, and is thus able to trace network evolution at the dyad level; it is the first, to my knowledge, to do so for small enterprise owners and in an urban setting. Comola and Prina (2018) focus on treatment and peer effect identification given network endogeneity, and provide an empirical illustration using panel network data and a randomized access-to-savings-accounts intervention in rural Nepal. Preliminary work by Fernando and Sharma (2016) investigates effects of a mobile-phone-based agricultural extension experiment in India on the popularity of treated farmers as sources of agricultural information. Banerjee et al. (2018a) find that the introduction of micro-finance in Indian villages induced a thinning of informal credit networks that spilled over to other dimensions of the social network. The authors' focus is the effect of the group-level introduction of a formal institution on pre-existing informal relationships within the group. I contribute a different angle, the consequences to the relationships within a group when only a subset receives an intervention.

Second, research into information diffusion and social learning in technology adoption has thus far focused primarily on social networks between households in a rural setting. However, incentives in the urban enterprise context may differ, and results may not be transferable. This paper joins a small set of very recent studies on urban business networks and contributes to an incipient debate on the role of competition and rival information (Cai and Szeidl (2018), Hardy and McCasland (2018)). While I do not observe indications of competitive behaviour in the patterns of network change induced by the training intervention studied, my results confirm that considering the incentives of both sides in a link is crucial in the firm setting (Hardy and McCasland (2018)). After early work on business networks had hinted at their important role in firm performance and economic development (Woodruff and MacMillan (1999), Barr (2000)), current work has introduced random variation in networks to pin down effects on diffusion of business practices (Fafchamps and Quinn (2018)) and on firm performance (Cai and Szeidl (2018)). To this emerging literature showing the causal effect of business networks on outcomes, I contribute evidence on the reverse causal effect of (perceived) outcomes on business networks.

Third, I contribute to a broader literature that highlights interactions between public policy and social networks. On the one hand, existing networks can affect program success. Barnhardt, Field and Pande (2017) ascribe low take-up and high drop-out rates for an access-to-housing program for slum dwellers in India to participants' perception and experience that moving is too costly in terms of loss of social ties. On the other hand, policy programs can affect networks, and in some cases this is their stated purpose, such as in the re-integration of ex-combatants of armed conflicts (UN (2014), Annan and Blattman (2016), Humphreys and Weinstein (2007)). Yet trying to engineer social interactions can backfire. An attempt by Carrell, Sacerdote and West (2013) to induce positive peer effects for low achieving college students had negative effects instead, because students self-sorted into homogenous groups within classes. To this body of work that highlights the complex interplay of public policy and group structure, I contribute further evidence that their networks matter to individuals, who will form and change them strategically (which may not always be expected by the policy maker), but who may also be nudged into collaborating through joint group assignments (Feigenberg, Field and Pande (2013)).

To sum up, this paper contributes to and innovates on several strands of the literature. It also harbours lessons for both public policy and research. First, in terms of policy implications, the findings presented suggest caution when designing interventions that build diffusion through networks into program design, as the network may rewire as an unintended consequence. This process can create new insiders and outsiders, with possibly negative equity implications – even though this does not seem to have been the case in the present context. Policy-induced network changes matter because networks matter for outcomes (e.g. Cai and Szeidl (2018) and Fafchamps and Quinn (2018) for enterprises).

Second, an implication for research into diffusion and social learning through networks is that relying on baseline (pre-intervention) networks may fail to capture the full effect of the intervention – any network rewiring it caused – and risks giving biased estimates of direct and indirect program effects.[4] For the case of a randomized intervention, the group we may consider as "pure" control group when only baseline data is available (untreated individuals whose baseline links also happen to be untreated) will in fact be "contaminated" if they have endogenously connected to treated individuals.

The remainder of this paper is organized as follows. Section 2 introduces the intervention and the data and Section 3 explores patterns of the baseline network. The conceptual framework and its predictions are discussed in Section 4, and the empirical strategy in Section 5. Section 6 presents my results, and Section 7 concludes.

2 Intervention and data

2.1 Intervention

The intervention that is at the heart of this paper is a business and technical skills training implemented in Kampala, Uganda, by Kassida, the Katwe Small Scale Industries Association for its members, with financial support from a World Bank project. These association members are owner-managers of small-scale manufacturing and service delivery firms, with three employees on average, that operate in nine very diverse sectors, from metal fabrication, foundry and forging, electronics, machining, and carpentry, to tailoring, shoe making, catering, and hair salons. All association members were eligible for the training. An impact evaluation of the training intervention was designed by Campos, Goldstein, Pimhidzai, Stein and Zia (2018). This paper uses this experiment to ask a different question: How do networks change as a consequence of a training intervention that targets only some members of an interconnected group.

The goal of the training was, firstly, to improve the technical skills and practices of business owners and workers – who both typically learned their trade through an informal apprenticeship rather than formal vocational training – and, secondly, to enhance the business owners' business skills and practices. The ultimate objective was to improve firm performance. The training was intended as a package of both technical and business trainings. Trainees received certificates of participation.

The technical training was sector specific and conducted at selected workshops of association members in groups of enterprise owners from the same sector and their employees. The training groups were formed based on practical considerations, that is, geographic proximity to the training workshop. The technical training included hands-on demonstrations and practice of production or service delivery techniques relevant to a given sector, to improve the quality of products, reduce

[4] This issue is a focus of Comola and Prina (2018) who propose an estimation strategy that accounts for network change.

resource wastage, and increase safety at work. Trainees were also introduced to new products and designs with increased value-added. This part of the training was delivered in the fall of 2011.

The business training was general (non-sector specific) and focused on rules of thumb in five main areas: separating household and business finances and protecting the business from one's family, market research and marketing, costing and pricing, business planning and record keeping, and saving for the business. It took place in 12 classroom-type sessions over the period of three months in late 2011 / early 2012 in mixed-sector groups of enterprise owners only. Groups were formed based on proximity and training venues were chosen nearby participants' business locations. The course was delivered by a certified instructor in a local training institute who had started out as a tailor in the association, before working his way to an MBA. Hence, he had a clear understanding of the baseline level of business management knowledge. Participants received a printed-out training manual and a do-it-yourself handout for each topic summarizing the key messages.

Firms in the association, and particularly those from the same sector, operate in very close proximity of each other, and are sometimes not separated by physical walls. For example, the business premises of most tailors do not comprise more than their sewing machine, and the next tailor is located less than a meter away. The same applies for women who prepare and sell food in a market where many caterers are bunched together. In other sectors, notably metal fabrication, machining, and hair salons, firms tend to have somewhat larger premises that are generally separated by walls, but they are still located right next to other firms of the same sector, who are often also association members.

Given this setting, to avoid spillovers from business owners literally looking over each other's shoulders, and instead be able to isolate intentional information sharing, Campos, Goldstein, Pimhidzai, Stein and Zia (2018) randomized the intervention not at the individual firm level, but at the level of same-sector clusters of firms.[5] Firms were assigned to within-sector clusters based on a perpetuated 20 meter rule using baseline GPS coordinates: Two firms are in the same cluster if they are in the same sector and within 20 meters of each other; the rule is applied until no other businesses of the same sector are within 20 meters of anyone already in the cluster. The 20-meter cut-off was chosen because the association considered it the minimal distance required to avoid imitation by direct observation between treatment and control businesses.

Clusters were then matched into pairs within sectors based on their observable characteristics, which included cluster size, characteristics of the businesses in the cluster and socio-economic characteristics of their owners. One cluster per pair was then randomly assigned to receive the training and the other to the control group. Figure A1 illustrates the outcome of this treatment assignment protocol.

[5]Cluster-level randomization was also more acceptable to the industry association, as they feared discontent if direct neighbors were assigned to different groups.

2.2 Data

Firms and firm owners

After a listing of all association members that applied for the training, the baseline survey was conducted on the universe of 733 firms in early 2011; they were in 273 clusters. In the endline survey in late 2013, 613 of the baseline firms were re-interviewed (83.7%); they were in 242 clusters.[6] From the firm-perspective, the average size of its cluster is 8.2 firms, and 26% of firms are in a cluster of their own (singletons) (Table 1). Baseline summary statistics on the 613 firms that are in the endline sample are presented in Table 1. Enterprise owners were on average 38 years old at baseline and had nine years of schooling. Fourty-five percent of enterprise owners in the sample are female.

The average business age was 12.6 years, and a bit over half of the firms were formally registered. Average firm size was three employees (25% had no employees and the largest firm in the sample had 45). The average revenues in the previous month were $2,330,000$ Ugandan shillings (approximately 975 USD in early 2011), and average monthly profits were $391,000$ Ugandan shillings (approximately 165 USD). The largest sectors in the sample are metal fabrication and tailoring, which together make up half the sample (27% and 24% respectively). Other larger sectors are catering (16%), hair salons (11%), foundry and forging (9%), and carpentry (7%). The smallest sectors are shoe making, electrical, and machining. Catering and hair salons are female dominated (98% and 92% female respectively), while the male-dominated sectors are machining, electrical, metal fabrication, shoe making and carpentry (100%, 99%, 97%, 97% and 93% male respectively), and tailoring and foundry and forging are mixed (67% and 26% female respectively).

Business links

At baseline, Campos, Goldstein, Pimhidzai, Stein and Zia (2018) asked all enterprise owners to provide the names and contact information of other enterprise owners they talk to about business issues or are engaged with in business collaboration. The question allowed for the nomination of any enterprise owner, from within or outside the study sample. On average, respondents gave 3 names (Table 1, "out-degree (full)"),[7] 1.24 of whom are from the main sample themselves ("out-degree (within)"). The average baseline "in-degree" is 1.29, which means respondents were nominated by on average that many others.[8]

[6] The likelihood to attrite from the study is not affected by treatment assignment.

[7] The question permitted a maximum of five names. The observed proportions of the possible values of the out-degree (0, 1, 2, 3, 4, 5) are, respectively, 1%, 19.4%, 24.8%, 18.4%, 10.4% and 25.9%. Implications of this survey-censoring are discussed below.

[8] Baseline summary statistics are for enterprise owners who are present in the endline sample. If we were looking at all baseline enterprise owners, the average baseline out-degree (within main sample) and the average baseline

This yields a sample of 1,813 directional baseline links, of which 762 links are with other respondents from the main sample. Of the latter, 446 (i.e. 58.5%) are within the same cluster and 316 are with firms from a different cluster (Table 2). Here and throughout, the term cluster refers to the within-sector clusters that are the unit of randomization (see Section 2.1). Reported links are directional in that one owner can nominate another without in turn being nominated by them. Occasionally, I will refer to the individual who reported a link as the "source", or "source node", and to the individual nominated as the "target" or "target node". We observe that 31.8% of the 762 directional baseline links are reciprocated (Table 2).[9]

For the purposes of this study, at endline I asked the same, open-ended, question as at baseline eliciting the names and contact information of other enterprise owners the respondent talks to about business issues or is engaged with in business collaboration. Respondents gave on average 3.19 endline contacts, of which 1.25 are from the main sample (Table 4, Columns 2 and 3, "Overall avg."). Fourty-five percent of the 1813 full-sample baseline links are mentioned again at endline, while 55.1% of the 762 within-main-sample baseline links are (Table 5, Columns 2 and 3, "Overall avg.").

The question of interest regards links at endline; the sample is 448,716 potential endline links (whether reported at baseline or not) within the main sample. This number is implied by 613 main respondents interviewed at endline, times 732, the total number of possible directional target nodes among 733 main sample respondents. Whether or not a specific one among the 536,556 potential baseline links (733 times 732) is in the endline sample of analysis depends only on whether the source node is in the endline sample, not on the target node. This is implied by the directional setup of the analysis, which uses endline information from the source node (their report on whether or not the link exists), but only baseline information from the target node (in particular treatment assignment).

In some of the analysis, I will separate links already reported at baseline (1813, of which 762 are from the main sample) and links newly reported at endline; the latter uses the sample of 447,954 potential newly reported endline links within the main sample.

in-degree would be equal by construction. Also note that 1.29 is the baseline in-degree from nominations by any baseline respondent; the baseline in-degree from nominations by respondents who do not themselves attrite is 1.11 (see Table 1), which we will use in Section 6.1 when discussing degree evolution.

[9]Framing this differently, we observe links between 641 distinct pairs of firm owners; for 121 of these pairs (or 20.4%) both sides report a link, while for 520 only one of the sides mentioned the other. Reciprocation is investigated in section 3.

3 Features of the baseline network

3.1 Baseline network characteristics

The baseline network is visualized in Figure 1 and baseline link characteristics are summarized in Table 2. I start by discussing links within the main sample. Source business owners have known the target for on average 10.7 years and are located on average 190 meters apart. For 89% of links, both partners have the same gender; this strong sorting by gender is at least partly driven by the fact that the large majority of links within the main sample is within the same line of business (93%, also see Figure 1) and there is strong sorting by gender in occupation choice. For 62% of links, the two sides are friends, for 17% business relations, for 16% family, and for 5% neighbors or other. These numbers are very similar in the full sample (i.e. the sample of links within and outside the main sample), except for the share of links that are within the same line of business, which drops to 76%;[10] the distance is unavailable for outside-sample links.

Information exchange on business issues between networking partners is frequent and business collaboration is intense and multi-dimensional. Sixty-four percent of pairs talk daily or weekly about business issues. Business collaboration involves sharing of employees, of materials and of equipment (45%, 53%, and 61% respectively have done so in the previous year). It extends to sharing of supplier and customer names (72% and 51% respectively), and may on occasion involve applying for larger contracts or tenders together (8%). When they are unable to fulfill an order themselves, entrepreneurs will refer the customer to their networking partner (reported outgoing for 65% of pairs and incoming for 63% of pairs). Financial interactions include direct lending to and borrowing from each other (33% and 34% respectively), in some cases entrepreneurs participating in the same savings or micro-finance group, and in rare cases jointly applying for external funds (8%, 8% and 4%) (see Table 2). No collaboration on any of these dimensions is reported for only 3.2% of baseline links. Given these patterns of collaboration, and the fact that links are largely within the same sector of business (93%), we can conclude that the business links in this sample are horizontal links rather than supply-chain links.

In the analysis, I will separate same-cluster links (source and target are from the same cluster) from different-cluster links. Figure 3 graphs the dimensions of business collaboration by these two groups, also breaking different-cluster links down by whether source and target are in the same sector. The average number of dimensions (out of nine) the pair collaborates on is smaller by about 1 for different-cluster links from the same sector compared to same-cluster links, but the nature of the business collaboration remains equally varied. For example, almost 40% for different-cluster pairs from the same sector share workers (understandably, it is lower for different-cluster links from

[10] This is expected, as all links with entrepreneurs in a sector of business that is not eligible for the training (e.g. retail) will be links outside the main sample.

different sectors), and the share referring customers to each other is close to 70%. The share who frequently exchange about business issues (daily or weekly) remains over 60% for different-cluster links from the same sector, and is still at 50% for those from different sectors. A similar break-down is shown in Appendix 3, Figure A2, by sector.

