When Proximity May Not Be Destiny: The Role of Existing Relationships

Sharique Hasan
Rembrand Koning

Working Paper 17-110
When Proximity May Not Be Destiny: The Role of Existing Relationships

Sharique Hasan
Stanford Graduate School of Business

Rembrand Koning
Harvard Business School

Working Paper 17-110
When Proximity May Not Be Destiny:
The Role of Existing Relationships *

Sharique Hasan
Stanford Graduate School of Business

Rembrand Koning
Harvard Business School

June 14, 2017

Abstract

Research on geography and knowledge spillovers is premised on the proposition that proximity reduces the cost of search and coordination. Thus, learning from proximate parties is easier than from more distant ones. As a consequence, nearby individuals, teams, and firms share overlapping knowledge and correlated outcomes. In this paper we theorize that spatial spillovers fundamentally depend on the presence of existing relationships. Using multi-dimensional network formation data from the random placement of teams at a startup bootcamp, we show that spatial spillovers decline if team members have existing ties within a particular social setting. For teams with preexisting ties within the bootcamp, localized spillovers appear small or non-existent. For teams without preexisting ties we find that outcomes improve if neighbors are high performing, but that outcomes worsen if neighboring teams are low performing. Our findings suggest that existing relationships do affect spillovers, primarily by capping downsides, but also by limiting the upsides of being near a high-performing team.

*Author ordering is alphabetical. Please direct correspondence to either sharique@stanford.edu or rem@hbs.edu.
This research has received funding and support from The Indian Software Product Industry Roundtable (iSPIRT), Stanford’s SEED Center, Stanford GSB, and the Kauffman Foundation.
Introduction

There has been a resurgence of interest in how microgeography—the location of individuals, teams and organizations in physical space—affects knowledge spillovers and performance outcomes. Building on the seminal work of Allen and Cohen (1969), research in a variety of contexts finds evidence of substantial localized spillovers. Catalini (2017) and Boudreau et al. (2017) report that proximity increases collaboration among scientists and, thus, increases knowledge transfer. Similarly, Murata et al. (2014) and Agrawal, Galasso and Oettl (2016) find strong evidence for localized spillovers among technologically intensive firms. Most existing models of geographic spillovers assume that proximity reduces search and coordination costs and, thus, increases information flow between adjacent parties, with the strongest spillovers occurring among those who share backgrounds, identities, or prior knowledge (e.g., Liu and Srivastava, 2015). Conventional wisdom, therefore, suggests that increasing proximity promotes spillovers. However, other concurrent social processes, such as an individual’s existing set of social relationships—e.g., prior friendships, work relations, and acquaintanceships—may alter the impact of proximity. In this study, we argue that these existing relationships limit when geographic spillovers occur. Specifically, we theorize, and then experimentally test, that the presence of existing network ties change an individual’s incentives to form relationships with neighbors and so breaks the correlation between proximity and performance.

While the literature on geographic knowledge spillovers and performance evokes network mechanisms, these processes often are implied rather than explicit (e.g., Stuart and Sorenson, 2003; Dahl and Sorenson, 2012; Agrawal, Galasso and Oettl, 2016). In this paper, we extend models of localized spillovers by explicitly considering their underlying network processes, focusing especially on the role of prior relationships in shaping the decision-making underlying new relationship formation. We draw on a rich network literature on the emergence and functioning of relationships that channel information; information that in turn shapes performance (e.g., Coleman, Katz and Menzel, 1957; Burt, 1980; Granovetter, 1985; Burt, 2004). In particular, a substantial body of theory argues that factors related to prior relationships, such as obligations, conformity, and limited access to non-redundant information, can curtail social action (Burt, 1992; Podolny and Baron, 1997; Krackhardt, 1999). Work by Gargiulo and Benassi (2000), for example, shows that preexisting relationships reduce the flexibility of managers placed in new and highly uncertain situations. Kim, Oh and Swaminathan (2006) introduce the idea of network
**inertia** which posits that prior relationships entail switching costs that limit an organization’s ability to dissolve established relationships as well as create new ones. In short, preexisting ties matter for future relationship formation.

Building on these insights we propose a simple actor-based model of network formation (e.g., Snijders, Van de Bunt and Steglich, 2010) that illustrates how preexisting ties impact geographic spillovers. We start with a well-established assumption that individuals have finite “social budgets” that restrict how many network connections they can maintain or how strong new connections are. (Granovetter, 1973; Hill and Dunbar, 2003). Given this social budget, an individual who meets someone new faces a relationship investment problem: Does she invest in a new relationship now or save her social capital in the hopes of establishing a potentially more valuable connection in the future? Propinquity structures this underlying search process; an individual is more likely to run into, meet, and socialize with their neighbors first before others who are farther away (Allen and Cohen, 1969). Some no doubt will skip over their neighbors in the hopes of a better “relationship return.” However, for the average individual the process of sequential relationship investment should lead to a greater prevalence of local relationships than distant ones.

At the heart of our model is the prediction that preexisting ties not only reduce the overall rate of new tie formation by depleting an actor’s social budget, but also disproportionately curb the impact of distance on relationship formation. These effects stem from the inertial, repeated and embedded nature of preexisting ties that greatly reduces uncertainty around their payoffs (Granovetter, 1985; Gulati, 1999; Dahlander and McFarland, 2013). Viewed in the language of relationship investment, existing ties provide known payoffs, especially when compared to the uncertain value of new relationships. We show that an actor who has “social insurance” from preexisting relationships will be more likely to skip over her neighbors and invest more of her finite social budget in more distant, riskier, but potentially more valuable connections. Moreover, our model suggests the existence of an important scope condition: if preexisting ties don’t reduce an actor’s social budget meaningfully, the impact of preexisting ties should diminish. Given that actor’s can form more weak ties than strong ties, we expect that the impact of preexisting ties on tie formation will be stronger for high-bandwidth strong ties than for low-bandwidth weak ties (Granovetter, 1973; Aral and Van Alstyne, 2011).

Furthermore, building on a substantial literature linking network ties to knowledge transfer and team performance (Argote, McEvily and Reagans, 2003; Reagans and McEvily, 2003),
we argue that our model of a spatial network formation not only describes who forms ties with whom, but also predicts the spatial distribution of performance. Having preexisting ties should insulate an actor from the performance of her neighbors. In contrast, actors with few existing relationships should be more likely to form ties with their neighbors and so similar performance outcomes.

To test our model we use data from a startup bootcamp for software entrepreneurs in India. Within the camp, team composition as well as team location was randomized allowing us to rule out bias from endogenous spatial positioning, a common concern in studies of both micro- and macro-geography (Chown and Liu, 2015). During the first week of the camp the participants were randomly assigned to a team of two or three people and then randomly assigned to one of 40 workstations in a large open-office space where they developed a prototype of a mobile app. Participants also were free to interact informally with and seek advice from other participants on a range of topics both pertinent and not pertinent to the design task. Participants communicated with one another using the bootcamp’s email system and Facebook group, generating digital traces of their communication patterns. Further, during the week participants used the bootcamp’s Google Apps account to develop a prototype of their product idea, enabling us to track revisions to their prototypes over time. At the end of the week, teams submitted a mock-up of their software prototype, a splash page, a paragraph describing the proposed software, and a one sentence description of their business idea. On Saturday morning, participants completed a series of surveys in which they provided information about their social networks, e.g., individuals that they got to know, sought advice from, became friends with, etc. Further, using a randomized double-blind process, each participant individually evaluated five submissions on several dimensions related to the quality of the idea, prototype, its novelty and business potential.