3.2 Reciprocation and relative status

As mentioned above, 31.8% of the 762 directional baseline links within the main sample are reciprocally reported by the target (or, put another way, for 20.4% of pairs with any reported directional link, the link is bi-directional). Yet, discussing business issues and business collaboration are inherently two-sided activities that both sides of the link should be aware of. Low congruence in link reporting thus raises a more general question about how to interpret the links observed in the data; finding an answer will be useful to support the conceptualization of linking behaviors and guide the interpretation of any empirical findings. In this sub-section, I show that links reported by only one side are likely business links that truly exist, and that non-reciprocation is due to self-censoring, with respondents reporting the links they care about most. In the next sub-section, 3.3, I will expand the analysis to link reporting in general.

Note first that observing reciprocation rates of similar magnitude as mine, for networks of two-sided activities, is common in the literature. To mention some examples, Leider et al. (2009) find reciprocation rates for friendship networks between college students in the US that are very similar to the numbers here – 36.6% of reported links are also reported by the other side, or differently put, 22.4% of undirectional links are symmetric links (own computations); the authors have a total number of source nodes similar to mine but a much smaller number of potential target nodes.[11] In the US National Longitudinal Study of Adolescent to Adult Health data from the mid-1990's ("AddHealth"), which is widely used in the literature as it includes friendship network data for 84 distinct networks of high school students, reciprocation rates are between 30% and 50% (Ball and Newman (2003)); even for romantic relationships the reciprocation rate does not reach 50% (Carver, Joyner and Udry (2003)).

Similarly, Banerjee et al. (2012) mention a low reciprocation rate for the reporting of kinship ties in 75 Indian villages. To cite an Eastern African example, for the risk-sharing network among households in a Tanzanian village (the data is used, among others, by De Weerdt and Dercon (2006) and De Weerdt and Fafchamps (2011)), the reciprocation rate is 44.4% of directional links,

[11] The reciprocation rate one can expect to observe empirically depends not only on the characteristics of the underlying network but also the link elicitation protocol; it is, for example, increasing in the number of links elicited per respondent and decreasing in the number of "permitted" target nodes. In my case, potential linking partners include any business owner known to respondents - in Kampala or, in principle, beyond. In Leider et al. (2009), respondents were asked to choose only from within the main sample of respondents; the authors furthermore incentivized concurrent reporting in their link elicitation.

or alternatively, reporting is concurrent for 28.6% of pairs with any reported link (own computations based on Comola and Fafchamps, 2014, 2017).

However, even if incomplete congruence in link reporting is a common feature of network data, it requires further investigation here, as understanding link reporting will be crucial going forward. Incomplete congruence could be either due to incomplete reporting, so that links reported by only one party are indeed existing business links, but the other party failed to report them for reasons to be investigated. Or, it could be an indication that a number of reported links are either non-existing and merely aspirational, or existing but social in nature.[12,13]

First note that for 81% of reported within-main-sample baseline connections the source provided a phone number, which was a correct phone number in at least 94% of cases.[14] As respondents thus largely know the phone numbers of those they nominated as connections, we can conclude that these are not made-up and purely aspirational but indeed existing links.[15] These links are, furthermore, not purely social in nature. While 62.1% of reported connections are indeed with someone the respondent characterizes as a friend, there is significant collaboration in terms of business activities: Respondents report working together in two or more of five dimensions of production collaboration for 69.8% of links, and in two or more of four dimensions of selling and customer acquisition collaboration for 62.0% of links; they report some financial collaboration for 46.8% of links (see Table 2).[16] No collaboration on any of these dimensions is reported for only 3.2% of baseline links. Respondents also report frequent information exchange on business issues, reporting to have received or provided business information at least daily or weekly from/to 64% of partners.

This leads to the conclusion that we can trust that reported links indeed describe existing business links, and observed low congruence is due to incomplete reporting. This could be due to random recall error. Alternatively, respondents' lists of networking partners may be censored, either by

[12] Data quality could be another explanation, in principle; if respondents gave their links with an unusual nickname and very imprecise contact information, insufficient to allow me to match them to main sample respondents, they would appear as outside sample contacts rather than (possibly reciprocated) within-main-sample contacts. However, great care was invested in the matching, including scrupulous investigation in the field with the help of enumerators, in order to minimize this possibility.

[13] Comola and Fafchamps (2014) use the pattern of concordant and discordant reporting of links from the two sides of the pair to infer whether reported links are existing or simply desired links, and if existing whether they are the product of a uni-lateral or bi-lateral link formation process. Their approach is not applicable here as the ability to distinguish existing from desired links crucially relies on the assumption that mis-reporting of an existing link is random, i.e. unrelated to source characteristics and relative characteristics between source and target. Distinguishing unilateral and bilateral link formation also relies on the (technical) assumption that under-and over-reporting are equally likely which implies that some reported links do not actually exist, which is implausible given the high levels of communication and business collaboration reported (see below).

[14] The share of correct numbers is a lower bound because it is possible that the source knows a phone number of the target that the latter did not reveal when interviewed.

[15] It is also not the case that some respondents gave some false links because they thought they had to exhaust the maximum allowed number of five contacts: for the 4th and 5th ranked in the order of reporting, the rate of phone number reporting is still over 70%.

[16] The fourteen dimensions of business collaboration elicited in the survey are grouped as follows. In terms of production: sharing of employees, of materials, of equipment, of supplier contacts, and buying materials together in bulk; in terms of selling and customer acquisition: referring customers to each other (asked directionally), sharing customer contacts, applying for contracts together; and in terms of finances: borrowing and lending (asked directionally), applying for loans together, being in a savings or micro-lending group together.

the survey question – which did not allow to name more than five partners, or self-censored if respondents only reported the links that are most salient to them, even without exhausting the limit of five. These hypotheses can be tested in the data. First, if censoring by the survey question is the reason for incomplete reporting, reciprocation rates should be lower for links reported with target nodes who exhausted their list of five (and whose ability to reciprocate is thus potentially censored). Instead, reciprocation rates for these links are slightly higher, 33.7% compared to 30.8% for links with target nodes who did not exhaust their list of five. We can thus dismiss the survey-censoring hypothesis.

Second, to distinguish between random recall error and self-censoring, it is useful to think about what factors would drive self-censoring. Not all of her business links will be equally valuable to an entrepreneur. Cai and Szeidl (2018), who experimentally identify positive effects of owner-manager's networks on firm outcomes, find causal evidence that these effects are larger when the network partner's firm is larger (where firm size is meant to proxy for the various dimensions of network partner "quality"). In the present context, it is indeed plausible that business links with certain individuals are more valuable, in particular, links with those with relatively more experience, knowledge and information, and with relatively higher-quality equipment, better trained employees, et cetera. In short, links with entrepreneurs of relatively higher "status" are more valuable. I thus hypothesize that individuals self-censor and only report those among their existing business links that matter most to them (regardless of whether that means exhausting the list of five permitted contacts).[17] We should then see that one-directionally reported links typically have a target node with higher status than the source node.[18]

To test this hypothesis against that of random recall error, I construct a relative status score at the dyad-level using principal component analysis, as follows. I compute the differences between each source node and each potential target node within the main sample in terms of each of nine elements of status in the dimensions business size and success, and owner knowledge and experience.[19] The differences are converted into z-scores, standardizing within sector for dyads from the same sector of business (as there are sector-level idiosyncrasies in elements such as typical firm size), and standardizing overall for dyads across sectors. I aggregate this information into a single relative status score computed as the linear combination of these z-scores using as weights the scoring coefficients of the first principal component. See Appendix 1 for details on the principal component analysis conducted, the distribution of the relative status score, and robustness checks.

A negative relative status score means that the source node has lower status than the target node.

[17]This is particularly likely as survey question specifically prompted respondents to mention contacts with whom they interact the most.

[18]Using a different approach to mine, Ball and Newman (2013) find, for the AddHealth data, that non-reciprocation in the friendship networks is a correlate of relative status, with the higher-ranked individual failing to report the existence of a friendship link.

[19]These are the baseline values of: number of employees, past month revenues, past month profits, number of customers per month, total assets, schooling, financial literacy, technical knowledge, owner age and business age.

By construction, the distribution of the relative status score in the sub-sample of reciprocated links is symmetric and centered on zero; this is because for these links, the relative status score from the point of view of one node is the inverse of that from the point of view of the other. If non-reciprocation is random, the relative status score for the sub-sample of one-directional links will be centered on zero as well. Comparing the distributions of the relative status score for one-directional links and reciprocated links, we see instead that the former is shifted to the left (see Figure 4). In the same vein, the average relative status score for one-directional links is significantly lower than that for reciprocated links (p-value: 0.0001);[20] in other words, on average the target node of a one-directional link has a higher status than the source node. This suggests that incomplete reporting of existing links is not due to random recall error, but instead self-censoring:[21] Respondents report those among their existing business links that they care about most; these are links with entrepreneurs of relatively higher status.

3.3 Link reporting and relative status

The pattern described in the previous sub-section is not specific to the issue of (non-)reciprocation but a general feature of baseline link reporting. I use standard dyadic regression analysis to show the correlation between baseline link reporting likelihood and status, in three closely related alternative specifications. The regressors are, for specification 1) the source and target status score, for 2) the relative status score discussed above, and for 3) a relative status indicator derived from the latter. The specification for 1) is $g_{ij} = \alpha + \theta_1 x_i + \theta_2 x_j + s_i + s_j + u_{ij}$, with source and target status entering as node-level regressors x_i and x_j, and for 2) and 3) it is $g_{ij} = \alpha + \zeta_1 z_{ij} + s_i + s_j + u_{ij}$, with the relative status score and relative status indicator respectively entering as dyad-level regressor z_{ij}. The dependent variable g_{ij} is equal to 1 if source i reports a baseline link with target j and 0 otherwise. Sector fixed effects s_i and s_j are included as the correlation between linking and status may depend on sector-specific linking patterns.[22] While the analysis in sub-section 3.3 used the sample of all links reported at baseline (to investigate correlates of non-reciprocal reporting given reporting by at least one side), here I use the sample of all potential links (to investigate which

[20]This test is conducted using dyadic standard errors as proposed by Fafchamps and Gubert (2006), which correct for potential correlations between observations that involve the same individual, whether in the role of source or target node. See Footnote 33 for more detail on dyadic standard errors in the context of this study. Here, the approach also corrects for the fact that for reciprocated links, a given pair of nodes is "observed" twice (and with mirrored status scores), while for one-directional links, the pair of nodes appears only once in the data.

[21]I perform another test, which again refutes random recall error: Under random mis-reporting of existing links, a link that is one-directional at baseline is as likely to be one-directional in the same direction at endline as it is in the opposite direction. Instead, of the 520 unreciprocated baseline links for which I have endline information from the source, 37.7% remain one-directional in the same direction, and only 4.0% become one-directional in the other direction (for 10.6%, I do not have endline information from the target, that is, reciprocation information is missing; 11.3% become reciprocated, 36.3% are not reported by either side).

[22]I use dyadic standard errors (see Footnote 33). The relative status score used in 2) is discussed above. The individual status scores in 1) are constructed in an analogous fashion. The relative status indicator used in 3) takes value one if the relative status score used in 2) is positive, and value zero otherwise. Due to the continuous nature of the relative status score, there are no dyadic observations with a score of precisely zero. See Appendix 1 for more detail on the construction of the scores and indicator.

ones among the full set of potential links within the sample were reported).[23]

I find that potential baseline links pointing towards a target with a higher status score are significantly more likely to be indeed reported, while higher status source nodes did not report significantly more or fewer connections than lower status source nodes (Table 3, Column 1); this suggests that source nodes, including those who are themselves of higher absolute status, tend to mention target nodes of higher status relative to them. This pattern is picked up both by the relative status score and the relative status indicator which both show highly significant point estimates (Table 3, Columns 2 and 3). Consistently, Figure 5 shows that the distribution of the relative status score for reported baseline links is skewed to the left compared to the (symmetric) distribution for all potential links. The fact that the source node status is not significant in Table 3, Column 1, clarifies that results for the relative score and indicator are not driven by higher status entrepreneurs reporting fewer links in total.[24] These results are again consistent with the interpretation that not all existing links are reported and that respondents self-censor, reporting the links they care about most.[25]

To summarize, the baseline data patterns analyzed in this sub-section and sub-section 3.2 suggest that relatively higher status entrepreneurs are more attractive as business network partners, and that links reported one-directionally (typically by the relatively lower status node in the pair) are truly existing business links, that are not important enough to the other node (typically the relatively higher status node) to report. In other words, higher absolute status entrepreneurs have more business links in total, but not all of them are equally salient to them.

The evidence from baseline linking patterns should be understood in a descriptive sense only, as the network at baseline is endogenously formed and an entrepreneur's observed popularity among her peers (hence observed network size) is jointly determined with the factors of status that make her popular: A successful entrepreneur may be attractive as networking partner to others, but at the same time her networks may be at the root of her success. But it will, in the next section, serve to inform the conceptual framework. I will derive predictions on the effect of an exogenously induced variation in relative status (from the training intervention), that I will then take to the empirical test.

[23]The number of all links reported at baseline is 914. The number of all potential links within the baseline sample is 536, 556 ($733 * 732$). These numbers are somewhat higher than what is used in the endline analysis (762 baseline links and 447, 954 potential links) due to attrition.

[24]As a consequence of these linking patterns, the average target node status in the sample of baseline links is positive, the average relative status score is negative, and 55% of baseline links have a source node with lower status than the target node (Table 2). If there was no relative-status pattern to baseline linking, the former two would be zero and for 50% of baseline links the source node would have a lower status than the target node.

[25]Results remain very similar, with identical significance levels, when omitting the sector fixed effects (unreported). Additional support to this interpretation is lent by the fact that in the order of reporting, the first link mentioned has on average the smallest relative status score (see Appendix 3, Figure A3), that is, the link for which the target node has the highest status relative to the given source node.

4 Conceptual framework

4.1 Baseline

I start by specifying individual i's utility from the (potential) link g_{ij} with a given individual j:

$$U_i(g_{ij}) = b_{ij} + ob_{ij} - c_{ij} - oc_{ij} \tag{1}$$

Individual i can derive positive utility from a link from two sources: the benefits of business collaboration b_{ij} (receiving useful information from j, ability to borrow j's employees etc.), and other benefits ob_{ij}, such as the enjoyment derived from working with a friend or family member. Forming and maintaining the link comes at a cost to i, c_{ij}, the cost of helping a potential competitor (through sharing valuable information, sharing trained employees etc.), and oc_{ij}, other costs which can vary at the individual level (for example through differential opportunity costs) and at the dyad level (likely increasing for businesses that are farther apart geographically, in terms of their business activities, or socially).

The value of business collaboration with j, b_{ij}, is a function of the knowledge, experience, and business size differential between the two nodes, or in short, the status differential $(s_i - s_j)$: the more j is experienced, knowledgeable, well equipped etc. compared to i, the larger b_{ij}.[26] Thus: $b_{ij} = f(s_i - s_j),\ with\ \frac{\delta b_{ij}}{\delta(s_i - s_j)} < 0$.[27]

Similarly, individual j's utility from establishing (or maintaining) the same link g_{ij} with individual i is given by:

$$U_j(g_{ij}) = b_{ji} + ob_{ji} - c_{ji} - oc_{ji}$$

Note that the link is denoted g_{ij} in both cases - the link between i and j either exists or it does not, but it does not have a "direction". On the other hand, the benefits from receiving information and engaging in business collaboration, and the costs of maintaining the link, can differ between i and j and are thus respectively denoted b_{ij} and b_{ji}, et cetera.

I assume that link formation is bilateral. Pairwise stability (Jackson and Wolinsky (1996)) then requires that a link will only be formed when both nodes weakly benefit compared to a situation where the link is not formed, at least one of them strictly; an existing link will only be maintained when both benefit weakly. Simplifying, link formation here requires $U_i(g_{ij}) \geqslant 0$ and $U_j(g_{ij}) \geqslant 0$. Modeling a bilateral link formation process rather than a unilateral link formation process seems

[26]It is plausible that there is a limit to the usefulness of collaborating with a (much) higher status individual; this idea can be captured through the cost oc_{ij} which may increase with various measures of distance, and may thus be prohibitively high for businesses that are too different.

[27]One could furthermore allow for decay in the value of some of the elements of j's status to i - in particular j's knowledge - from one period to the next to incorporate the idea that the value to i of the link depreciates as more and more of j's knowledge is transferred.

appropriate, as it is implausible that one entrepreneur can initiate a business collaboration without the other entrepreneur's consent. Forming a new link then is a bilateral decision while breaking an existing link can be unilateral and will happen as soon as one of the nodes derives negative utility from the link.

However, I assume that while business link formation is bilateral, information sharing, as well as at least some of the dimensions of business collaboration, can be initiated unilaterally among linked pairs, that is, once a link exists, neither party can refuse to share information in their possession with the other. The only way to stop sharing (and stop bearing the competition cost c_{ij}) is to break the link.

In this framework, some links that exist at a given moment will be dropped over time because either $U_i(g_{ij}) < 0$ or $U_j(g_{ij}) < 0$, due to depreciation of the value of j's knowledge to i (or of i's knowledge to j), or shocks to or changes in any of the elements of costs and benefits: changes in s_i or s_j (one node acquires new knowledge, new capital, experiences business growth, etc.), and changes in ob_{ij} (e.g. the two nodes fall out socially), c_{ij} (e.g. i acquires valuable information that will be costly to share in a competitive environment), or oc_{ij} (e.g. one node moves farther away such that the cost of maintaining the link goes up). Similarly, such changes and shocks will induce the formation of new links. In other words, the network will be subject to *endogenous* change over time even in the absence of any policy interventions.[28]

4.2 Effect of the intervention on individual utility

In this framework, the business and technical training intervention may introduce an *exogenous* shock to the elements of costs and benefits. In particular, it may change the benefits of linking b_{ij} and b_{ji}, by changing the status differential between the two nodes of a pair, and/or changing the competition costs of sharing, c_{ij} and c_{ji}. The training may also have a specific effect when both nodes happen to be in the treatment group, through ob_{ij} and oc_{ij}, as discussed below. I adapt the potential outcomes framework to the network context to illustrate the counterfactual utilities from linking:

$$U_i(g_{ij}) = (1 - T_i)(1 - T_j)U_i^{00}(g_{ij}) + T_iT_jU_i^{11}(g_{ij}) + T_i(1 - T_j)U_i^{10}(g_{ij}) + (1 - T_i)T_jU_i^{01}(g_{ij}) \quad (2)$$

where $U_i(g_{ij})$ is the utility to i that will be realized after the intervention if i maintains or creates a link with j; it will be equal to one of the four potential outcomes depending on the binary treatment assignment of both nodes, T_i and T_j. U_i^{00} indicates the potential utility to i from a business link with j when both are in the control group, U_i^{11} that when both are in the treatment group; U_i^{10}

[28]This "background" level of change is endogenous as e.g. business growth may be the result of having good quality network partners (see Cai and Szeidl (2018)) but also cause network changes through s_i.

and U_i^{01}represent the potential utilities to i when respectively only i or j are treated. Analogous superscripts indicate the b_{ij}, ob_{ij}, c_{ij}, and oc_{ij} corresponding to each of the potential utilities.

When both nodes are treated, their status differential remains unchanged compared to the case where neither is, so $b_{ij}^{00} = b_{ij}^{11}$, henceforth denoted b_{ij}^{0}. When only one node is treated, the status differential shrinks when the treated node was of relatively lower status and increases when the treated node was of relatively higher status; the benefit to the untreated node increases and that to the treated node decreases. That is, for i: $b_{ij}^{10} = (s_i^1 - s_j^0) < b_{ij}^0$ when i is treated and $b_{ij}^{01} = (s_i^0 - s_j^1) > b_{ij}^0$ when j is treated, where s_i^1 and s_i^0 denote i's status in the state of the world where i is treated and untreated respectively, and likewise for j.

The competition cost to i when i is treated and j is not is c_{ij}^{10}, which captures the cost of sharing exclusive knowledge with a potential competitor. The competition cost to i in i's untreated state and when both are treated is c_{ij}^0, i.e. I set $c_{ij}^{00} = c_{ij}^{01} = c_{ij}^{11}$, henceforth denoted c_{ij}^0, where the first equality is because i does not have any additional exclusive knowledge, and the second equality is because it is not costly in a competition sense to share knowledge the other person already knows. We have $c_{ij}^{10} \geqslant c_{ij}^0$.

I assume that the treatment can only affect ob_{ij} and oc_{ij} when both nodes are treated, which means, for example, that the training does not induce participants to move away. So $oc_{ij}^{00} = oc_{ij}^{10} = oc_{ij}^{01}$ and $ob_{ij}^{00} = ob_{ij}^{10} = ob_{ij}^{01}$, henceforth denoted oc_{ij}^0 and ob_{ij}^0 respectively. The training may increase the benefits from collaboration that are not related to the status differential, ob_{ij}, for pairs where both are treated. This could be if there is a benefit from continuing to discuss the issues discussed in the training with fellow trainees, or if going to the training together increased their social interactions[29] such that they now draw utility from collaborating beyond the pure business-related benefit; $ob_{ij}^{11} \geqslant ob_{ij}^0$. Alternatively, the training may reduce the non-competition related costs of linking, oc_{ij}, e.g. if it builds trust that the business practices and standards followed by the partner are similar to own practices and standards; $oc_{ij}^{11} \leqslant oc_{ij}^0$.

4.3 Effect of the intervention on linking likelihoods

Pulling these elements together, I now discuss the effect of the intervention on the likelihood that a given potential link is formed (or baseline link is maintained), under bilateral link formation. This requires distinguishing between the possible treatment assignments of the pair, formed by the possible combinations of the binary T_i and T_j: neither of the nodes is treated, only one is (where from i's point of view it matters whether it is i or j), or both are.

[29]Feigenberg, Field and Pande (2013) show that frequent micro-finance group meetings induced increased social interaction in the long run, which goes together with increased collaboration, in their case an increased willingness to pool risk.

Treatment assignment $T_i = T_j = 0$

From the elements of i's utility discussed above, $U_i(g_{ij}) = U_i^{00}(g_{ij}) = b_{ij}^0 + ob_{ij}^0 - c_{ij}^0 - oc_{ij}^0$, and symmetrically for $U_j(g_{ij})$. In the absence of externalities, and in particular if i's choice to link with j does not negatively impact the likelihood that i will also link with another node k, utilities in this group and the resulting linking patterns represent the utilities and the linking patterns that would have been realized for everyone in the absence of the intervention, i.e. if everyone is in the potential untreated state. $U_i^{00} = U_i^N$, where U_i^N denotes the overall counterfactual "No intervention". All links that would have been formed in the absence of the intervention are formed, and all links that would not have been formed in the absence of the intervention are not formed. The pair treatment assignment $T_i = T_j = 0$ then is the comparison group.

Treatment assignment $T_i = 0$, $T_j = 1$

In this case, $b_{ij} = b_{ij}^{01}$ and $c_{ij} = c_{ij}^0$; $b_{ji} = b_{ji}^{10}$ and $c_{ji} = c_{ji}^{10}$, while ob and oc are not affected (see above). Thus $U_i(g_{ij}) = U_i^{01}(g_{ij}) = b_{ij}^{01} + ob_{ij}^0 - c_{ij}^0 - oc_{ij}^0$ and $U_j(g_{ij}) = U_j^{10}(g_{ij}) = b_{ji}^{10} + ob_{ji}^0 - c_{ji}^{10} - oc_{ji}^0$. To understand the effect of this pair-level treatment assignment on the linking likelihood, it will be useful to distinguish four pair types. These are defined by whether the potential utilities of i and j in the absence of the intervention, $U_i^N(g_{ij})$ and $U_j^N(g_{ij})$, are positive or negative, that is, by whether i and/or j would have wanted to form the link in the absence of the intervention. Note that this definition is at the pair level; among each node's set of potential links, there will be some of each of the pair types - some links neither side would have wanted, some that only i or only j would have wanted, and some that both would have wanted.

Pair type A: $U_i^N(g_{ij}) < 0$ and $U_j^N(g_{ij}) < 0$; in the absence of the intervention, neither i or j want the link, and it does not get formed; if it exists it will be dropped. With the intervention, as $c_{ji}^{10} \geqslant c_{ji}^0$ and $b_{ji}^{10} \leqslant b_{ji}^0$, $U_j(g_{ij}) = U_j^{10}(g_{ij}) \leqslant U_j^N(g_{ij}) < 0$. As j still does not want this link, under bilateral link formation it will not be formed, regardless of how i's utility is affected. Thus, for pair type A, the intervention has no effect on the linking likelihood, $Pr(g_{ij}^A = 1) = Pr^N(g_{ij}^A = 1) = 0$, where g_{ij}^A stands for a link of pair type A, and N for the overall counterfactual of the intervention not taking place, as before.

Pair type B: $U_i^N(g_{ij}) \geqslant 0$ and $U_j^N(g_{ij}) < 0$; in the absence of the intervention, i wants the link but j does not, and the link does not get formed. By the same reasoning as for pair type A, with the intervention, $U_j(g_{ij}) = U_j^{10}(g_{ij}) \leqslant U_j^N(g_{ij}) < 0$ and the link will not be formed because j does not want it; $Pr(g_{ij}^B = 1) = Pr^N(g_{ij}^B = 1) = 0$.

Pair type C: $U_i^N(g_{ij}) < 0$ and $U_j^N(g_{ij}) \geqslant 0$; in the absence of the intervention, j wants the link but i does not, and the link does not get formed. From i's point of view, as $b_{ij}^{01} \geqslant b_{ij}^0$ while

c_{ij} is not affected, we have $U_i(g_{ij}) = U_i^{01}(g_{ij}) \geqslant U_i^N(g_{ij})$. From j's point of view, as $c_{ji}^{10} \geqslant c_{ji}^0$ and $b_{ji}^{10} \leqslant b_{ji}^0$, we have $U_j(g_{ij}) = U_j^{10}(g_{ij}) \leqslant U_j^N(g_{ij})$. If for at least some of the i for which $U_i^{01}(g_{ij}) > 0$, the increase in costs and reduction in benefits for the associated j is small enough such that $U_j^{10}(g_{ij}) \geqslant 0$, then the intervention will increase the likelihood that a link of this pair type is formed. It can never reduce it as $\Pr^N(g_{ij}^C = 1) = 0$. Thus $\Pr(g_{ij}^C = 1) \geqslant \Pr^N(g_{ij}^C = 1)$.

Pair type D: $U_i^N(g_{ij}) \geqslant 0$ and $U_j^N(g_{ij}) \geqslant 0$, with at least one strict inequality in the case of new links; this is the only pair type that will form a link in the absence of the intervention because, as required by bilateral link formation, both want the link; $\Pr^N(g_{ij}^D = 1) = 1$. As $b_{ij}^{01} \geqslant b_{ij}^0$ while c_{ij} is not affected, $U_i(g_{ij}) = U_i^{01}(g_{ij}) \geqslant U_i^N(g_{ij}) \geqslant 0$ and thus i will continue to want the link in the presence of the intervention, but as $c_{ji}^{10} \geqslant c_{ji}^0$ and $b_{ji}^{10} \leqslant b_{ji}^0$, we get $U_j(g_{ij}) = U_j^{10}(g_{ij}) \leqslant U_j^N(g_{ij})$, and j may not any more. Thus $\Pr(g_{ij}^D = 1) \leqslant \Pr^N(g_{ij}^D = 1)$ depending on whether the competition cost of providing training-related information is high enough to push some j into not wanting the link.

To summarize, the conceptual framework suggests that for the case that one node in the pair is treated and the other is not, the intervention could have a positive or a negative effect on the endline linking likelihood. Any positive effect will come from pairs where in the absence of the intervention the link would not be formed (or maintained) because the untreated node would not want it although the treated node would (pair type C). These are pairs where the relatively higher baseline status node is untreated. A link will be formed for pair type C if the intervention increases the relative status (and thus desirability as networking partner) of the treated node to the point where the untreated node wants the link, and if the utility to the treated note (although potentially lowered due to an intervention-induced reduction in benefits and increase in costs) remains weakly positive.

Any negative effect on the linking likelihood of pairs with a treated and an untreated node has to come from pairs that would form (or maintain) a link in the absence of the intervention because both nodes would want it (pair type D). These are pairs where both nodes' desire to link is driven by high other benefits and/or low costs rather than by status differentials, as the benefit related to relative status is by definition asymmetric. The intervention may cause some of these links not to be formed if increased competition costs make links with untreated nodes less desirable to treated nodes: some treated nodes may decide not to form the link to keep the new knowledge (and, e.g., trained employees) to themselves. This is because once linked, neither partner can refuse to share information in their possession or collaborate in certain business activities.

Treatment assignment $T_i = 1$, $T_j = 0$

While from the point of an individual node it matters whether they are the treated or the untreated node of the pair, it is not necessary to discuss treatment assignment case $T_i = 1$, $T_j = 0$ here, as it is the same as the previous case with the roles of i and j reversed.

Treatment assignment $T_i = T_j = 1$

In this case, $b_{ij} = b^0_{ij}$ and $c_{ij} = c^0_{ij}$, but $ob^{11}_{ij} \geqslant ob^0_{ij}$ and $oc^{11}_{ij} \leqslant oc^0_{ij}$. Thus $U_i(g_{ij}) = U^{11}_i(g_{ij}) = b^0_{ij} + ob^{11}_{ij} - c^0_{ij} - oc^{11}_{ij}$, and symmetrically for $U_j(g_{ij})$. Therefore, the intervention cannot have a negative effect on linking likelihood for any of the pair types. For pair type D, the intervention then has no effect, a link will be formed (or maintained) with or without the intervention. Pair types B and C are identical because of the symmetry of this treatment assignment. For pair types A and B/C, the intervention has a weakly positive effect on the linking likelihood, depending on whether the utility from linking of the node(s) unwilling to link in the absence of the intervention is pushed above zero by the reduction in other linking costs oc or increase in other linking benefits ob.

To summarize, the framework predicts that for pairs where both nodes are treated, the intervention may have a positive effect on the linking likelihood, if both attending the training, possibly together, increases the benefits of being linked and/or reduces the costs. The effect would be driven by types of pairs where either one or both nodes would not have wanted the link in the absence of the intervention.

4.4 Reporting of links

The conceptual framework discussed here concerns link existence, while the data provides link reporting. What can be learned from the latter on the former? In section 3 we saw evidence that not all existing links are reported, but rather those that matter most to a respondent. Through the lens of the conceptual framework, individuals report those among their links that they derive the highest utility from.[30] This explains why a substantial fraction of links are only reported by one side at baseline; but it also implies that we can expect that some links existing at baseline will not have been reported by either side. Appendix 2 discusses to what extend effects on link reporting identified empirically are indicative of effects on individual utilities and on link formation; I report here results on the cases most relevant to my empirical results.