The results of our empirical analysis support the predictions from our theory. Consistent with our model, while geographic proximity increases the rate of tie formation—for knowing, advice, friendship, and electronic messaging—individuals possessing preexisting ties to others within the bootcamp are less likely to form new ties to proximate others. The primary exception, consistent with our model, is the formation of weaker “knowing” ties that are largely invariant to the presence of existing ties. Further, we find evidence that our model explains the spatial distribution of team performance. Teams with members with preexisting ties have performance outcomes that are uncorrelated with their neighbors. Those teams without existing relationships
to others at the bootcamp appear to acquire information from nearby teams on how to develop their prototypes, how to describe their product, and even appear more likely to work on their prototypes at the same time. These spillovers are double-edged: poor quality neighbors reduce a focal team’s quality and high-quality teams improve quality in some dimensions and not others. Overall, our findings suggest that existing relationships affect spillovers, primarily by capping downsides, but also limiting the upsides of fortuitously working near a high performing team.

Our findings contribute to the debates about the promise of microgeography as a strategic tool both within and across organizations. Within organizations, prior research has suggested that spatial positioning of employees can be used to increase competition between teams (Chan, Li and Pierce, 2014a), promote learning and innovation across teams (Chan, Li and Pierce, 2014b; Catalini, 2017), and potentially increase the chance of integration post-merger (Allatta and Singh, 2011). However, our results highlight how prior social ties may limit a manager’s ability to engineer new strong ties using propinquity alone. Across organizations, recent work has documented a trend towards co-working spaces, accelerators, and incubators that offer startups co-location with one another, often with the explicit promise of knowledge spillovers (Cohen and Hochberg, 2014). Building on our findings, when startup ecosystems are nascent with few preexisting ties, these organizations are likely to foster new connections and promote knowledge transfer. However, as ties within each community develop over time, their impact on creating new connections is likely to diminish.

**Theory and Hypotheses**

Conventional models of geographic spillovers are premised on a simple idea: proximity is a cost reduction device. Being close reduces the cost of finding potential connections—*search costs* (e.g., Boudreau et al., 2017)—and decreases frictions in interaction and coordination—*coordination costs* (e.g., Catalini, 2017; Allen and Cohen, 1969). The drop in these costs increases the likelihood that a social relationship will form between two parties. The lower the coordination costs, the higher the likelihood of interaction and, therefore, the higher the likelihood of a stronger relationship that can serve as a readily available source of information and knowledge. This information, when higher in quality, increases performance; lower quality decreases it. As a consequence, the outcomes of proximate parties become correlated. Next, we articulate this baseline model further, and then introduce into it our argument that preexisting
Microgeography, Networks and Spillovers

The development of a network tie connecting two individuals $A$ and $B$ begins with the simple fact of knowing a potential contact (e.g., $B$) exists. Building the consideration set of potential contacts is a limiting factor in network formation. People who are proximate will be more likely to know each other, learn about each other’s qualities, and as a consequence be more likely to form a social relationship that can serve as a conduit for information. In contrast, those more geographically distant are often unknown and, thus, not even in the consideration set. Boudreau et al. (2017) finds that scientists within the same disciplinary area at the same university often do not know each other. They show that merely placing two individuals in a room for a presentation increases the likelihood of co-authoring a grant. The greatest increase occurs for those in the same discipline. Thus, as proximity increases, individuals become members of a consideration set of potential contacts, and the likelihood of forming a network connection with them increases.

The formation of a consideration set is the first stage linking proximity to the formation of a new network tie. A second cost that proximity reduces are the costs of interaction and coordination (Catalini, 2017). Although an individual may know that a distant party exists, the likelihood of both serendipitous as well as planned interaction with them will be lower. Because proximate parties occupy the same geographic area, they will “bump” into each other more often, creating both familiarity as well as increasing the volume of information that they exchange through casual conversation. Further, planning for interactions with more distant connections is costly in terms of process. Interacting with a more distant connection requires greater effort in coordinating where and when to meet as well as potentially specifying a reason to meet. In contrast, if a focal individual has a question for a neighbor she can “walk in” and ask it with fewer formalities. That is, the costs for both parties are lower. Thus, proximity also increases the likelihood of a relationship forming, especially relationships with higher bandwidth.

**Hypothesis 1** Individuals are more likely to form a social relationship with those more geographically proximate to them.

Tie formation with neighbors creates a channel for information and knowledge transfer. This conduit can convey two primary sorts of information: benchmarks regarding performance and
knowledge that serves as inputs to performance. A substantial body of work in management and sociology demonstrates that network alters function as “reference groups” that anchor expectations about appropriate behaviors and desirable outcomes (Lawrence, 2006). For instance, Cohn et al. (2014) show that reference groups significantly affect effort provision among sales employees. Similarly, in their study of performance feedback, Kuhnen and Tymula (2012) find that having a reference group ranked higher than themselves increases performance by intensifying competitive effects. By promoting both input intensity, i.e., effort and time spent on production, and output quality, e.g., defining acceptable quality or performance, proximate parties will perform at similar levels.

A second mechanism leading to similar levels of performance among proximate parties is knowledge transfer and learning (Argote, McEvily and Reagans, 2003; Tsai, 2001). Neighbors have stocks of knowledge that vary in terms of both quantity and quality. This knowledge could include specific technologies, e.g., a programming language, visual design software, or statistical technique, or insights into specific processes, consumer insights, and other business related know how. In addition to possessing a higher volume of knowledge, neighbors may vary in the quality of information; some neighbors may have greater expertise in these areas than others (Hasan and Bagde, 2013). Moreover, a higher quality neighbor should be able to provide better feedback on the intermediate outputs of the focal party (Wooten and Ulrich, 2017). Conversely, a neighbor with lower stocks of knowledge or poorer quality skills will be an impediment to higher performance.

Together, performance benchmarking and knowledge transfer mechanisms should lead to a greater degree of correlated outcomes for parties that are geographically proximate. Thus, in the context of our startup bootcamp neighboring teams are likely to perform similarly.

**Hypothesis 2** The performance of a team is more correlated to those of proximate teams than to distant ones.

However, this flow of information (both positive and negative) may be disrupted if other frictions, besides search and coordination costs, inhibit the positive effect of geographic proximity on relationship formation.
The Impact of Preexisting Relationships

According to our theory of spatial spillovers outlined above, the primary frictions in interparty information exchange stem from the costs imposed by distance. Other frictions postulated by current theory arise from the heterogeneity in the traits, backgrounds and status of the parties involved (e.g., McPherson, Smith-Lovin and Cook, 2001; Liu and Srivastava, 2015; Chown and Liu, 2015). We argue that having existing relationships in a specific context counteracts the reduced costs of interaction owing to proximity.

In particular, existing relationships entail trade-offs that minimize the likelihood of forming new ties, even ones with proximate neighbors. First and foremost, existing relationships drain an individual’s social budget. Existing relationships involve social commitments (e.g., lunches, meetings, and other informal interactions) that reduce the amount of time and attention available to meet new people and form new relationships (Granovetter, 1973). Furthermore, because seeking information from others with whom one has built trust already requires less effort, individuals are more likely to exploit existing connections instead of exploring new ones (Gargiulo and Benassi, 2000). In addition to being costly in terms of time and effort, new relations also have uncertain payoffs. An individual deciding to invest time forming a new relation must determine whether or not the alter is high-quality, helpful and trustworthy. As a consequence, individuals with existing networks are less likely to form new ties, be they to neighbors or to more distant individuals:

Hypothesis 3  Individuals with preexisting relationships are less likely to form new ties.

Not only do preexisting ties reduce the rate of new tie formation, but we argue that having existing relationships disproportionately reduces the chance that an individual forms ties to her immediate neighbors. A large scholarly literature highlights how an individual’s preexisting ties provide stable and known payoffs, be they monetary or social (Uzzi, 1996; Dahlander and McFarland, 2013; Hasan and Bagde, 2015). In many ways, these ties can be considered a form of social insurance, providing value no matter whether any new relationship she invests in succeeds or fails. Indeed, a growing body of research finds that network ties can serve as a form of insurance in lending markets (Fafchamps and Gubert, 2007; Karlan et al., 2009).

This networks-as-insurance view implies that an individual with preexisting ties should be less likely to form connections with neighbors. Assuming individuals first learn about the value available from neighbors, the decision to skip over one’s neighbors can be thought of as a bet that
a distant and still uncertain future relationship may provide more value. Given that individuals with preexisting relationships have a form of “social insurance,” we expect that they will be more likely to take a risk and skip over their neighbors in the hopes of making especially valuable connections in the future. Individuals with preexisting ties have a safety net that enables them to ignore neighbors in the hopes of finding more valuable, and more distant, social connections.

Our argument has parallels in more formal theories of investment decisions under uncertainty (Markowitz, 1952). In Appendix A we develop a simple formal model of relationship formation to test the logic of our argumentation more rigorously. The formal model, which treats relationship formation as a two-stage investment problem between known neighbors and unknown distant partners, and the arguments above lead to the following prediction.

**Hypothesis 4** Individuals with preexisting relationships are less likely to form a social relationship with proximate peers than individuals with no preexisting relationships.

The mathematical and informal logic underlying Hypotheses 3 and 4 rests on the idea that individuals face a social budget constraint that drives how they make future connections. In the case of Hypothesis 3, we would expect that as an individual’s social budget grows the impact of preexisting ties on relationship formation decreases. In the case of Hypothesis 4, an individual with a larger social budget must no longer skip over neighbors. She can invest in both riskier distant relationships and relationships with her neighbors. Both arguments are formally illustrated in Appendix A. To assess how changes in an individual’s budget shape our predictions, we rely on variation in how weak or strong each tie is. Individuals have a much greater budget for weak ties, like knowing someone, than they have for strong ties, like friendship or wanting to co-found a company with someone (Granovetter, 1973; Hill and Dunbar, 2003). Variation in tie strength indicates variation in social budgets which we use to identify this effect.

**Hypothesis 5** The predictions made in Hypotheses 3 and 4 will be stronger for strong-ties, and weaker for weak ties.

Notice that this hypothesis does not necessarily imply that spatial distance will not matter for the formation of weak ties. If local ties are less costly and/or risky, an individual will still prefer to make more ties to neighbors even as her social budget increases. Appendix A shows that a social budget increase pushes individuals with preexisting ties to connect again with neighbors. Thus the main effect of distance should remain even for weaker ties, but the effects
Finally, we build on the benchmarking and knowledge transfer arguments used in developing Hypothesis 2 to link the presence of preexisting ties to the spatial distribution of performance. If individuals with preexisting ties are less likely to connect with their neighbors, they also are less likely to develop the local channels needed for benchmarking and knowledge transfer. Therefore, someone with preexisting ties or this individual’s team is less likely to learn from neighboring stocks of knowledge or benchmark against the effort of their neighbors. The lack of local ties may be detrimental to performance if neighbors have high quality skills and deep stocks of knowledge. In contrast, scarce local ties may be beneficial if neighbors have few skills and limited knowledge. Regardless, our model predicts a decrease in the spatial correlation of a team’s performance when the team is composed of agents with more preexisting social ties.

**Hypothesis 6** The performance of a team with few preexisting ties will be more correlated to proximate teams than teams with more preexisting ties.

**Research Design**

To investigate the effects of distance and preexisting ties on relationship formation and the spatial distribution of team performance, we conducted research at a three-week long startup bootcamp. This camp helped 112 aspiring Indian entrepreneurs learn to generate ideas, develop prototypes, and test business models. The demographic background of our participants is similar to bootcamps, incubators, and accelerators in the US. The program consisted of three week-long modules that mixed hands-on practice with instruction from successful Indian entrepreneurs, designers, and venture capitalists. The program required full-time attendance and operated Monday through Saturday.

We focus on experimental interventions deployed and data collected during the first week of the program. In the initial week participants were randomly assigned into teams and, after

---

1 The participants were young ranging from 18 to 36, with a mean age just over 22. The participants all had a college degree or were enrolled in college, 8 had advanced degrees including PhDs in Electrical Engineering, and nearly 70% had degrees in technical fields. Unlike many accelerators in the US, the program had a sizeable number, 25, of female participants.

2 The second and third week were similar to the first, but increasingly less structured. During the second week, we again placed participants into teams of three, but instead of focusing on product development each team applied lean startup frameworks to validate their app’s business model. The best teams again won prizes similar in size to the first week’s prizes. The final and third week was the least structured; participants self-selected their teams, used any framework they wished and competed to receive up to 8,000 USD in funding and support to implement their startup idea. We also embedded experimental intervention during the second week. The third week involved no designed...
an orientation day on Monday, spent the rest of the week working in these teams to develop a mobile app prototype while learning design-thinking frameworks. To ensure that prototypes were commensurable across teams, each group was assigned the task of developing an app for the Indian wedding industry. In addition to randomizing team composition, teams also were randomly assigned to workbenches distributed within the bootcamp’s large, open office space. On the Saturday concluding the first week, participants independently filled out network surveys and evaluated one another’s prototypes through the bootcamp’s learning management system. Participants also completed a network survey prior to the first day of the camp. To incentive effort and provide resources for the most promising teams, the three highest rated proposals and prototypes from this week won prizes totaling 45,000 Indian rupees (INR), that is, about 789.47 USD.

The first week of the program is selected for both theoretical and empirical reasons. Since our model studies network formation in new social contexts, it considers an actor with preexisting social ties who evaluates how to allocate her social budget between new neighbors or more distant others. The first week of the boot camp reflects just such a process: some participants bring with them preexisting social ties and all are forming new relationships within the bootcamp.

Empirically, the first week of the camp provides the concrete measures we need to test our team-level and individual-level hypotheses. The double-randomization of people to teams and teams to workbenches enables us to generate a measure of distance between any two participants (or teams) that is exogenous to individual qualities, abilities, and characteristics. At the individual level, we use the data from Week 1 to build a dyad-level person-to-person dataset to test (1) if relationships are less likely to form with more distant others, and (2) if this effect dissipates for individuals entering the camp with preexisting ties. At the team level, the data from Week 1 enables us to build a dyad-level, team-to-team dataset to check (1) if neighboring teams’ performances are correlated and (2) if this correlation weakens if team members’ have experimental interventions.

3We chose the Indian wedding industry for three reasons. First, in conversations with Indian entrepreneurs and venture capitalists, we learned that the Indian wedding industry is large and has significant market potential with several venture capital firms actively investing in the space. Second, the “Indian wedding” was something that the all of the participants had some experience with, but it was an industry where a subset of individuals were unlikely to have insurmountable knowledge advantages. Third, we chose this industry because while focused it is still a relatively diverse space, with problems ranging from finding a spouse to buying wedding dresses to honeymoon selection and even post-marital counseling. This focus helped ensure that the prototypes generated by each team were comparable and could be evaluated against one another.

4The first prize was 20,000 INR, the second was 10,000 INR, and the third was 7,500 INR.
more preexisting ties.