I show that the direction of an effect on reporting always (weakly) corresponds to that of the underlying effect onlink formation. For the case with both nodes treated, while we cannot distinguish whether a positive effect on link reporting comes from new links being formed or existing links becoming more important to the respondent, a positive effect on reporting is unambiguously due

[30] Plausible channels are that the links most salient to a respondent are those he or she derives the highest utility from, or that links are more likely to be reported by their initiator - which tends to be the node that derives the relatively higher utility from it.

to a positive effect on individual utility. For the case with one node treated, I show that a positive effect (from pair type C) on link reporting can be interpreted as a positive effect on link formation: even though not all newly formed links are reported, all newly reported links are newly formed links. If these are reported by the untreated node, it confirms that the new link was formed due to an increase in utility to that node.

Empirically, I will thus conduct *directional* analysis of effects of the intervention on link reporting patterns, which allows to keep the information that is contained in discordant reporting of links. This is particularly opportune as the conceptual framework indicates that for the case that only one node is treated, effects of the intervention on utilities of the two nodes go in opposite directions.

5 Empirical strategy

A main implication from my conceptual framework is that if additional links are formed as an effect of the intervention, this is because links with treated entrepreneurs become more desirable, either to untreated entrepreneurs or to other treated entrepreneurs. If, on the other hand, the intervention causes fewer links to be formed, this is because treated entrepreneurs are unwilling to share their new knowledge and expertise with untreated entrepreneurs who are potential competitors. In both cases, control group entrepreneurs would receive fewer nominations as business network partners than those in the treatment group.[31] I thus start by testing whether there are differential changes in entrepreneurs' in-degree between baseline and endline, by treatment assignment. Still at the individual-level, I then test for a difference in the evolution of the out-degree, the number of entrepreneurs in the sample a respondent mentioned as networking partners. This will give a first hint at whether any changes in the in-degree are driven by nominations from the treatment or control group. I use the following individual-level specification.

$$y_i = \alpha + \delta_1 \mathtt{treat}_i + \delta_2 y_i^{BL} + \eta_1 x_i + \nu_i \tag{3}$$

Where y_i is the statistic of interest – individual i's in-degree or out-degree –, $\mathtt{treat}_i$ is individual i's treatment assignment, and y_i^{BL} is the baseline value of the statistic; in x_i I control for cluster size.[32] Standard errors are clustered at the enterprise-cluster level, the level of randomization.

However, individual-level analysis using specification (3) can only give limited answers on the effect of the intervention. Finding different in-degrees for treated and control at endline could either mean links with treated entrepreneurs became more attractive to control entrepreneurs or links with control entrepreneurs became less attractive to treated entrepreneurs.

[31] A negative effect on the in-degree of the untreated would come from a reporting differential for existing links for pair type D induced by the intervention; see Appendix 2.

[32] I do so because while cluster size is balanced at the cluster-level there is an imbalance at the individual-level as some large clusters happen to be assigned to the treatment group.

The main analysis therefore has to be at the dyad level, which allows to separately identify the effect of target node treatment assignment for source nodes in treatment and control groups. Here and throughout, I use the random treatment assignment, not actual treatment, and effects thus represent the "intent-to-treat" effects. I use directional analysis to capture the intricacies of the reactions by the two sides as predicted by the conceptual framework. Noting that links within and across clusters require a different empirical treatment (see below), I estimate the following dyadic model with five mutually exclusive dummies:

$$g_{ij} = \alpha + \beta_1 \mathtt{diff11}_{ij} + \beta_2 \mathtt{diff10}_{ij} + \beta_3 \mathtt{diff01}_{ij} + \beta_4 same1_{ij} + \beta_5 same0_{ij} \qquad (4)$$
$$+ \theta_1 x_i + \theta_2 x_j + u_{ij}$$

This model is equivalent to an interacted model of source node treatment indicator, target node treatment indicator, and same-cluster indicator. The omitted reference category is $\mathtt{diff00}_{ij}$. The source business owner is denoted i, and the target business owner j. The dependent variable g_{ij} takes value 1 if i gave j as a business networking partner at endline and 0 otherwise. Source- and target-node-level controls can be included as x_i and x_j, respectively; I control for cluster size on both sides. Standard errors are adjusted for dyadic error correlations using the method proposed by Fafchamps and Gubert (2006).[33]

The types of node pairs that most closely correspond to the conceptual framework, and that are thus of most interest, are those from different clusters. Here, the empirical strategy relies on the randomness of source node and target node treatment assignments implied by the experiment, and the fact that these treatment assignments are orthogonal to each other. In the style of the notation used in the conceptual framework, $\mathtt{diff11}_{ij}$ denotes pairs of entrepreneurs from different clusters where both source and target node happen to be treated, $\mathtt{diff10}_{ij}$ those where the source node is treated and the target node is untreated, and $\mathtt{diff01}_{ij}$ the reverse. The omitted reference category $\mathtt{diff00}_{ij}$ refers to pairs where both nodes happen to be untreated; it describes link evolution in the absence of the intervention (see section 4).

For pairs with only one node treated, the coefficients on $\mathtt{diff01}_{ij}$ and $\mathtt{diff10}_{ij}$ thus respectively indicate whether the intervention induces the untreated node and the treated node to report a link. If the intervention increases the desirability to control group entrepreneurs of links with treatment group entrepreneurs, this will be indicated by a positive coefficient on $\mathtt{diff01}_{ij}$. If it reduces the desirability to treatment group entrepreneurs of links with control group entrepreneurs, this will

[33] This error correction accounts for the fact that dyadic data involve multiple observations per entrepreneur in the roles of source and target. Two-way clustering – simultaneously clustering by source node entrepreneur and by target node entrepreneur – would be insufficient as each individual will appear in the dataset both in the roles of source and target entrepreneur. The method proposed by Fafchamps and Gubert (2006) allows for error correlations across observations involving the same entrepreneur across roles. An alternative correction approach used with a sufficient number of separate networks (e.g. villages with non-overlapping networks, such as in Banerjee et al. (2012)) is network-level clustering, which allows for arbitrary error correlations between observations within network, and thus between all observations that involve the same individual. It does not apply here as the network can only be understood as a single network: links across neighborhoods and across sectors of business are possible and do exist.

manifest in a negative coefficient on $\mathtt{diff10}_{ij}$. The coefficient on $\mathtt{diff11}_{ij}$ indicates whether, as a result of the intervention, treated entrepreneurs report more links with other treated entrepreneurs from a different sector.

For pairs from the same cluster, whether or not that cluster is treated remains random, but treatment assignment of the two nodes is not orthogonal to each other, but instead identical. Same-cluster links then require separate treatment from different-cluster links. Here, $same1_{ij}$ refers to pairs within the same, treated cluster, and $same0_{ij}$ to pairs within the same, untreated, cluster. The coefficients on $same1_{ij}$ and $same0_{ij}$ are not interesting as such, as the reference category is $\mathtt{diff00}_{ij}$ (and being in the same or a different cluster is of course not randomly assigned). Rather, we are interested in their difference, this is, in $(\beta_4 - \beta_5)$. This difference indicates whether links within the same cluster are more likely to be kept (or new potential ones to be forged) if the cluster randomly got treated.

Specification (4) is the general specification to investigate effects on link reporting at endline. When looking at the full sample of potential links at endline, I also control for whether a link was reported at baseline (i.e. g_{ij}^{BL}, the baseline value of the dependent variable). When I look at the sample of links reported at baseline, I instead include two additional categories of baseline links, $nonsamp1$ and $nonsamp0$, which indicate links outside the main sample when the source node is assigned to treatment group and control group respectively; the difference in their coefficients can be interpreted as the effect of treatment on the likelihood to keep links outside the main sample. When I look at the sample of potential links that were not baseline links ("new links"), specification (4) remains as is.

As robustness checks, I also estimate the dyad-level equivalent of equation (3) for all three samples, and a simpler version of specification (4), including in the place of the three "$\mathtt{diff}$" dummies a single dummy, $\mathtt{diff1}_{ij}$, that indicates different-cluster links of the treated, with omitted category $\mathtt{diff0}_{ij}$ (different-cluster links of the untreated). The coefficient on $\mathtt{diff1}_{ij}$ will indicate whether effects are driven by a shift away from or towards links outside the own cluster.

Finally, the conceptual framework predicts that if there is a positive effect for pairs with one treated node on link reporting by the untreated node, this will be driven by pairs with a relatively higher status untreated node. To test this prediction, I estimate a version of equation (4) that splits each group into sub-groups by whether the source node is of higher status than the target node (denoted with "H"), or lower status (denoted with "L"):[34]

[34]This is done using the relative baseline status indicator constructed for the purposes of sub-section 3.3. Due to the continuous nature of the relative status score that this indicator is based on, there is, for each pair of nodes, exactly one observation for which the indicator takes value one (the observation that represents reporting by the relatively higher-status node), and exactly one observation for which the indicator takes value zero (the observation that represents reporting by the relatively lower-status node).

$$g_{ij} = \alpha + \beta_1 \mathtt{Hdiff11}_{ij} + \beta_2 \mathtt{Hdiff10}_{ij} + \beta_3 \mathtt{Hdiff01}_{ij} + \beta_4 \mathtt{Hsame1}_{ij} + \beta_5 \mathtt{Hsame0}_{ij} \tag{5}$$
$$+ \beta_6 \mathtt{Ldiff11}_{ij} + \beta_7 \mathtt{Ldiff10}_{ij} + \beta_8 \mathtt{Ldiff01}_{ij} + \beta_9 \mathtt{Ldiff00}_{ij} + \beta_{10} \mathtt{Lsame1}_{ij} + \beta_{11} \mathtt{Lsame0}_{ij}$$
$$+ \theta_1 x_i + \theta_2 x_j + u_{ij}$$

Due to its symmetry, this specification allows to investigate link reporting by both sides of a potential link at the same time: when both nodes of a potential different-cluster link happen to be treated, reporting by the relatively higher status node is captured by Hdiff11, and reporting by the relatively lower status node by Ldiff11. The case where neither of the nodes is treated, Hdiff00 and Ldiff00, as well as the cases involving same-cluster links, can be interpreted analogously. When only one node of the pair is treated, the coefficients that represent reporting from the two nodes on the same relationship are, for the case that the higher-status individual is treated, Hdiff10 and Ldiff01. For the case that the lower-status individual is treated they are Hdiff01 and Ldiff10.

6 Results

Descriptively, one can observe that the network does not remain static between baseline and endline (see Figure 2). Overall, among the within-main-sample baseline links of source nodes that are observed at endline, 55.1% are reported again; this number is 45.0% for the full sample of baseline links, which includes links with target nodes within and outside the main sample (Table 5. Columns 3 and 2, "Overall avg.").

6.1 In-degree and out-degree

I use specification (3) to test for a treatment effect on the in-degree, the number of other firm owners that report a business link with a given entrepreneur. I find that the in-degree of treated entrepreneurs at endline is significantly higher than that of control entrepreneurs: 1.30 compared to 1.10, controlling for baseline values (p-value 0.035, Table 4, Column 1). Treatment group entrepreneurs became more desirable as business networking partners compared to control group entrepreneurs. The in-degree of the control group at endline is the same as the average baseline in-degree,[35] hinting that the difference at endline may be driven by the *increased* attractiveness of treated entrepreneurs (from the point of view of the treated and/or untreated) rather than a *reduced* attractiveness of the untreated from the point of view of the treated. But as the counterfactual

[35]For a correct comparison, one has to use the baseline in-degree of target nodes that have not attrited, from source nodes that have not attrited either (as these represent the pool nominations can come from at endline). This baseline in-degree is 1.11 (see Table 1).

evolution of the entrepreneurs' in-degrees is unknown, this result needs to be confirmed in the dyad-level analysis using my main specification (4).

This difference in in-degrees is not simply driven by treated entrepreneurs reporting additional links to other treated entrepreneurs, as there is no significant difference between the out-degree of treatment and control group entrepreneurs (the number of business links a given entrepreneur reported), in total or within-main-sample (Table 4, Columns 2 and 3). This is confirmed in the most simple dyad-level specification: treated source nodes are not significantly more likely to report links at endline (Table 5, Column 1); this holds both for baseline links overall and from within the main sample, and for potential links that were not mentioned at baseline (Table 5, Column 2-4). Again, this result requires confirmation through the main specification (4).

6.2 Network change

Different-cluster links

I turn to the main specification of interest, starting with the analysis of links across clusters. Recall that for this subset of links, the experiment ensures that source and target treatment assignments are both random and orthogonal to each other. This corresponds to the conceptual framework setup. I find that control group entrepreneurs are significantly more likely to report a link when the target node is treated: the endline likelihood of reporting a link is increased by 33% compared to the base category average – the average link reporting likelihood when both are control group (p-value 0.083; Table 6, Column 1, `diff01`).[36] This indicates that the intervention has increased the desirability of links with treated entrepreneurs to untreated ones. I will discuss potential channels through which this may have occurred below.

This effect is driven by a significant effect on keeping baseline links: at endline, control group enterprise owners re-nominate on average 46.4% of their baseline links outside the own cluster if the target owner happens to be untreated; target treatment increases this likelihood by 13.1 percentage points, or 28% (p-value 0.084; Table 6, Column 3, `diff01`). This indicates that control group entrepreneurs keep up interactions with some of their treated baseline contacts that they would have stopped interacting with in the absence of the intervention. For new links – the set of all potential endline links that were not already given at baseline – the coefficient is positive and of similar (in fact somewhat larger) relative magnitude, but not significant (Table 6, Column 5, `diff01`).

[36] Note that both the average link reporting likelihood in the base category (see bottom of the table) and the effect in absolute terms are small. This is because the likelihood for a given link in the sample of $448,716$ potential links to be reported is small (in other words, the network is sparse). The effect relative to the base category average, however, is sizeable.

Link reporting by treated nodes, on the other hand, does not react to the intervention. There is no net negative effect on the desirability of links with untreated entrepreneurs, whether for all potential endline links, for baseline links, or for potential new links (Table 6, Columns 1, 3, and 5, `diff10`). It does not seem to be the case that the intervention induces treated entrepreneurs to avoid links with untreated entrepreneurs; in other words, there is no appreciable negative competition effects to the treated of linking with untreated. There is also no positive effect on the desirability of links with other treated entrepreneurs outside the own cluster (Table 6, Columns 1, 3, and 5, `diff11`).

I turn to results from estimating differential effects by nodes' relative status, using equation (5). We have seen that the training intervention has a positive effect for untreated nodes on the desirability of links with treated nodes. The conceptual framework predicts that the channel for this effect is the change in the relative status induced by the intervention: to relatively higher baseline-status untreated nodes, previously undesirable links with relatively lower baseline-status nodes become more attractive if the latter received the training. The observed overall effect then has to come from pairs where the untreated node has relatively higher status. Table 7, Panel A, gives a succinct representation of the results from this regression for potential endline links outside the own cluster.[37] Note that the coefficients on the parameters with an "H" in the name represent the report of the relatively higher baseline-status individual on the existence of the link, while the coefficients on the parameters with an "L" in the name represent the report of the relatively lower baseline-status individual.