**Individual-level Data**

To test how distance and preexisting ties shape relationship formation, we build a person-to-person, dyad-level dataset with 12,222 inter-team dyad observations. Each row in this dataset represents a potential relationship between ego $i$ and alter $j$. While the entire set of potential directed dyads for the 112 participants is 12,432 ($112 \times 111$), we exclude the 210 within-team dyads because team members were required to interact and so tie formation is not merely a function of distance. After excluding the within-team dyads, this dyad-level dataset is suitable for testing if individuals are more likely to form ties to neighboring others without concern that team assignment is driving our results. Relatedly, given that individuals are assigned to teams, and teams to a location, our data suffers from within-team non-independence. We account for this problem in our models by clustering at the team level. Moreover, to account for the non-independence present in all dyad-level data, we robustly cluster our standard errors at the level of an ego’s team, the alter’s team, and at the team-dyad level (Cameron and Miller, 2015).

**Randomized Distance** During the first week of the bootcamp, participants were randomly assigned into thirty-two teams of three people and eight teams of two people. These teams were then randomly assigned to one of forty workstations within the bootcamp’s open office space (see Figure 1). Teams were required to work at these assigned workstations starting on Tuesday through end of day Friday. Due to this double randomization, we can construct a measure of distance that is exogenous to each participant’s qualities, abilities, and characteristics. Given that endogenous location choices are rarely random but involve clustering on both observed and unobserved variables, having an exogenous measure of distance is critical. Non-random spatial clustering would lead to severe selection bias and obscure our ability to determine whether distance is the underlying mechanism responsible for our results.

![Figure 1 about here.](image)

We create an indicator of distance between teams by first measuring the shortest walkable route between each table. Since the space had a large wall in the middle (see Figure 1) that prevented interaction between the teams on either side, we include this single barrier in our

---

5Results are substantively unchanged when clustering at the individual ego, alter and dyad level.
distance calculations, assuming participants have to walk around the wall to interact with teams on the other side. Other than this wall, the space was entirely free of physical obstacles that would prevent proximate individuals from interacting. The walking distance between the tables ranges from 89cm (3ft) to 2,627cm (86ft). Allen and Cohen (1969) seminal work on micro-geography in R&D labs has found that the effects of distance decay exponentially. Therefore, for linear models the correct measure is not the raw distance, but the log of distance. This gives us our randomized distance measure $\log(Distance\ between\ i\ and\ j)$. This measure captures both the distance between teams and between individuals on those teams. Table 1 presents dyad-level summary statistics for our distance measure along with the other individual-level measures described below.

[Table 1 about here.]

**Preexisting Ties** Before the bootcamp, each participant identified the other participants he or she knew in a roster-based network survey (see the next subsection for a description of roster interface). Using this data we construct a dichotomous variable, *i has preexisting ties*, that equals 1 if the individual has preexisting ties within the bootcamp and is 0 otherwise. To confirm that individuals have at least one meaningful preexisting tie, we specify that a focal individual has a preexisting tie if the person has a connection to another bootcamp participant and that another individual indicates that he or she knows the focal participant. The presence of both outgoing and incoming connections increases the likelihood that preexisting tie(s) are more likely to be durable, sticky, and involve commitment, consistent with our theoretical model. Our measure indicates that 56 (50%) of the participants entered the camp with no preexisting ties.

To control for existing relationships, we also construct a dyad-level indicator *i knows j pre-bootcamp*. Table 1 shows that 50% of participants have preexisting ties but only 2% of the set of potential knowing relationships are present before the camp. Beyond improving our statistical power by accounting for pre-bootcamp heterogeneity, introducing *i knows j pre-bootcamp* allows us to parse out tie inertia from the impact of having preexisting ties on the formation of new relationships. There is little doubt that if *i knows j* they will continue knowing them. Our interest lies in how participants with preexisting ties form new relationships.

**Relationship Formation** To capture the multiplex nature of tie formation, we use five different relationship measures to triangulate the effects of spatial proximity on tie formation. Our
first four measures of tie-formation are derived from a roster-based network survey conducted on the Saturday that concluded the first week of the bootcamp. The survey presented each participant with the name and a photo of each of the other participants. Similar to social networking websites, the respondent selected whom they had a relationship with. Following standard practice in the social networks literature, they then selected who they know at the bootcamp, who they go to for advice, and who they consider a friend.

Complementing these standard questions, we collected two other types of network data. The first, called Future Team, was developed from the roster survey question that asked participants who they would like to have as part of their startup team in the future. One of the explicit goals of the camp was to provide a forum for the nascent entrepreneurs to meet possible co-founders. This question provides evidence for how the participants are searching for potential co-founders. Furthermore, the forward-looking nature of the question demonstrates that this measure captures strategic networking behavior. All of our roster-based measures are binary variables.

The final network dependent variable, Message, is a count of the email and Facebook activity between any two participants. Specifically, we sum the number of emails from $i$ to $j$ in our bootcamp email system, likes by $i$ of $j$’s posts in the bootcamp’s private Facebook group, and the comments by $i$ on $j$’s posts in the same group. Given that participants communicated over different digital platforms and through different means, we aggregate across the three types of “messaging” to create a less sparse and potentially more robust measure. Unlike the binary roster survey measures, this indicator is a count of the number of messages from $i$ to $j$.

**Team-level data**

To investigate how distance and preexisting ties shape the spatial distribution of team performance, we build a team-to-team dyadic-level dataset with 1,560 inter-team dyad observations. Each row in this dataset represents a focal team $i$ and an alter team $j$. This data represents the entire set of potential team-to-team pairings for the 40 Week 1 teams ($40 \times 39 = 1,560$). Using this data structure we can test if neighboring teams have performance outcomes that are more similar to one another than for teams that worked farther away from each another. We account for the non-independence present in the dyadic data by robustly clustering our standard errors at the ego-team, alter-team, and team-dyad level. Build-
ing on our individual-level data, we generate a team-level measure of preexisting ties, \( Team_i \) has preexisting ties (mean), by simply taking the teammate average within each team \( i \).

**Team Performance** We rely on a double-blinded peer evaluation process to generate measures of team performance. By midnight on Friday teams submitted a digital “packet” containing a one-sentence description of their software app, a longer one paragraph description of their product, a splash page and a slide deck walk-through describing how users would interact with their product. On Saturday morning, each participant rated submission packets for five randomly assigned teams (excluding their own team). Ratings were assigned to 12 project dimensions on a 5-point Likert scale: novelty, insightfulness, empathy with user, predicted demand, purchase intent of the rater, purchase intent of ideal user, feasibility of the product, business potential, prototype quality, splash page quality, paragraph quality, and sentence quality.\(^6\) Given there are 40 teams and each person evaluated 5 projects, we end up with 14 evaluations for each project. We construct an aggregate Team performance \( i \) measure by averaging across the 14 evaluations and then across the 12 dimensions. Our performance indicator ranges from a low of 2.44 to a high of 3.29 with a mean of 2.79.

To test if a focal team’s performance is correlated with their neighbors’ we follow the literature on knowledge spillovers and spatial agglomeration to construct an inverse distance-weighted performance measure for each of the other 39 teams. Specifically, we divide the Week 1 performance of a team \( j \) by the log-distance between team \( j \) and our focal team \( i \). If proximate teams influence each other’s performance, then this inverse distance-weighted measure, Team \( j \) ’s performance / log(Distance), should be positively correlated with the performance of the focal team \( i \). If there is no influence, then we should find little correlation between the two variables.\(^7\) Since we are interested in the spatial clustering of performance and not in the

---

\(^6\)We drop one dimension of evaluation, the quality of the product’s name, from this analysis. While the other dimensions are largely correlated with one another, the quality of the name is both less correlated with the other dimensions and exhibits much less internal rating consistency. Nevertheless, results are very similar in magnitude when using all 13 dimensions, though the statistical significance is somewhat weakened, dropping from the 5% level to the 10%.

\(^7\)As we have argued, the randomization of teams to tables implies that any observed correlation in performance between proximate teams emerges because of spillovers and not because of selection by the teams into spatial clusters. Furthermore, since the open office plan and workstations were designed to be similar to one another, we greatly dampen the possibility that environmental differences (quieter workspace, more light, etc...) explain our results. Compared to past work that has examined spatial effects at the micro-level, our design takes place in one open room and so greatly reduces the chance that environmental confounds are responsible for our results. Finally, our estimates do suffer from a reflection problem. Since we use the Week 1 performance of peer teams to predict a focal team’s performance, we are limited in our ability to identify the underlying spillover mechanism as emerging from
absolute level of performance, we are less concerned about evaluation biases or errors. To unduly affect our estimates, the evaluation bias or error must correlate with our randomized distance measure. Critically, we find little evidence that distance predicts how an individual evaluates a team. Thus, we believe our measure, though imperfect, provides us with a proxy that can test for spatial clustering within the bootcamp if not absolute performance differences. Table 2 displays summary statistics for our team level measures at the dyad level.

[Table 2 about here.]

**Prototype Development** To study spillovers we also analyze spatial dependence across a particularly salient channel: prototype development. The primary deliverable for each team at the end of the first week was a Google Slides deck demonstrating how a user would interact with their product and a splash page explaining their business and app. These sorts of wireframe mockups are common in the design-thinking methodology, and often are used by product managers and designers in industry. Thus, the participants spent most of their time during the first week prototyping a mobile app and corresponding splash page. If teams are impacted by their neighbors we expect such spillovers to be strongest for prototype development.

We test for prototype development spillovers using three different outcomes. The first two variables, the inverse distance-weighted measures *Prototype Score* and *Splash Page Score*, are individual indicators derived from the peer evaluation process that capture the evaluation of the prototype and related splash page. The third variable is constructed from the development of the prototype itself. Specifically, we use the revision data from the Google Slides documents to test if teams near one another are more likely to make edits at similar times during the day. Being more likely to edit at the same time as a near neighbor is consistent with the proposition that neighbors are interacting with one another and so potentially sharing information, providing feedback, and benchmarking against one another.

Our third outcome variable *Team i edited prototype* reflects the timing of when Google Slides revisions occurred. We generate a panel dataset that records if a team made a revision during each 30-minute period from the moment the documents were first generated on the Wednesday performance multipliers or from some other background characteristic. That said, the theoretical goal of our paper is to identify the presence of such spillovers and their dependence on existing networks. We hope future work explore the underlying mechanisms in more detail.
morning until the end of day Friday when submissions were due. Like our peer evaluation analysis, we also construct an Team j edited prototype / log(Distance) measure which is our revision indicator divided by log-distance. In contrast to our peer evaluation analysis, the panel structure of revision data allows us to include fixed effects for both the 30-minute time period and the focal team. The time period fixed effects control for variation in across-bootcamp activity levels, and the focal team fixed effects control for average differences in activity across teams. With these fixed effects our estimate is identified off variation in alter team editing within a given time period, a relatively demanding specification.

Results

Was the Spatial Randomization Successful?

Before turning to our six hypotheses, we check that our randomization was successful. We conduct a balance test that analyzes if observed pre-treatment characteristics are uncorrelated with treatment status. Given that our treatment is the randomized distance between any two participants, we check for balance using our individual-level dyadic data. Specifically, we regress the log(Distance between i and j) on our two pre-treatment variables, the indicator for if the focal actor i knows j pre-bootcamp and the indicator i has preexisting ties within the bootcamp overall. This regression finds no evidence for imbalance (see Table 3). Our randomization of individuals to teams, and teams to workbenches, suggests that distances are uncorrelated with the existing social networks of the participants.

[Table 3 about here.]

Does Randomized Distance Matter for Tie Formation?

We test our hypotheses sequentially starting with Hypothesis 1 which says that individuals are more likely to form a social relationship with those more geographically proximate to them. This test is essentially a replication analysis of the recent work that uses natural and designed randomization to show that exogenous distances at level of meters can shape scientific collaboration networks (Catalini, 2017; Boudreau et al., 2017). Like these studies, we regress our

---

All but four teams developed their prototype in a single Google Slides document over the course of the week. The four teams that did not develop in Google Slides made their presentations in Powerpoint and uploaded the deck to Google Drive at the end of the week. Thus, we do not have revision histories for these four teams.
network formation measures—Know, Advice, Friend, Future Team, Message—on our measure of log-distance between any two individuals $i$ and $j$. To make this test more robust and to ensure that our preexisting tie measure is not merely capturing a tendency to maintain existing ties, we include the dummy variable $i$ knows $j$ pre-bootcamp. If $i$ already knows $j$, the control captures this, thus parsing out existing relationships to isolate the effect of preexisting ties on the formation of new relationships and not the maintenance of old relationships.

[Table 4 about here.]

Table 4 reports estimates from logistic (Models 1-4) and quasi-poisson (Model 5) regressions that test Hypothesis 1. Estimates are the log-odds of tie-formation for each relationship type. Consistent with prior work on relationship decay and maintenance (Dahlander and McFarland, 2013), the coefficients for our control variable $i$ knows $j$ pre-bootcamp are positive, large, and highly statistically significant in all five models. The estimates of randomized distance on network formation offer evidence in support of our first hypothesis. In all models the coefficients on distance are negative indicating that as distance increases the probability of tie formation decreases. Our strongest evidence comes from Model 1 ($p < 0.05$), estimating who knows who, and Model 3, estimating who is friends with whom ($p < 0.01$). The effects on relationship formation are non-trivial. For example, in Model 3, we find that moving a person from a neighboring workbench to the farthest workbench possible reduces the probability of friendship formation by nearly 50%. In Model 1, a one-unit decrease in log-distance from the mean distance (a decrease in raw distance of 270%) increases the probability of $i$ knowing $j$ nearly 13% from a base rate of just under 16%. Critically, this effect multiplies as distance increases. Moving from being a neighbor to the farthest workbench possible reduces the probability that $i$ knows $j$ by just over 66%.

Does Randomized Distance Shape the Spatial Distribution of Performance?

If social ties are conduits for knowledge transfer and performance benchmarks, and social ties are spatially determined, then team performance should be spatially determined. To test if

---

9Given that our first four network formation measures are dichotomous, we use logistic regression; for our fifth measure, the count of digital messages sent from $i$ to $j$, we use a quasi-poisson regression to account for over-dispersion. As we described earlier, given that distances are randomized at the team level, we robustly cluster standard errors at the focal-team level, the alter-team level, and the team-dyad level.
the performance of a team is more correlated to the scores of proximate teams than to distant ones’ (H2), we regress a team’s own performance on the distance-weighted performance of the other teams at the bootcamp (see Table 5). However, regressing the outcome of one team on the outcome of other teams introduces a subtle negative bias due to the problem of exclusion (Caeyers and Fafchamps, 2016). This bias can be understood through a simple example. Imagine if our experiment only had 3 teams, one whose performance score is 1, one with a score of 0, and one with a low score of −1. If we consider the correlation between a focal team’s score and the scores of the remaining population we would find a negative correlation because the focal team is excluded and so the population tends to have a lower average score than then higher scoring focal team; or a higher average score than the lower scoring focal team. While the exclusion bias dissipates as the sample size grows, at the team level we have a relatively modest sample size at 40 teams, and thus are estimates are likely to be biased by this exclusion effect (Caeyers and Fafchamps, 2016). However, the bias works against the direction of our predicted relationship. Hypothesis 2 predicts a positive correlation between a team’s own performance and the inverse distance-weighted performance of other teams. Thus, while our estimated effects are likely downwardly biased, since the bias works against us, our estimates serve as lower bounds for the true effect.