Two main results stand out. First, the intervention almost doubles the likelihood that a relatively higher baseline-status node reports a link at endline, for the case that the lower baseline-status node is treated and the higher status node is not (+89%, p-value 0.011, Column 7, Row i). This is in line with the conceptual framework prediction. As expected, there is no effect on link reporting by the relatively higher status node for any of the other treatment assignment cases (that is, when both are treated or when the higher status node is the treated node (Columns 5 and 6, Row i)). As expected, the intervention also has no effect on link reporting by the relatively lower status node (Columns 5-7, Row ii). In fact, reporting likelihoods are remarkably similar within relative status type, and the only coefficient that stands out as unusual is reports by untreated higher status nodes of links with lower status treated nodes (Columns 1-4, Rows i and ii). This leads to the conclusion that the positive effect on reporting of links with treated entrepreneurs by the untreated is indeed driven by relatively higher baseline-status untreated entrepreneurs, as predicted by the conceptual framework.

Second, for the comparison group (both nodes of the pair are in the control group), the lower

[37] The full set of coefficients with standard errors is shown in the Appendix 3, Table A1, Column 1. The full sets of coefficients with standard errors that correspond to Table 7, Panels B and C discussed below are in Table A1, Columns 2 and 3 respectively.

baseline-status node is significantly more likely to report a link (Column 1, Row iii). This is in line with the directional link reporting pattern observed for the baseline network (see section *3*). The pattern remains unchanged when both nodes are in the treatment group or only the relatively higher status node is treated (Columns 2 and 3, Rows i to iii). Remarkably, however, there is no difference in the reporting likelihood by node relative status when only the relatively lower-status node is treated (Column 4, Row iii). In light of the conceptual framework, this leads to the conclusion that being assigned treatment allows some relatively lower baseline-status nodes to catch up in their desirability as linking partners with some higher baseline-status nodes. In other words, it gives lower status entrepreneurs access to better networks.

Splitting the analysis sample into potential links that were reported at baseline and potential links that were not (i.e. "new links"), I find the overall results mirrored for the latter sample (Table 7, Panel C). Indeed, while in the main specification, the point estimate for new links reported by the untreated node in a mixed treatment status pair was large but not significant (Table 6, Column 5, `diff01`), it becomes significant in the subgroup that is expected to drive the effect: the intervention more than doubles the likelihood that a relatively higher baseline-status node reports a new link (+141%, p-value 0.019, Column 7, Row vii). As in the overall sample, this effect is large enough that it eliminates the difference in reporting of new links between relatively higher and lower baseline-status nodes that is otherwise observed. Crucially, this effect shows that the intervention not only affects whether baseline links are kept, but also the way new links are formed, at least for the subgroup that should drive such an effect.

For the sample of baseline links, while some of the coefficients and differences of interest are large economically, none of them are statistically significant, so that it is not possible to draw robust conclusions (Table 7, Panel B). This is due to low power to conduct differential analysis on different-cluster baseline links, owing to an already fairly low number of observations in each of the cells.

These findings raise the question what it is that makes the treated more popular as networking partners to the control group. Is it that the control group firms wish to get access to the information provided in the training and learn the practices taught from the treated, is it that this information and training permitted the treated to improve their business performance, which makes them more attractive for business collaboration, or is it the simple perception that are doing better?

In the impact evaluation of the intervention, Campos, Goldstein, Pimhidzai, Stein and Zia (2018) find small effects on the adoption of practices taught in the training but not on measured knowledge. They find no effects on firm revenues or profits but cannot exclude effects on owner perceptions of the business as successful and on the number of employees. The do-it-yourself handouts, which training participants received for each topic at the end of the training sessions, spread through the baseline network. Given these results, which include elements of all three mechanisms but no strong evidence on either, it is not possible to draw clear conclusions on what factor drives the effect. Anecdotally, control group firms were aware of the training while it was happening, and were reminded of it afterwards as the treated put up their training certificates in their workshops and shared the training handouts. Thus, it is possible that the status of the treated increased from the point of view of the control group without the tangible correlates of status (such as business size, owner knowledge) changing. Whichever the mechanism, it is clear that status is part of the picture. When control group firms seek out treated firms, for any of the motives discussed above, they (successfully) approach a treated of relatively lower baseline status.

Same-cluster links

Recall that treatment assignment of a given cluster is random, but because treatment assignment does not vary within cluster, it is not possible to separately estimate the effect of source and target treatment assignment for baseline links within the same cluster; one can, however, compare treatment and control clusters. Overall and for new links, I find no effect of treatment on same-cluster links at endline (Table 6, Columns 1 and 5, difference $\mathtt{same1}$–$\mathtt{same0}$ at the bottom of the table).[38] This is consistent with the absence of an effect on $\mathtt{diff11}$ above, which too compares pairs with two treated nodes with pairs with two control nodes, but across clusters.

For baseline links, though, I find that treatment has a significant positive effect: 62.1% of links within treatment group clusters are kept at endline, compared to only 51.2% of links within control group clusters (Table 6, Column 3, difference $\mathtt{same1}$–$\mathtt{same0}$ at the bottom of the table, p-value 0.050). The most plausible explanation is that this is a (side-)effect of jointly attending the training; for individuals who were already collaborating before the intervention, going to the same training sessions increased the trust in each other's business standards and practices (reducing the "other costs" of being linked) or created social ties (increasing the "other benefits" of being linked). Continuing discussions with other participants may also have allowed trainees to better understand certain elements of the training (another form of "other benefits" of being linked).

Support for this explanation comes from the fact that for *different*-cluster links, it is not the case that the treated keep more of their baseline links with other treated (see above and Table 6, Column 3, $\mathtt{diff11}$). Indeed, by construction of the experiment, it is close-by entrepreneurs from the same sector that share a cluster. These are also more likely to have attended the same training sessions (which were, for the technical training, sector-specific and for both trainings in groups formed based on geographic proximity).[39]

[38] Given the setup of the specification as an interacted model, the former result is given by the sum of the "Comparison group average" and the coefficient on $\mathtt{same1}$, and the latter by the sum of the "Comparison group average" and the coefficient on $\mathtt{same0}$. The fact that the coefficients on $\mathtt{same1}$ and $\mathtt{same0}$ are different between the two columns of each sample results from different omitted categories; the quantity of interest – the difference between the two coefficients – is the same.

[39] An alternative explanation would be a direct training effect, if owners learned at the training that collaboration

Robustness

As a robustness check, I exclude the possibility that the observed effect patterns within and across clusters are the result of treated entrepreneurs shifting from different-cluster links to same-cluster links. I estimate a simpler version of specification (4), including instead of the three "diff" dummies a single dummy, diff1, that indicates different-cluster links of the treated; the omitted category is now diff0 (different-cluster links of the untreated). I find that treatment and control group enterprise owners are equally likely to report links with someone from a different cluster, be it for all potential links, baseline links, or new links (Table 6, Columns 2, 4 and 6,diff1). Similarly, I confirm that respondents do not shift towards keeping baseline links with entrepreneurs from the main sample, and away from keeping non-sample baseline links, which are by definition untreated (Table 6, Column 4, difference "$\mathtt{nonsamp1} - \mathtt{nonsamp0}$" at the bottom of the table).

Both findings suggests that while the intervention induces treated firms to keep more same-cluster links (with partners who are by definition treated), this is not at the expense of links with partners outside the cluster, be it different-cluster partners (that are less likely to be treated) or non-sample partners (that are untreated). This is further evidence against the idea that treated entrepreneurs bear a competition cost from maintaining their links with untreated entrepreneurs.

7 Conclusion

I use novel panel network data, in combination with random variation in the treatment assignment of the two sides involved in a link that is induced by a skills training intervention, to investigate whether and how business links between urban micro-enterprises in Kampala, Uganda, changed in response to the intervention. I find that the network is indeed endogenous to this public policy intervention.

I find evidence of strategic behavior in the endline linking decision of untreated entrepreneurs. The intervention modifies the relative status differentials between entrepreneurs, and increases the desirability to untreated entrepreneurs of links with relatively lower baseline status treated entrepreneurs. This is as predicted by the conceptual framework. Baseline links within clusters of treated enterprises are strengthened, which is not due to a strategic replacement of untreated with treated partners but seems to be an effect of jointly attending the training, by increasing trust in each other's business practices and standards, or tightening social ties. Treated entrepreneurs do not seem to bear a competition cost from maintaining their links with untreated entrepreneurs.

with other entrepreneurs from the same sector or in close proximity can be beneficial. However, the training did not include any such suggestions.

The finding that networks change in response to a public policy intervention has significant policy relevance, as understanding the mechanisms of information diffusion and social learning through networks may help to better target interventions. If the analysis of these mechanisms fails to take into account the possibility that networks change in response to the intervention studied, and instead takes the observed pre-intervention network as given, as is common in the literature, it will fail to capture the full effect of the intervention and risks giving biased estimates of the extent of social learning. The findings presented in particular suggest caution when designing interventions that build diffusion through networks into program design, as this approach requires targeting the intervention to some individuals in a group but not others, which may change the incentives underlying link formation. Such unintentional network rewiring can create new insiders and outsiders, with possibly negative equity implications.

References

[1] Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna and B. A. Olken. 2016. "Network structure and the aggregation of information: Theory and evidence from Indonesia," American Economic Review, 106(7): 1663-1704.

[2] Annan, J. and C. Blattman. 2016. "Can Employment Reduce Lawlessness and Rebellion? A Field Experiment with High-Risk Men in a Fragile State," American Political Science Review, 110(1): 1-17.

[3] Ball, B. and M. E. J. Newman. 2013. "Friendship networks and social status," Network Science, 1(1): 16–30.

[4] Bandiera, O. and I. Rasul. 2006. "Social Networks and Technology Adoption in Northern Mozambique," Economic Journal, 116: 869–902.

[5] Banerjee, A., A. G. Chandrasekhar, E. Duflo and M. O. Jackson. 2012. "The Diffusion of Microfinance," NBER Working Paper No. 17743, National Bureau of Economic Research.

[6] Banerjee, A., A. G. Chandrasekhar, E. Duflo and M. O. Jackson. 2013. "The Diffusion of Microfinance," Science, 341(6144).

[7] Banerjee, A., A. G. Chandrasekhar, E. Duflo and M. O. Jackson. 2018a. "Changes in social network structure in response to exposure to formal credit markets," Working Paper.

[8] Banerjee, A., A. G. Chandrasekhar, E. Duflo and M. O. Jackson. 2018b. "Using Gossips to Spread Information - Theory and Evidence from two Randomized Controlled Trials," Working Paper.

[9] Barnhardt, S., E. Field, and R. Pande. 2017. "Moving to Opportunity or Isolation? Network Effects of a Randomized Housing Lottery in Urban India," American Economic Journal: Applied Economics, 9(1): 1-32.

[10] Barr, A. 2000. "Social Capital and technical information flows in the Ghanaian manufacturing sector," Oxford Economic Papers, 52(3): 539–559.

[11] Beaman, L., A. BenYishay, J. Magruder, and M. Mobarak. 2018. "Can Network Theory-based Targeting Increase Technology Adoption?," Working Paper.

[12] Beaman, L. and A. Dillon. 2018. "Diffusion of Agricultural Information within Social Networks: Evidence on Gender Inequalities from Mali," Journal of Development Economics, 133: 147-161.

[13] BenYishay, A., and A. M. Mobarak. 2018. "Social Learning and Incentives for Experimentation and Communication," Review of Economic Studies, forthcoming.

[14] Cai, J., A. de Janvry, and E. Sadoulet. 2015. "Social Networks and the Decision to Insure," American Economic Journal: Applied Economics, 7(2): 81–108.

[15] Cai, J. and A. Szeidl. 2018. "Interfirm Relationships and Business Performance," Quarterly Journal of Economics, 133(3): 1229–1282.

[16] Campos, F., M. Goldstein, O. Pimhidzai, M. Stein and B. Zia. 2018. Social learning or spillovers? Indirect effects of technical and business skills training for small entrepreneurs. Preliminary Working Paper.

[17] Carrell, S., B. Sacerdote and J. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," Econometrica, 81(3): 855-882.

[18] Carter M. R., R. Laajaj and D. Yang. 2014. "Subsidies and the Persistence of Technology Adoption: Field Experimental Evidence from Mozambique," NBER Working Paper No. 20465. National Bureau of Economic Research.

[19] Carver, K., K. Joyner, and J. R. Udry. 2003. "National estimates of adolescent romantic relationships", in P. Florsheim, ed.: Adolescent romantic relations and sexual behavior: Theory, research, and practical implications, pp. 23-56. Mahwah: Lawrence Erlbaum Associates Publishers.

[20] Comola, M. and M. Fafchamps. 2014. "Testing Unilateral and Bilateral Link Formation," Economic Journal, 124(579): 954-975.

[21] Comola, M. and M. Fafchamps. 2017. "The Missing Transfers: Estimating Mis-reporting in Dyadic Data," Economic Development and Cultural Change, 65(3): 549-582.

[22] Comola, M. and S. Prina. 2018. "Treatment Effects Accounting for Network Changes," Working Paper.

[23] Conley, T. G. and C. R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana," American Economic Review, 100(1): 35-69.

[24] De Weerdt, J. and S. Dercon. 2006. "Risk-sharing networks and insurance against illness," Journal of Development Economics, 81(2): 337–56.

[25] De Weerdt, J. and M. Fafchamps. 2011. "Social identity and the formation of health insurance networks," Journal of Development Studies, 47(8): 1152–77.

[26] Fafchamps, M. and F. Gubert. 2006. "The formation of risk sharing networks," Journal of Development Economics, 83: 326 – 350.

[27] Fafchamps, M., and S. Quinn. 2018. "Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment," World Bank Economic Review, 32(3): 656–675.

[28] Fafchamps, M. and M. Soderbom. 2014. “Network Proximity and Business Practices in African Manufacturing,” World Bank Economic Review, World Bank Group, 28(1): 99-129.

[29] Feigenberg, B., E. Field, and R. Pande. 2013. “The Economic Returns to Social Interaction: Experimental Evidence from Microfinance,” Review of Economic Studies, 80: 1459–1483.

[30] Fernando, A. N. and G. Sharma. 2016. “Seeking Treatment: The Effects of Technology on Social Interactions and Network Structure in Agriculture,” Working Paper.

[31] Foster, A. and M. Rosenzweig. 1995. “Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture,” Journal of Political Economy, 103(6): 1176-1209.

[32] Granovetter, M. 2005. “The Impact of Social Structure on Economic Outcomes,” Journal of Economic Perspectives, 19(1): 33-50.

[33] Hardy, M. and J. McCasland. 2018. “It Takes Two: Experimental Evidence on the Determinants of Technology Diffusion,” Working Paper.

[34] Hotelling, H. 1933. “Analysis of a Complex of Statistical Variables into Principal Components,” Journal of Educational Psychology, 24(6): 417–441.

[35] Humphreys, M. and J. M. Weinstein. 2007. “Demobilization and Reintegration,” Journal of Conflict Resolution, 51(4): 531-567.

[36] Jackson, M. O. 2008. Social and Economic Networks. Princeton: Princeton University Press.