Turning back Table 5, we again use dyadic data but unlike Table 4 the dyads are team-team observations and not individual-individual observations. Model 1 regresses a focal team’s performance on the team performance of each peer team. We find a small, but statistically significant, negative effect, consistent with our arguments concerning exclusion bias. Model 2 regresses team i’s performance on our inverse distance-weighted performance measure of team j’s performance. We find some evidence for spatial clustering of performance, though the effect is only significant at the 10% level. The coefficient implies that if my immediate neighbor moved from being the worst to the best performer, my team’s performance would be expected to rise approximately one-fifth of a standard deviation. In Model 3 we include our inverse distance-weighted measure and the raw peer performance measure to account for exclusion bias better when estimating our inverse distance-weighted measure. Doing so yields stronger evidence for spatial clustering. Both the coefficient and statistical significance increases to

\[ \text{effect size} = \frac{\text{lowest score} - \text{highest score}}{\log\text{distance}} \times \frac{1}{\text{standard deviation}} \]

\[ \text{effect size} = \frac{2.44 - 3.29}{4.49} \times \frac{1}{0.20} \]

\[ \text{effect size} \approx 0.196 \]

\[ \text{effect size} \approx (0.218 \times 2.29 - 2.44 \times 4.49) \times \frac{1}{0.20} \]

\[ \text{effect size} \approx 0.196 \]
0.387 and to the 5% level, respectively. The coefficient implies that the correlation between a focal team’s performance and that of a peer team drops by roughly 50% when the peer moves from a neighboring bench to the other side of the bootcamp’s space. Furthermore, since the denominator is logarithmic, the drop in correlation occurs most rapidly as two teams begin to become more distant. Taken together, the results in Tables 4 and 5 support Hypotheses 1 and 2 that the distance between teams shapes network formation and the subsequent knowledge spillovers.

Do Preexisting Ties Shape Network Formation in a Pattern Consistent with the Proposed Model?

Does the preexisting network matter? While Table 4 offers strong evidence that individual $i$ maintains preexisting ties, it does not reveal if preexisting ties shape how individual $i$ goes about forming new relationships. To test if new relationships are impacted by an individual’s preexisting ties, we build on the regressions in Table 4 by interacting our dummy variable $i$ has preexisting ties with our distance measure. This regression tests Hypothesis 3, that individuals with preexisting relationships are less likely to form new ties, and Hypothesis 4, that individuals with preexisting relationships are less likely to form ties to spatially proximate peers.

Table 6 reports the same set of regressions as in Table 4 but with our preexisting tie variable and its interaction with our distance measure. Across all five models we find that the main effect of preexisting ties is negative and statistically significant, providing support for Hypothesis 3 that individuals with preexisting ties are less likely to form new relationships. The coefficients of the interaction terms indicate positive effects in four of the five regression models. Since the main effects for distance are negative, these interaction estimates imply that the effects of distance on relationship formation are mitigated for individuals with preexisting ties. Consistent with our arguments that proximity has less impact on tie formation for those with preexisting ties, we find that the absolute size of the interaction coefficient nearly matches the absolute size of the main distance effect in all models. For advice, the main effect is -0.287 and the interaction is 0.332; for friendship, the main effect is -0.365 and the interaction is 0.250; for future team, the
main effect is -0.331 and the interaction is 0.365; and for messaging, the main effect is -0.060 and the interaction is 0.084. These results demonstrate that the distance effect essentially vanishes for individuals with preexisting ties.

Next we turn to Hypothesis 5, that the effects of preexisting ties on relationship formation are weaker for weak ties than strong ties. Model 1, which tests the effects of distance on relatively weak knowing ties, generates the least amount of support for Hypotheses 3 and 4. The coefficient for $i$ has preexisting ties is one-fourth the size of the equivalent coefficient for advice, friendship, and future team ties. Furthermore, this result is only significant at the 10% level and is significantly smaller than the equivalent advice coefficient. The same conclusion holds for the interaction term. The coefficient in the knowing model is roughly one-fourth the size in terms of magnitude and is statistically insignificant. To test for differences between strong and weak ties formally, we append the advice and knowing design matrices to one another, and then interact each of our variables with a dummy that is 1 if the observation reflects a weak knowing tie row and is 0 when the observation represents stronger advice ties. Again, we cluster at the ego team, alter team, and team dyad levels. This “stacked” regression (available upon request) indicates that both the main effect of preexisting ties and the interaction with distance are not just smaller but significantly so for the weak knowing ties compared to stronger advice ties ($p < 0.05$).

Do Preexisting Ties and Randomized Distance Better Describe the Spatial Distribution of Performance?

Again, if social ties are conduits for knowledge transfer and performance benchmarks, and social ties are determined by spatial position when preexisting networks are absent, then team performance should be spatially determined only when teams have limited preexisting networks. To test our final prediction, Hypothesis 6, we again regress a focal team’s performance on an alter team performance, inverse distance-weighted performance, but also include our team-level preexisting tie measure along with its interaction with the distance-weighted performance measure. If teams with members having preexisting ties are less likely to form relationships with neighbors, then the team’s performance should be less correlated with its neighbors.

(Table 7 about here.)

Model 1 in Table 7 displays the estimates from this regression. First, the control variables
show little evidence that teams with members who have preexisting ties perform better (or worse) than teams with fewer members with preexisting ties. We again find evidence for exclusion bias with a relatively small, but negative and significant, coefficient on Team $j$’s performance. Our distance-weighted measure is positive and statistically significant, implying that neighboring teams have correlated performance outcomes. Moreover, the coefficient of the interaction term is negative and statistically significant ($p < 0.05$). At -0.174 it is about roughly one-half the size of the estimated main distance-weighted alter performance effect. Furthermore, the standard errors on the interaction term and main distance-weighted peer performance effect reveal that the coefficient sizes are not statistically distinguishable from one another. Taken together, these results suggest that teams with members who have preexisting ties have performance outcomes that are less correlated with their neighbors.

Models 2-4 check for spillovers in the most salient activity of the first week, prototype development (see Table 7). Models 2 and 3 parallel the regression in Model 1, but instead test for spatial clustering in terms of a team’s disaggregated prototype score and splash page score. If anything, the results for these two prototyping-related measures are even stronger. Consistent with our model, teams appear to be sharing knowledge with neighbors about how to build and refine their prototypes, but only when the team members have few preexisting ties.

Paralleling the earlier regressions in Table 7 but with our edited prototype measures, we further check this working with neighbors mechanism in Model 4. Beyond serving as a “behavioral” check on our peer evaluation scores, the longitudinal nature of this data allows us to specify a more demanding regression. Specifically, we include both 30-minute time period and ego team fixed effects. The former controls for the fact that certain time periods have more activity (3pm) than others (4am). The latter controls for team-level differences in activity and identifies spillovers using variation in when spatial proximate teams are working while controlling for each team’s propensity to work overall.

Model 4 in Table 7 reports results from this regression using a linear probability model. Since we include team-level fixed effects, the main effect of preexisting is absorbed and so not estimable. Though insignificant the coefficient on Team $j$ edited prototype is small and negative, consistent with our exclusion bias arguments. In contrast, the coefficient of Team $j$ edited prototype /

---

12 Results are substantively unchanged when using different time period widths. However, as the periods begin to get longer than an hour, or shorter than 10 minutes, the estimates lose statistical significance as the wider and shorter periods wipe out the variation that identifies our effects.
**log(Distance)** is a positive coefficient significant at the 10% level. If a neighboring team makes an edit the focal team is roughly 12.5% (0.561/4.49) more likely to edit their prototype in the same period. The interaction term coefficient is negative and statistically significant effect ($p < 0.01$) that is roughly comparable in absolute magnitude at -0.769 compared to the main effect of 0.561. Thus, it appears that neighboring teams work at similar times, but that this clustering disappears when a team has more members with preexisting ties.

**Conclusion**

Our findings identify an important scope condition limiting the impact geography has on the outcomes of individuals and firms (e.g., Chown and Liu, 2015; Murata et al., 2014; Catalini, 2017; Agrawal, Galasso and Oettl, 2016; Boudreau et al., 2017). We argue for and empirically test the role that preexisting relationships play in the decision to form new relationships that channel knowledge and can result in geographically correlated performance (Kim, Oh and Swaminathan, 2006; Gargiulo and Benassi, 2000). Our theory suggests two reasons for the reduction in localized relationship formation: a reduced social budget and the role that preexisting ties play as insurance that incentivizes “holding off” for more distant, and potentially higher quality, relationships.

Our analysis supports these general claims and offers more nuanced insight into the process of geographic spillovers. First, we find evidence that teams whose members have preexisting relationships with others at the bootcamp are significantly less biased towards forming relationships with proximate neighbors. They appear to search more broadly and have a network that consists of more distant connections. Second, this dampening of localized tie formation exists primarily for “stronger” ties of friendship and advice, but not for weaker “knowing” ties of acquaintanceship. As a consequence, our findings reveal significantly reduced geographic spillovers for teams whose members have preexisting relationships. Conversely, teams with no preexisting relationships experience diminishing performance outcomes if they are located near low-performing teams and improved performance if they are near high-performing teams. For teams that experience spillovers, our results suggest that this is primarily driven by improvements in the visual appeal of a prototype, splash page and description of the product. These improvements may stem from either learning specific skills (e.g., design) or setting benchmarks on observable dimensions of performance, or both. Overall, our research suggests that existing
relationships do affect spillovers, primarily by capping downsides, but also limiting the upsides of fortuitously working near a high-performing team.

Before examining the implications of our findings for the study of geographic spillovers, networks and innovation, we discuss our study’s shortcomings that restrict how our findings may be generalized. One particularly important limitation is the setting: a one-week long product design competition at a startup bootcamp. Our site is limited by the transitory nature of teams, the narrower geographic area, and short timespan of the innovation task. Generalizing to more durable teams or organizations, a wider geographic span, and longer duration opportunities for interaction may change the results. For example, established teams may develop more preexisting ties and thus reduce the likelihood of proximate tie formation further (Granovetter, 1985). On the other hand, a wider geographic region may discourage forming more distant ties that likely are more costly and thus counteract the effect of preexisting ties. Relatedly, a longer duration may give those with no preexisting ties wider latitude to search further, thus reducing the differences with those with preexisting relationships.

Despite these limitations, our research proposes and identifies an important scope condition on a key theoretical and policy-relevant issue. Geographic spillovers are a central component of models of individual and organizational innovation, as well as of regional economic development (Jaffe, Trajtenberg and Henderson, 1993; Anselin, Varga and Acs, 1997). Our findings suggest that once networks are established, geographic spillovers may be less predictive of performance than earlier on in an ecosystem’s history. Further, we suggest an important mechanism for why older, more established firms in innovative regions may be less innovative than their newer peers. In an effort to minimize downsides, preexisting network ties may prevent firms from exploiting localized knowledge. These concerns also apply to organizations where spillovers among employees are critical for transferring knowledge across departments or business units (Allen and Cohen, 1969). Our findings suggest that once networks are “locked in” inducing tie formation by reorganizing office locations, layouts, or other geographic reshuffling may not succeed.

Several opportunities exist for extending our results to explore the mechanisms we propose. While preexisting relationships may limit geographic spillovers through a variety of mechanisms, our model and empirics cannot completely distinguish between them. Future research should investigate whether preexisting networks act to limit spillovers by serving as constraints (Granovetter, 1985; Krackhardt, 1999), as insurance (Karlan et al., 2009), or through social
disincentives to form new ties (Gargiulo and Benassi, 2000). Moreover, future work should examine possible interventions that could overcome the inertial force of preexisting ties. Finally, at the macro-level researchers should consider how preexisting relationships and new opportunities interact in geographic space, and how the optimal organization of proximity can be engineered to increase spillovers (Carrell, Sacerdote and West, 2013).
Appendix A: A formal model of spatially sequential tie formation

To highlight the trade-off between preexisting ties and the formation of spatially proximate ties more fully, we develop a simple, formal model that incorporates the presence of preexisting ties along with the uncertain and sequential nature of network formation. Conceptually, we treat the decision to form a tie with a proximate neighbor as a two-stage sequential portfolio investment decision (Markowitz, 1952). The decision problem of the agent is to determine whether to invest her limited social budget in forming relationships in the current period or wait to form potentially better, but uncertain, relationships in the future. Building on the propinquity literature, we treat the sequence as spatially determined with an agent’s search process starting with local neighbors and subsequently extending out to more distant others.

Formally, the agent first learns about the potential value $v_1$ of forming a relationship with her immediate neighbors. To ease interpretation, this variable can represent the expected value of interacting with an alter, but could also represent the value of a host of observed and unobserved factors such as positive affect (Casciaro and Lobo, 2008), status (Chown and Liu, 2015) or potential connections (Hasan and Bagde, 2015). Yet, unlike her neighbors, the agent does not know the exact value of connecting with more distant partners. Instead, she only knows that the value of these more distant connections have a normally distributed payoff $v_2 \sim N(\mu, \sigma^2)$, where $\mu$ is the average value of the relationship and $\sigma^2$ the variance of this value. These values yield utility that is proportional to the amount the agent invests in each type of relationship. Specifically, the agent must decide how much of her finite social budget $k$ to devote to her neighbors ($w_1$) and how much to invest in forming future relationships with more distant partners ($w_2$).

Even if the expected value of connecting with neighbors or distant others is the same (e.g. $v_1 = v_2$), prior research indicates that individuals still prefer a relationship with their immediate and known neighbors over the riskier and unknown outsider (Gargiulo and Benassi, 2000). To capture this preference for the immediate and known cleanly, we assume the agent is risk averse with some positive constant-absolute risk-aversion parameter $A$ (Arrow, 1974). Under this

---

13 Building on a larger literature modeling the dynamics of social networks formation we assume the decision to form a tie is unilateral (e.g., Snijders, Van de Bunt and Steglich, 2010). See Jackson (2008) for examples of strategic bilateral network formation models.