[37] Jackson, M. O. 2011. “An overview of social networks and economic applications,” in J. Benhabib, A. Bisin, and M. O. Jackson, eds.: Handbook of Social Economics, Vol. 1, pp. 511–86. Amsterdam: North Holland.

[38] Jackson, M. O. and A. Wolinsky. 1996. “A Strategic Model of Social and Economic Networks,” Journal of Economic Theory, 71(1): 44-74.

[39] Jackson, M. O., and L. Yariv. 2011. “Diffusion, Strategic Interaction, and Social Structure,” in J. Benhabib, A. Bisin, and M. O. Jackson, eds.: Handbook of Social Economics, Vol. 1, pp. 645-678. Amsterdam: North Holland.

[40] Kim D. A., A. R. Hwong, D. Stafford, D. A. Hughes, A. J. O’Malley, J. H. Fowler and N. A. Christakis. 2015. “Social Network Targeting to Maximise Population Behaviour Change: a Cluster Randomised Controlled Trial,” Lancet, 386(9989): 145-153.

[41] Kremer, M. and E. Miguel. 2007. “The Illusion of Sustainability,” Quarterly Journal of Economics, 122: 1007-1065.

[42] Leider, S., M. M. Möbius, T. Rosenblat and Q.-A. Do. 2009. “Directed Altruism and Enforced Reciprocity in Social Networks,” Quarterly Journal of Economics, 24(4): 1815-1851.

[43] McMillan, J. and C. Woodruff. 1999. "Interfirm Relationships and Informal Credit in Vietnam", Quarterly Journal of Economics, 114(4): 1285–1320.

[44] Magnan, N., D. J. Spielman, T. J. Lybbert and K. Gulati. 2015. "Leveling with Friends: Social Networks and Indian Farmers' Demand for a Technology with Heterogenous Benefits," Journal of Development Economics, 116: 223-251.

[45] Miller, G. and A. M. Mobarak. 2015. "Learning about New Technologies through Social Networks: Experimental Evidence on Nontraditional Stoves in Bangladesh." Marketing Science, 34(4): 480-499.

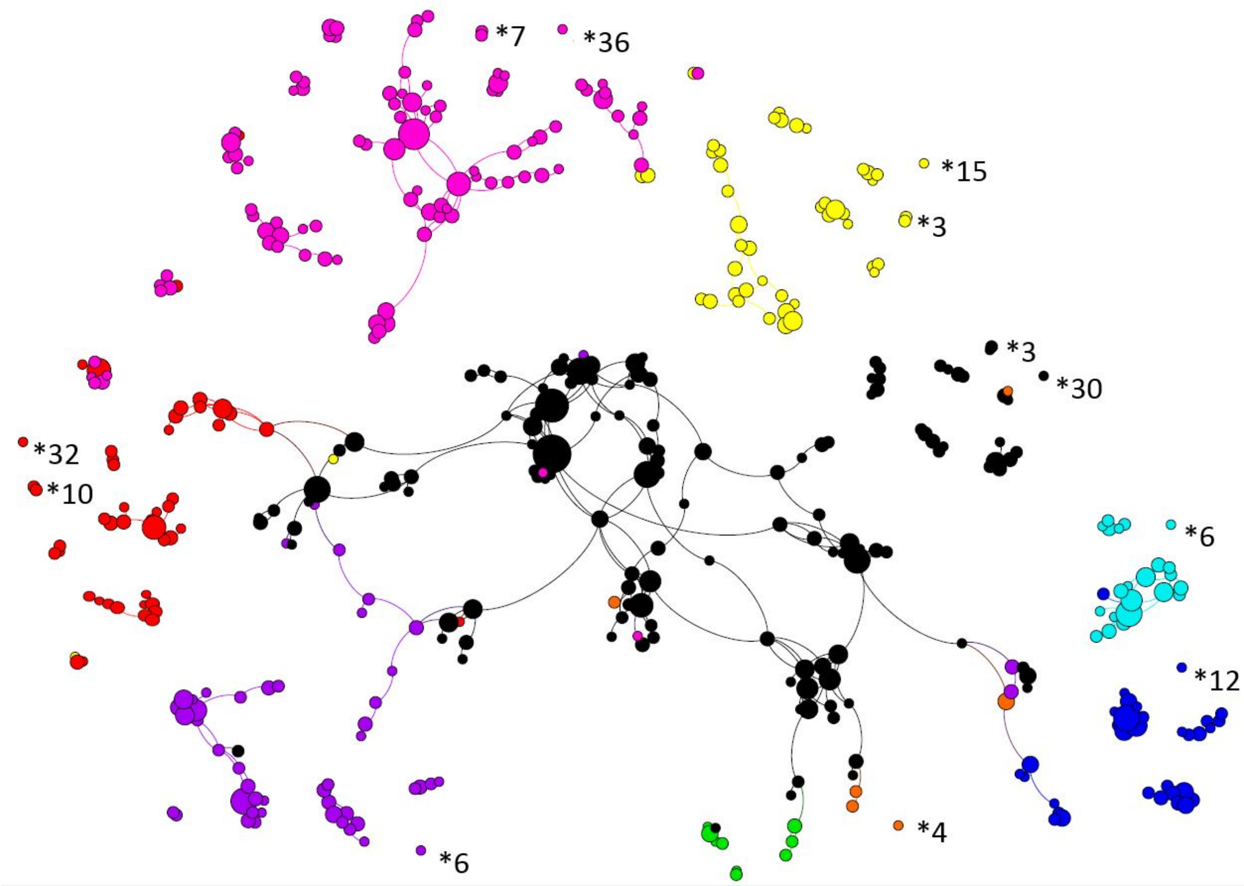
[46] Mobius, M. and T. Rosenblat. 2014. "Social Learning in Economics," Annual Review of Economics, 6: 827-847.

[47] Oster, E. and R. Thornton. 2012. "Determinants of Technology Adoption: Private Value and Peer Effects in Menstrual Cup Take-Up," Journal of the European Economic Association, 10(6): 1263-1293.

[48] United Nations. 2014. Operational Guide to the Integrated Disarmament, Demobilization and Reintegration Standards (IDDRS), http://unddr.org/uploads/documents/Operational%20Guide.pdf.

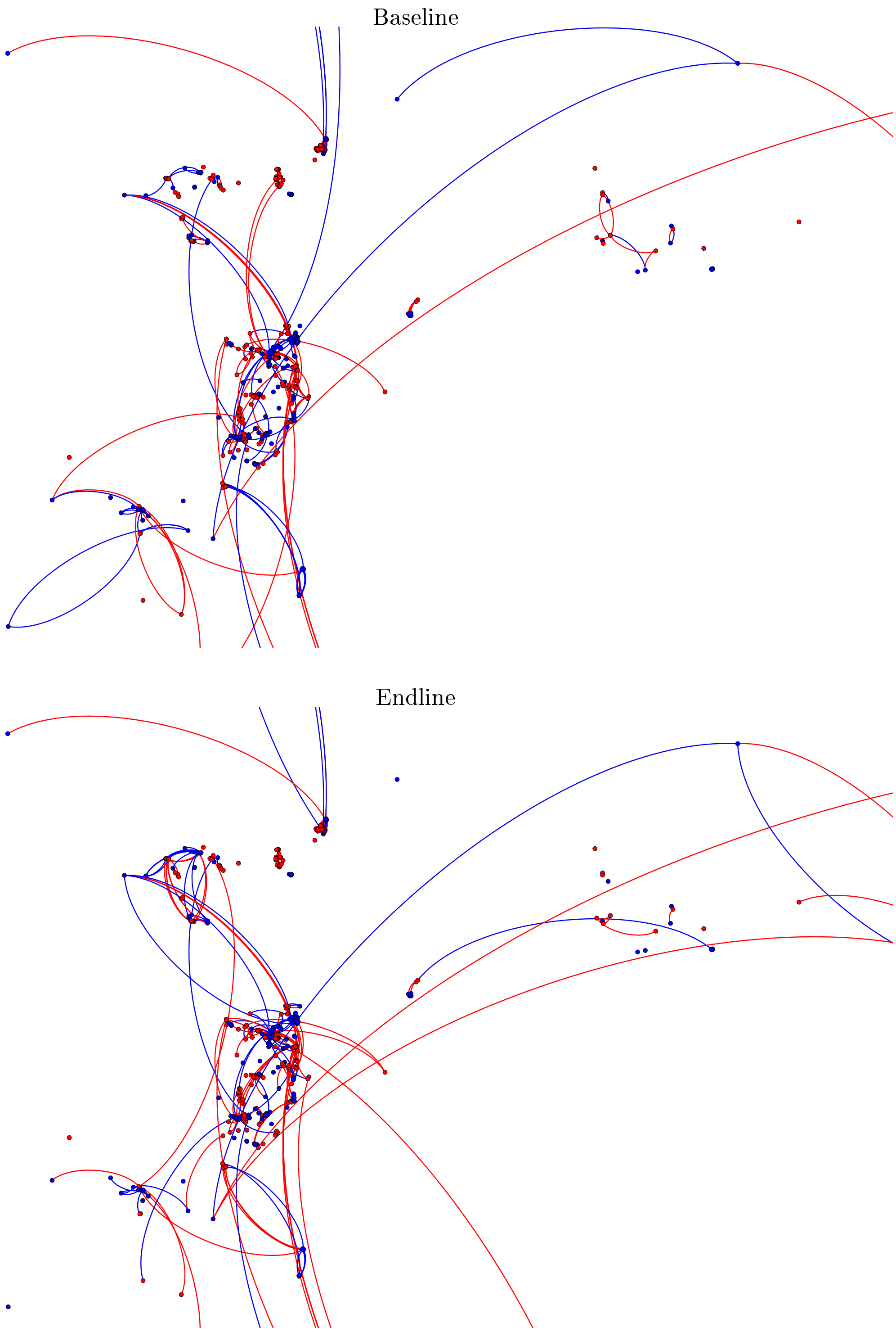
[49] Vyas, S. and L. Kumaranayake. 2006. "Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis," Health Policy and Planning, 21(6): 459–468.

Figure 1: **Links at baseline**



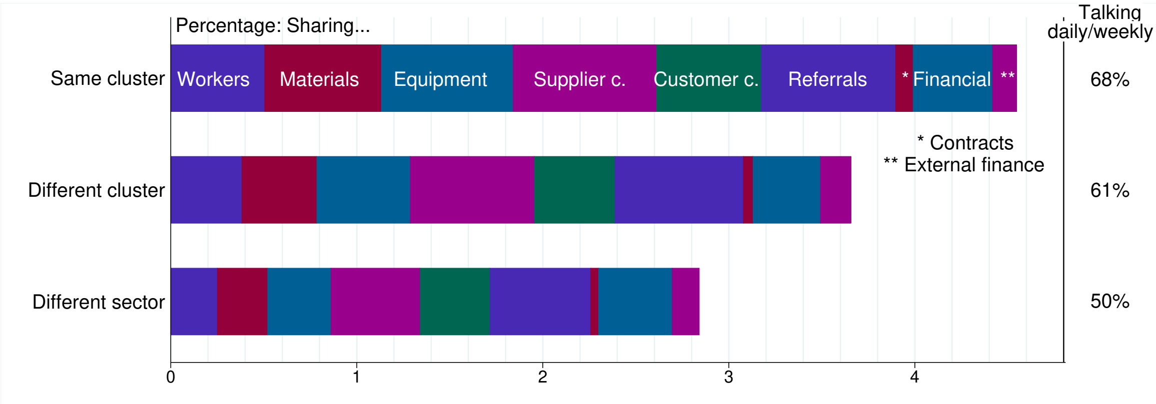
Note: Node size indicates in-degree ("popularity"). Node color indicates node's sector of business; Link color indicates source node sector of business. Black: metal fabrication, pink: tailoring, red: catering, yellow: hair salon, purple: foundry & forging, dark blue: carpentry, light blue: shoe making, green: electrical, and orange: machining.

Figure 2: **Links at baseline and endline**



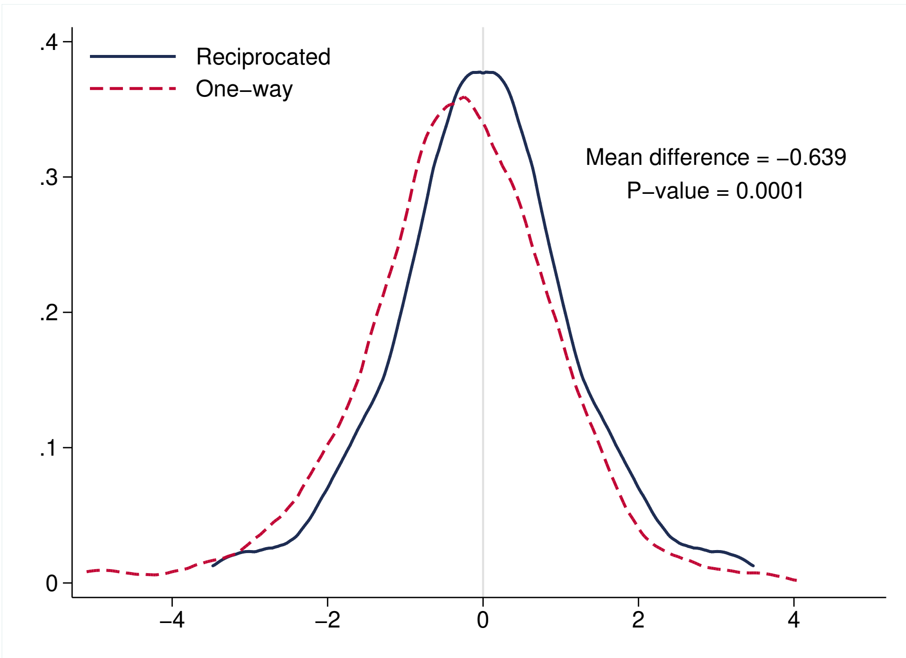
Note: Node location: GPS coordinate based. Node color indicates node's treatment assignment; Link color indicates target node treatment assignment. Red: treatment; Blue: control.

Figure 3: **Business collaboration and information exchange (baseline)**



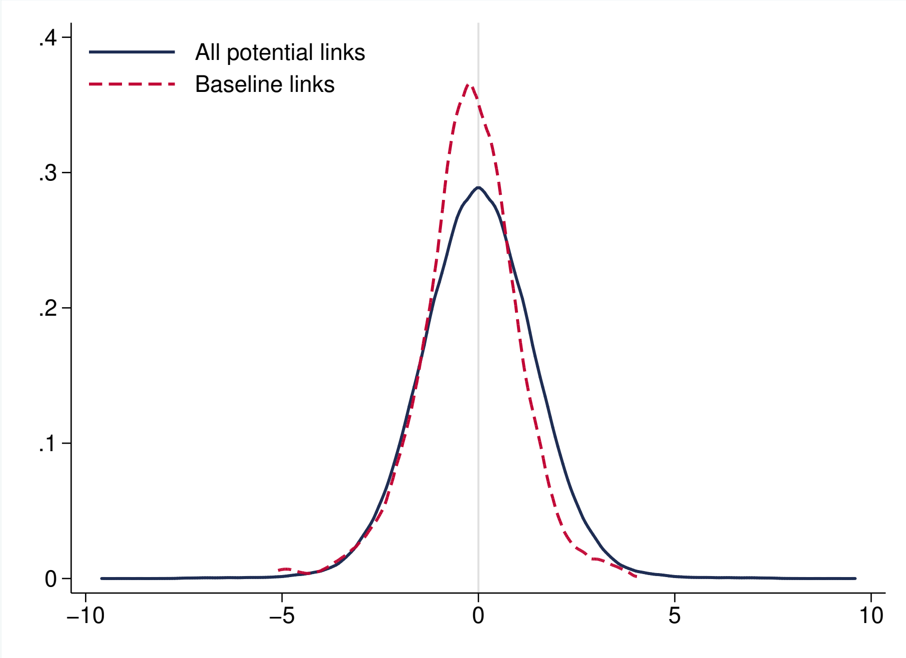
Note: Link-level business collaboration along nine dimensions and frequency of information exchange. The first bar is for same-cluster links (438 observations), the second for different-cluster links from the same sector (260 observations) and the third fir different-sector links which are always from different clusters (48 observations). The segment width represents the share of source nodes reporting collaboration in the given dimension in the past 12 months (hence the maximum segment width is 1). The total bar width represents the average number of collaboration dimensions (out of nine). "Supplier c." stands for "Supplier contacts", "Customer c." for "Customer contacts", "Referrals" for "Referred customers to each other", and "Financial" for "Financial interaction (borrow/lend)". The right-most column indicates the share of links within group that share business information at least daily or weekly.

Figure 4: **Relative status score by baseline link type**



Note: Kernel density plot of the link-level relative status score, by baseline link type: Reciprocated and one-way. Sample includes all 914 within-main-sample baseline links. The mean difference and its p-value are from a regression of the relative status score on a "one-way"-dummy, using dyadic standard errors.

Figure 5: **Relative status score, all potential links vs. baseline links**



Note: Kernel density plots of the link-level relative status score, of all potential links (536,556 observations) and of baseline links (914 observations).

Table 1: **Baseline summary statistics: Enterprise-level**

	N	Mean	Sd	Min	Max
Enterprise-level, individual characteristics					
Female	613	0.452	0.498	0	1
School years	591	9.003	3.117	0	16
Age	609	38.353	9.906	18	81
Enterprise-level, network characteristics					
In-degree (within)	613	1.294	1.702	0	15
In-degree (within, non-attrited sources)	613	1.114	1.514	0	12
Out-degree (full)	613	2.958	1.502	0	5
Out-degree (within)	613	1.243	1.242	0	5
Out-degree (same cluster)	613	0.728	1.014	0	5
Enterprise-level, firm characteristics					
Business age	609	12.588	9.180	1	53
Registered	613	0.568	0.496	0	1
Number of employees	613	2.966	3.789	0	45
Revenues (last month, million UGX)	606	2.330	5.024	0	60
Profits (last month, million UGX)	603	0.391	0.615	0	5
Sector: Metal fabrication	613	0.266	0.442	0	1
" : Tailoring	613	0.238	0.426	0	1
" : Catering	613	0.157	0.364	0	1
" : Hair salon	613	0.106	0.308	0	1
" : Foundry and Forging	613	0.086	0.281	0	1
" : Carpentry	613	0.072	0.258	0	1
" : Shoe making	613	0.044	0.205	0	1
" : Machining	613	0.016	0.127	0	1
" : Electrical	613	0.015	0.120	0	1
Cluster size	613	8.183	7.970	1	33

Note: Baseline summary statistics for the 613 firms that are still in the endline sample.

Table 2: **Baseline summary statistics: Link-level**

	Within main sample					*Full sample*	
	N	Mean	Sd	Min	Max	N	Mean
Link is reciprocated	762	0.318	0.466	0	1		
Distance (in 100m)	762	1.906	8.269	0	97.2		
Same cluster	762	0.585	0.493	0	1		
Same gender	762	0.891	0.312	0	1	1813	0.868
Same line of business	753	0.932	0.251	0	1	1786	0.756
Relationship: friend	759	0.621	0.486	0	1	1800	0.629
" : business relation	759	0.173	0.378	0	1	1800	0.146
" : spouse, other family	759	0.158	0.365	0	1	1800	0.176
" : neighbor, other	759	0.049	0.215	0	1	1800	0.049
Knows target for ... (in years)	762	10.748	8.552	.167	50	1813	10.573
Share information every day or week	740	0.639	0.481	0	1	1768	0.591
Any business collaboration (out of following 9)	744	0.968	0.177	0	1	1728	0.907
Shared (past 12 months): Employees	759	0.445	0.497	0	1	1804	0.329
" : Materials	762	0.526	.500	0	1	1813	0.405
" : Equipment	757	0.613	0.487	0	1	1801	0.446
" : Supplier contacts	759	0.717	0.451	0	1	1802	0.574
" : Customer contacts	759	0.505	0.500	0	1	1806	0.447
" : Contracts	761	0.076	0.266	0	1	1808	0.074
Past 12 months: Referred customers to each other	762	0.699	0.459	0	1	1813	0.634
" : Financial interaction (borrow/lend)	759	0.402	0.491	0	1	1810	0.369
" : External finance	762	0.143	0.350	0	1	1813	0.119
Source status score	762	0.123	1.459	-5.309	5.677	1813	0.124
Target status score	762	0.336	1.501	-5.552	5.677		
Relative status score	762	-0.163	1.240	-5.100	4.036		
Source has higher status score than target	762	0.450	0.498	0	1		

Note: Baseline link characteristics; variables in rows 1-3 and 2-25 constructed by the author, variables in rows rows 4-21 as reported by the source entrepreneur.

Table 3: **Baseline: correlation of linking and status**

	(1)	(2)	(3)
Source status score	0.000009		
	(0.000062)		
Target status score	0.000259***		
	(0.000077)		
Relative status score		-0.000223***	
		(0.000056)	
Source has higher status score than target			-0.000537***
			(0.000136)
Constant	0.000617	0.001712	0.000613
	(0.000776)	(0.001132)	(0.000708)
Controls	Sector FEs	Sector FEs	Sector FEs
Observations	536,556	536,556	536,556

*** p<0.01, ** p<0.05, * p<0.1. All models estimated by OLS. Dyadic standard errors in parentheses. The dependent variable is a dummy indicating that the existence of a link was reported by the source entrepreneur.

Table 4: **Entrepreneur-level: Impact on in-degree and out-degree**

	(1) In-degree	(2) Out-degree (full)	(3) Out-degree (within)
Treat	0.205**	0.019	0.059
	(0.096)	(0.136)	(0.103)
BL-value of dependent variable	0.838***	0.204***	0.454***
	(0.061)	(0.043)	(0.034)
Constant	0.017	2.425***	0.345***
	(0.083)	(0.180)	(0.080)
Control group avg.	1.098	3.182	1.213
Overall avg.	1.178	3.189	1.254
Controls	lc	lc	lc
Observations	613	613	613

Note: *** p<0.01, ** p<0.05, * p<0.1. All models estimated by OLS. Standard errors clustered by firm cluster. Column (1): Baseline value of the dependent variable is the in-degree from source nodes that are non-attrited at endline. Columns (2), (3): Out-degree (full) is the total number of links reported at endline, out-degree (within) the total number reported that is from within the main sample.

Table 5: **Dyad-level equivalent of out-degree**

	(1) Link at endline	(2) Baseline links (full)	(3) Baseline links (within)	(4) New links
Treat	0.00009	0.033	0.050	0.00001
	(0.00012)	(0.030)	(0.043)	(0.00010)
Baseline link	0.55025***			
	(0.02217)			
Constant	0.00001	0.391***	0.467***	0.00017
	(0.00017)	(0.029)	(0.048)	(0.00014)
Control group avg.	0.00166	0.430	0.524	0.00076
Overall avg.	0.00171	0.450	0.551	0.00078
Controls	lc nlc	lc nlc	lc nlc	lc nlc
Observations	448716	1813	762	447954

Note: *** p<0.01, ** p<0.05, * p<0.1. All models estimated by OLS. Dyadic standard errors in parentheses. Columns (1), (3), (4) use the full sample of baseline links of endline sample firms; Column (2) uses their within-main-sample links.

Table 6: **Dyad-level impact**

	Link at endline		Baseline links		New links	
	(1)	(2)	(3)	(4)	(5)	(6)
diff-1		-0.00007		-0.027		-0.00005
		(0.00008)		(0.064)		(0.00006)
diff-11	0.00005		0.039		0.00003	
	(0.00014)		(0.090)		(0.00011)	
diff-10	0.00004		0.044		0.00001	
	(0.00011)		(0.077)		(0.00009)	
diff-01	0.00022*		0.131*		0.00013	
	(0.00013)		(0.076)		(0.00010)	
same-1	0.05189***	0.05177***	0.157**	0.086	0.04699***	0.04691***
	(0.00768)	(0.00767)	(0.072)	(0.065)	(0.00683)	(0.00682)
same-0	0.04293***	0.04281***	0.048	-0.022	0.04995***	0.04987***
	(0.00801)	(0.00800)	(0.071)	(0.062)	(0.00791)	(0.00791)
Baseline link	0.52320***	0.52320***				
	(0.02252)	(0.02252)				
Constant	0.00080***	0.00091***	0.456***	0.518***	0.00088***	0.00094***
	(0.00016)	(0.00015)	(0.055)	(0.052)	(0.00013)	(0.00012)
same1 - same0	0.00896	0.00896	0.109	0.109	-0.00296	-0.00296
...p-value	0.425	0.425	0.050	0.051	0.777	0.777
diff11 - diff10	0.00001		-0.005		0.00001	
...p-value	0.964		0.956		0.920	
nonsamp1 - nonsamp0			0.025	0.025		
...p-value			0.493	0.495		
Comparison grp. avg.	0.00067	0.00076	0.464	0.532	0.00035	0.00040
Controls	lc nlc	lc nlc	lc nlc	lc nlc	lc nlc	lc nlc
Observations	448716	448716	1813	1813	447954	447954

Note: *** p<0.01, ** p<0.05, * p<0.1. All models estimated by OLS. Dyadic standard errors in parentheses.

Table 7: **Dyad-level: Differential effects by node relative status for different-cluster links**

		Link reporting likelihoods						
		(1) Both control	(2) Both treatment	(3) Only H treament	(4) Only L treatment	(5) (2) - (1)	(6) (3) - (1)	(7) (4) - (1)
Panel A								
(i)	Report by H	0.000472	0.000528	0.000511	0.000891	0.000056	0.000039	0.000419**
(ii)	Report by L	0.000849	0.000882	0.000849	0.000885	0.000033	0.000000	0.000037
(iii)	(ii)-(i)	0.000377**	0.000354**	0.000338**	-0.000006			
Panel B								
(iv)	Report by H	0.382	0.333	0.487	0.549	-0.049	0.104	0.167
(v)	Report by L	0.549	0.629	0.630	0.532	0.080	0.081	-0.017
(vi)	(v)-(iv)	0.167	0.296***	0.144	-0.017			
Panel C								
(vii)	Report by H	0.000217	0.000293	0.000186	0.000523	0.000076	-0.000031	0.000306**
(viii)	Report by L	0.000485	0.000450	0.000422	0.000536	-0.000035	-0.000063	0.000051
(ix)	(viii)-(vii)	0.000268**	0.000157	0.000236**	0.000013			

Note: Panel A: Coefficients in columns (1) to (4), rows (i) and (ii), are derived from Table A1, for links outside the own cluster (coefficients whose namesinclude "diff"), denoting "H" the relatively higher-status node in a pair and "L" the relatively lower-status node. Row (iii) shows the difference between the coefficients in rows (ii) and (i); stars denote significance level of the difference. Columns (5) to (7) show the difference between the coefficients of columns (2), (3), and (4) respectively, and those from column (1); stars denote significance level of the difference. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Estimated by OLS on the sample of 448,716 potential endline links, using dyadic standard errors. Panel B: Rows (iv) to (vi) are the analogues of rows (i) to (iii) for the sample of 762 baseline links. Panel C: Rows (vii) to (ix) are the analogues of rows (i) to (iii) for the sample of 447954 potential new links.

Appendix 1: Construction of the relative status score and indicator

To construct the dyad-level relative status score and indicator, I proceed in five steps. First, I compute the differences between source and target node of all dyads (that is all potential links) in the main sample, in terms of the baseline value of each of nine elements of status in the dimensions business size and success (number of employees, past month revenues, past month profits, total assets), and owner knowledge and experience (schooling, financial literacy, technical knowledge, owner age and business age). For revenues, profits and assets, I use the inverse hyperbolic sine transformation as an alternative to the log transformation in the presence of zero values, in order to limit the influence of outliers; results are robust to using their levels (available upon request). Second, I convert the differences into z-scores, standardizing within sector for dyads from the same sector of business (as there are sector-level idiosyncrasies in elements such as typical firm size), and standardizing overall for dyads across sectors.

Third, I conduct principal component analysis to aggregate the nine z-scores into the relative status score while maintaining the maximum amount of information (Hotelling (1933), Vyas and Kumaranayake (2006)). This method derives the principal components of the set of original variables, that is, the orthogonal linear combinations of these variables that (successively) explain a maximum proportion of the total variance. Results are in Appendix 3, Table A2. I use the first principal component; thus, fourth, each dyad's relative status *score* is constructed as the linear combination of its values of the nine z-scores using the first principal component's scoring coefficients as weights. Five, the relative status *indicator* is defined to take value 1 when the relative status score is positive (that is, when the source node has higher status compared to the target node) and to take value 0 otherwise. There are no ambiguities in coding this variable as, due to to the continuous nature of the relative status score, no dyad in the data has a relative status score of exactly 0.

The principal component analysis in step three above is conducted on the sample of $461,720$ dyads that have no missing values for any of the nine status difference z-scores (in other words, dyads for which neither node has any missing values). In step four above, I compute the relative status score also for the remaining $74,836$ dyads in order to avoid dropping them from any analysis that uses the relative status score. This follows the same intuition that may lead a researcher intending to construct an un-weighed score or index to compute it as the observation-level average over non-missing elements, which implies re-scaling their weights. Applying this approach to a weighed linear combination using first principal component weights requires taking into account the fact that it is their squares that sum to 1, not their levels. Thus, for dyads with one or more missing elements, I re-scale each non-missing element e's weight w_e to $w_e^{'} = \sqrt{w_e^2 / \sum_{k=1}^{K} w_k^2}$, with K the dyad's number of non-missing elements.

The relative status score thus constructed on the full sample of dyads (potential links) is symmetric around 0, with standard deviation 1.452 and maximum 9.593. Its distribution (Figure 5, repeated in Appendix 3, Figure A4, panel A) looks near identical to that using only the sample of dyads contributing to the principal component analysis, i.e. when dropping dyads with missing values among the status elements (Appendix 3, Figure A3, panel B).

Appendix 2: Link formation and link reporting

Treatment assignment $T_i = T_j = 0$

We expect a similar pattern of more reporting of links by the relatively lower status individual as we observed for the baseline. Given our assumptions, the intervention does not affect linking likelihoods in this case; it also does not affect reporting. Thus pairs in this group remain the comparison group.

Treatment assignment $T_i = 0$, $T_j = 1$

For pair types A and B, the intervention has no effect on the linking likelihood and no links are formed; consequently, no links are reported by either node. For pair type C, the intervention has a weakly positive effect on the linking likelihood, which is driven by relatively higher baseline-status untreated nodes who now initiate (and subsequently report) links with relatively lower baseline-status treated nodes. As the utility to the latter from this type of link is reduced by the intervention, they may, on the other hand, not report these links. While not all newly formed links may be reported by the untreated node, all newly reported links are newly formed links (because in the absence of the intervention no links of this type exist).

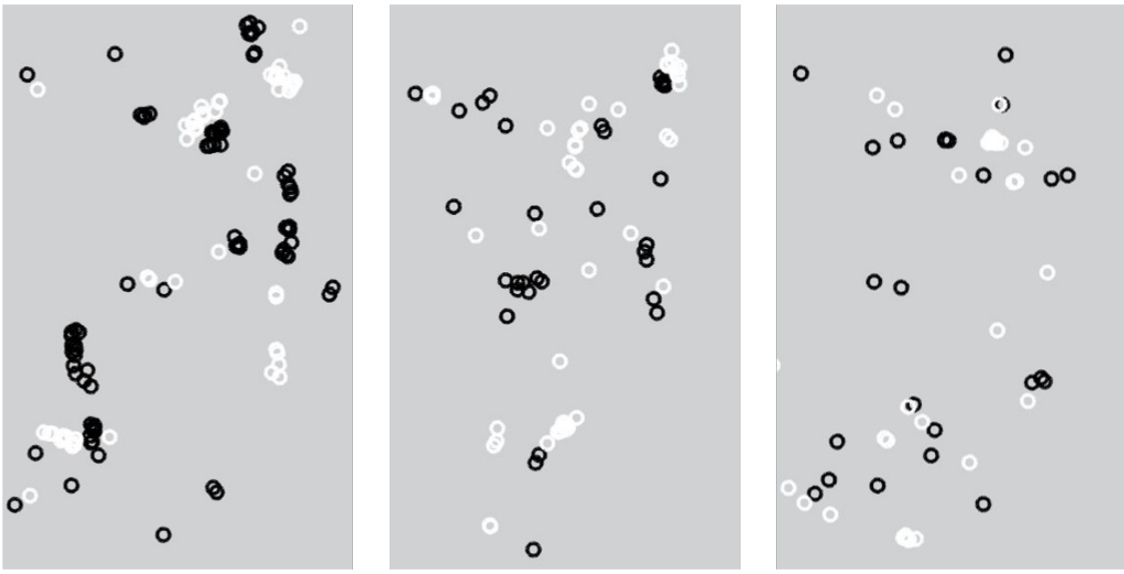
For pair type D on the other hand, changes in link reporting can be driven by both changes in link formation (weakly fewer links formed) or by changes in reporting of existing links (fewer of the existing links reported by the treated node as their utility weakly decreases even for nodes it keeps). Links of type D are those where the desire to link of both nodes is not driven by status differentials but by high other benefits and/or low costs; the intervention may then affect linking and reporting through the competition cost c_{ji}, but not through b_{ij} and b_{ji}. If there is a competition cost to the treated node, then the intervention has a weakly negative effect on the linking likelihood for this pair type; none of the dropped links will be reported by the treated or untreated nodes, but the untreated node will in addition not report some of the nodes that do get formed. If the negative effect on the utility of the treated node is small enough, there may be no effect on link formation even though we see reduced reporting by the treated node.

Treatment assignment $T_i = T_j = 1$

For pairs of type D, there is no effect on the linking likelihood (they get formed regardless of the intervention). But there may be an effect on reporting: because the utility from linking weakly increases for both nodes, they are more likely to report it. For pairs of types A to C the intervention may cause the formation of links that would not otherwise have been formed, and not all formed links may be reported. In particular, for pairs of type B/C that form a link, it is more likely to be reported by the relatively lower status individual, as the relative status was not affected by the intervention. To summarize, for the case that both nodes are treated, newly reported nodes may be either newly formed nodes (types A-C) or existing nodes that are newly reported on (type D). However, in all cases, an increase in reporting unambiguously signals an increase in individual utility from a given link type.

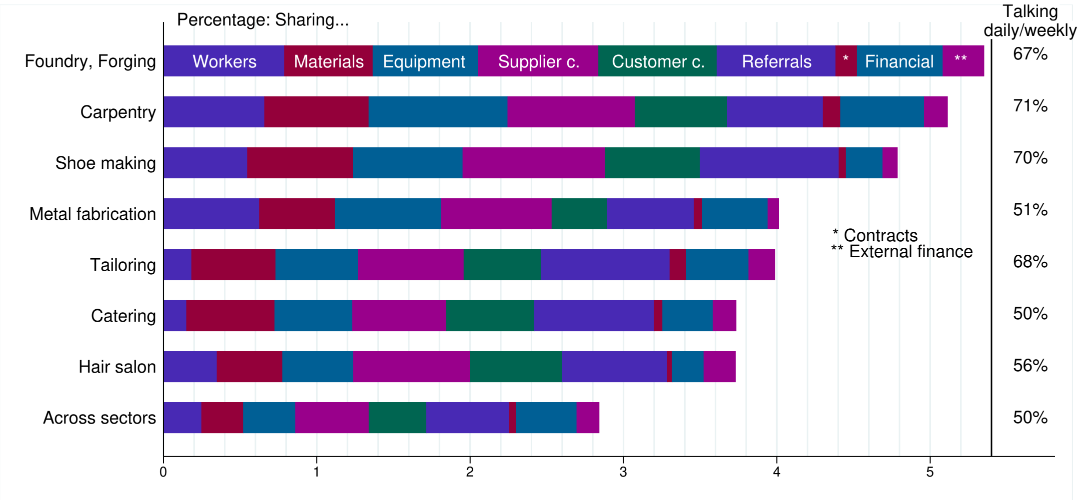
Appendix 3: Additional figures and tables

Figure A1: **Clustered treatment assignment by sector**



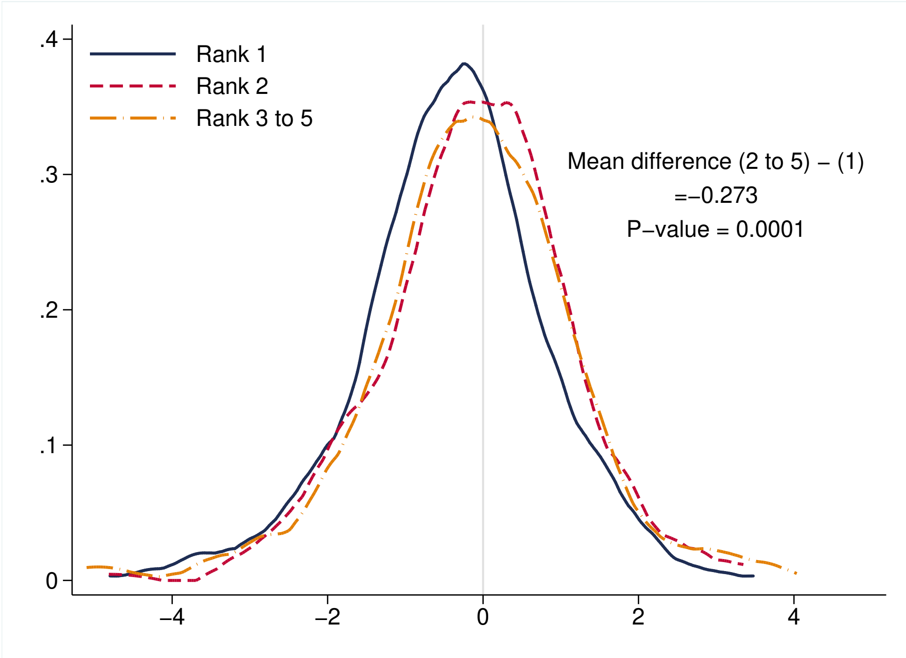
Note: These three panels show the same area within Kampala where a sizeable share of firms in the sample are located; its area is about 560 by 900 meters. The first panel shows metal fabricators, the second caterers and the third all other sectors present in the area combined. Firms were assigned to within-sector clusters based on a perpetuated 20 meter rule (see main text). This rule is illustrated here by drawing a circle of 10 meter radius around each firm's baseline GPS-coordinate-based location. If the circles drawn around two firms from the same sector intersect, they are assigned to the same cluster; this rule is repeated until there are no further intersections with any of the circles drawn around firms from the same sector. Treatment assignment is at the cluster level. Treatment clusters are shown in black, while control clusters are shown in white.

Figure A2: **Business collaboration and information exchange, by sector (baseline)**



Note: Link-level business collaboration along nine dimensions and frequency of information exchange, by sector. The first seven bars are for within-sector links (714 out of 762), the eighth is for across-sector links. The segment width represents the share of source nodes reporting collaboration in the given dimension in the past 12 months (hence the maximum segment width is 1). The total bar width represents the average number of collaboration dimensions (out of nine). "Supplier c." stands for "Supplier contacts", "Customer c." for "Customer contacts", "Referrals" for "Referred customers to each other", and "Financial" for "Financial interaction (borrow/lend)" Number of observations: Foundry, Forging 71, Carpentry 53, Shoe making 42, Tailoring 175, Metal fabrication 225, Catering 73, Hair salons 63; Across sectors 48. Sectors Electrical and Machining are included in the Metal fabrication bar due to small number of observations (11 and 1 respectively). The right-most column indicates the share of links within group that share business information at least daily or weekly.

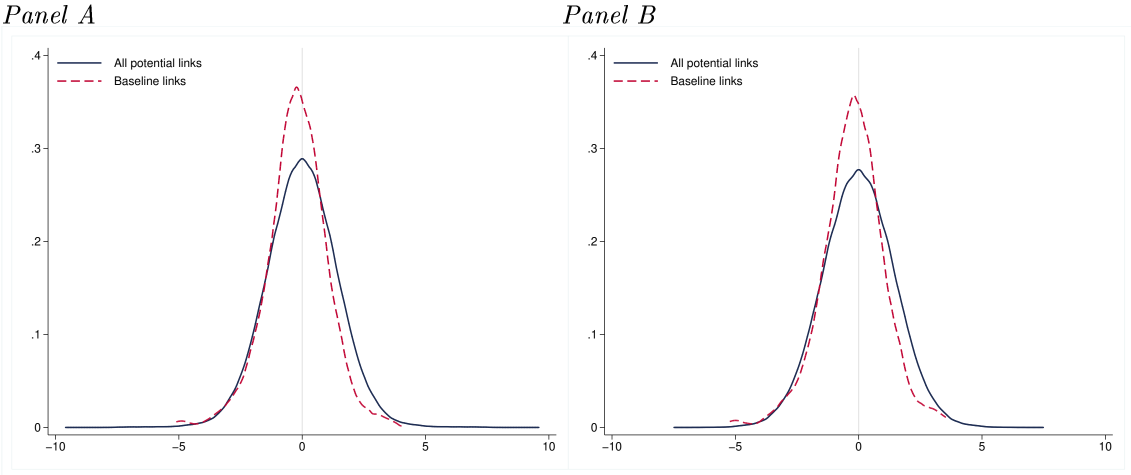
Figure A3: **Relative status score by reporting rank**



Note: Kernel density plot of the link-level relative status score, by rank in respondent's reporting order. Sample includes all 914 within-main-sample baseline links. The mean difference and its p-value are from a regression of the relative status score on a "rank 1"-dummy, using dyadic standard errors.

Figure A4: **Relative status score, all potential links vs. baseline links**

Panel A *Panel B*



Note: Kernel density plots of the link-level relative status score of potential links and baseline links; Panel A: all observations (536,556 observations for potential links, 914 for baseline links), Panel B: only observations used in the principal component analysis (461,720 for potential links, 786 for baseline links).

Table A1: **Dyad-level results: Differential effects by node relative status**

	Link at endline	Baseline links	New links
	(1)	(2)	(3)
Hdiff11	0.000056	-0.049312	0.000076
	(0.000123)	(0.107261)	(0.000109)
Hdiff10	0.000039	0.104372	-0.000031
	(0.000121)	(0.124134)	(0.000092)
Hdiff01	0.000419**	0.167107	0.000306**
	(0.000166)	(0.118150)	(0.000130)
Hsame1	0.044392***	0.227727**	0.039193***
	(0.007953)	(0.105724)	(0.007053)
Hsame0	0.036564***	0.084958	0.045829***
	(0.009040)	(0.106086)	(0.009395)
Ldiff11	0.000410**	0.246462*	0.000233
	(0.000195)	(0.131748)	(0.000146)
Ldiff10	0.000413**	0.149903	0.000319**
	(0.000172)	(0.103079)	(0.000144)
Ldiff01	0.000377**	0.247982**	0.000205*
	(0.000146)	(0.120410)	(0.000120)
Ldiff00	0.000377**	0.166706	0.000268**
	(0.000150)	(0.141116)	(0.000124)
Lsame1	0.060118***	0.224541**	0.055457***
	(0.009617)	(0.103355)	(0.008475)
Lsame0	0.050202***	0.145528	0.054763***
	(0.011482)	(0.108960)	(0.010397)
Baseline link	0.522701***		
	(0.022549)		
Constant	0.000618***	0.365576***	0.000749***
	(0.000169)	(0.088755)	(0.000135)
Controls	lc nlc	lc nlc	lc nlc
Ldiff11 - Ldiff00	0.000033	0.079756	-0.000035
...p-value	0.881	0.534	0.839
Ldiff10 - Ldiff00	0.000037	-0.016804	0.000051
...p-value	0.835	0.880	0.732
Ldiff01 - Ldiff00	0.000000	0.081276	-0.000063
...p-value	1.000	0.488	0.665
Ldiff11 - Hdiff11	0.000354	0.295774	0.000157
...p-value	0.047	0.002	0.261
Ldiff01 - Hdiff10	0.000338	0.143610	0.000236
...p-value	0.011	0.222	0.012
Ldiff10 - Hdiff01	-0.000006	-0.017204	0.000013
...p-value	0.974	0.875	0.927
Lsame1 - Lsame0	0.009916	0.079013	0.000693
...p-value	0.509	0.258	0.959
Hsame1 - Hsame0	0.007828	0.142769	-0.006636
...p-value	0.522	0.047	0.572
diff of diff	-0.002088	0.063756	-0.007329
...p-value	0.893	0.469	0.601
Comparison grp. avg.	0.000472	0.382353	0.000217
Observations	448,716	762	447,954

Note: *** p<0.01, ** p<0.05, * p<0.1. All models estimated by OLS. Dyadic standard errors in parentheses.

Table A2: **Construction of absolute and relative status score: Principal component analysis**

Panel A

Principal component (PC)	Eigenvalue	Total variance accounted for: Proportion	Cumulative
1	2.120	0.236	0.236
2	1.619	0.180	0.416
3	1.250	0.139	0.554
4	0.929	0.103	0.658
5	0.853	0.095	0.752
6	0.759	0.084	0.837
7	0.661	0.073	0.910
8	0.441	0.049	0.959
9	0.369	0.041	1

Panel B

Variable		1st PC scoring coefficient
Z-score of difference in:	number of employees	0.419
”	: past month revenues (IST)	0.397
”	: past month profits (IST)	0.327
”	: total assets (IST)	0.407
”	: years of schooling	0.214
”	: financial literacy	0.191
”	: technical knowledge	0.255
”	: owner age	0.342
”	: business age	0.361

Note: Number of observations: 461,720 dyads with no missing values in the elements of the status score.