14 We use CARA preferences for both substantive and analytical reasons. Substantively, it seems reasonable to assume that as an individual’s social budget increases she should have increasing, or at least constant, tolerance for network formation “risk.” Analytically, while models with decreasing-absolute risk-aversion lead to the same set of predictions, CARA preferences provide simpler and more solutions.
assumption the expected utility of the agent is \( u_i(\cdot) = w_1v_1 + w_2v_2 - \frac{A}{2} \text{Var}(w_2v_2) \). In the context of our two-stage model, \( A \) reflects not just pure risk aversion, but also represents a wide array of mechanisms that lead individuals to prefer neighbors over distant others. For example, if individuals discount the future, then \( A \) can be thought of as capturing not just risk aversion, but also the amount an agent discounts the value of potential future relationships vis-a-vis forming ties in the present. Similarly, if search itself is costly—perhaps because it is unpleasant or simply takes time away from other activities—then a larger \( A \) can be interpreted as capturing these costs. To avoid unneeded complexity, we do not include these mechanisms as separate parameters our model.\(^{15}\)

To incorporate our main theoretical construct—*existing relationships*—we introduce a parameter \((w_0)\) in the utility function of the agent that captures the idea that some agents have more, or fewer, existing relationships. That is, we assume each agent has “pre-committed” \( w_0 \) to existing relationships with value \( v_0 \). We hold constant each agent’s social budget \( k \) such that \( w_0 + w_1 + w_2 = k \) for all agents. With these constraints and assumptions, the agent maximizes the following utility function:

\[
u_i(\cdot) = w_0v_0 + w_1v_1 + w_2v_2 - \frac{A}{2} \text{Var}(w_2v_2)\tag{1}\]

Rearranging this expression and normalizing the budget constraint to \( \frac{w_0}{k} + \frac{w_1}{k} + \frac{w_2}{k} = 1 \), we can see more clearly see how variation in the budget constraint \( k \) may alter the agent’s \( w_2 \) investment decision.

\[
\frac{w_0}{k}v_0 + (1 - \frac{w_0}{k} - \frac{w_2}{k})v_1 + \frac{w_2}{k}v_2 - \frac{A}{2} \text{Var}(\frac{w_2}{k}v_2)\tag{2}\]

Taking the F.O.C. with respect to \( w_2 \) gives us the following expression for the agent’s optimal \( w_2 \):

\[
w_2^* = \frac{k(v_2 - v_1)}{A\sigma^2}\tag{3}\]

Substituting and again rearranging yields the following expression for how much time and effort the agent puts into forming relationships with her immediate neighbors:

---

\(^{15}\)[Including these parameters as separate terms does not meaningfully change any of the model’s predictions.](#)
\[ w^*_1 = k(1 - \frac{w_0}{k} - \frac{v_2 - v_1}{A\sigma^2}) \] (4)

The expressions for \( w^*_1 \) and \( w^*_2 \) imply three observable patterns about the link between preexisting relationships, proximity and new relationship formation. First, if agents bring equivalent social budgets, then an agent with more existing relationships must form fewer new relationships overall. That is, as \( w_0 \) increases in our model, the sum of \( w_1 \) and \( w_2 \) decreases. This leads to Hypothesis 3.

Second, the expressions for \( w^*_1 \) and \( w^*_2 \) indicate that when the agent has more existing relationships, she will invest relatively less effort and time connecting with her neighbors than with outsiders. Moreover, the expression for \( w^*_1 \) reveals that as \( w_0 \) increases the amount of time and effort invested in forming relationships with neighbors decreases. While the overall social budget constraint may limit how much effort and time the agent can put into distant relationships \( w^*_2 \), she will always try to reduce \( w^*_1 \) first. This leads to Hypothesis 4.

Third and finally, the term \( \frac{2w_0}{k} \) in the expression for \( w^*_1 \) suggests that as the agent’s social budget increases the impact of preexisting ties on relationship formation declines. With an increase in \( k \) the social budget constraint never binds, and so the agent does not have to reduce her ties to neighbors to maintain her optimal level of network exploration. This mitigates the prediction of Hypothesis 4. Furthermore, as \( k \) increases any level of \( w_0 \) becomes relatively less consequential, and so as an agent’s social budget increases the predicted effect in Hypothesis 3 diminishes. While it is clearly difficult to assign, or even observe, an individual’s latent social budget, different types of relationships exhibit markedly different budget constraints, that is, weak ties have much larger budgets than strong ties. This leads to Hypothesis 5, that the predicted outcomes in Hypotheses 3 and 4 diminish for weak ties.

\[ ^{16}\text{We set aside predictions that involve variation in } v_0, v_1, \text{ and } v_2 \text{ for measurement reasons. First, they are fundamentally hard to observe because they may be dyad specific and not universal measures of the payoff to relationship formation. An individual } i \text{ may “click” with person } j, \text{ whereas } k \text{ may find } j \text{ provides very little value. Thus, even if we use measures of individual ability like exam scores or GPA, they may not capture the true payoffs to relationship formation. Second, in our empirical context individuals are randomized to teams and locations, which leads to each person having a spatial sequence of potential partners that is, at least in expectation, the same in terms of } v_1 \text{ and } v_2. \text{ Thus the research design randomizes away endogenous differences in the quality of neighbors and more distant alters. Relatively, we also avoid predictions that involve } A \text{ as our data does not provide good measures of risk-aversion and other economic preferences. Finally, we focus on } w_0, w_1, \text{ and } w_2 \text{ because of our interest in network formation processes and because we can tightly measure these variables in our context using roster-based network surveys and digital communication data.} \]
References


Figure 1: Floor plan of the bootcamp space. Workstation tables are indicated by the purple numbers. Table 40 (not shown on the diagram) was to the left of Table 39.
Table 1: Individual-level dyad summary statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>i knows j pre-bootcamp</td>
<td>12,222</td>
<td>0.02</td>
<td>0.16</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Distance between i and j</td>
<td>12,222</td>
<td>1,322.49</td>
<td>724.06</td>
<td>89.00</td>
<td>2,627.10</td>
</tr>
<tr>
<td>log(Distance)</td>
<td>12,222</td>
<td>6.97</td>
<td>0.75</td>
<td>4.49</td>
<td>7.87</td>
</tr>
<tr>
<td>i has preexisting ties</td>
<td>12,222</td>
<td>0.50</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Know</td>
<td>12,222</td>
<td>0.18</td>
<td>0.38</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Advice</td>
<td>12,222</td>
<td>0.05</td>
<td>0.21</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Friend</td>
<td>12,222</td>
<td>0.02</td>
<td>0.15</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Future Team</td>
<td>12,222</td>
<td>0.04</td>
<td>0.20</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Message</td>
<td>12,222</td>
<td>0.36</td>
<td>1.41</td>
<td>0</td>
<td>35</td>
</tr>
</tbody>
</table>

Note: Observation is the ego i-alter j individual-individual dyad.
Table 2: Team-level dyad summary statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team $i$ has preexisting ties (mean)</td>
<td>1,560</td>
<td>0.50</td>
<td>0.29</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Team $j$’s performance</td>
<td>1,560</td>
<td>2.79</td>
<td>0.20</td>
<td>2.44</td>
<td>3.29</td>
</tr>
<tr>
<td>Team $j$’s performance / log(Distance)</td>
<td>1,560</td>
<td>0.41</td>
<td>0.06</td>
<td>0.31</td>
<td>0.66</td>
</tr>
<tr>
<td>Team $j$’s prototype score</td>
<td>1,560</td>
<td>3.01</td>
<td>0.34</td>
<td>2.29</td>
<td>3.87</td>
</tr>
<tr>
<td>Team $j$’s prototype score / log(Distance)</td>
<td>1,560</td>
<td>0.44</td>
<td>0.08</td>
<td>0.29</td>
<td>0.82</td>
</tr>
<tr>
<td>Team $j$’s splash page score</td>
<td>1,560</td>
<td>2.85</td>
<td>0.42</td>
<td>2.07</td>
<td>3.60</td>
</tr>
<tr>
<td>Team $j$’s splash page score / log(Distance)</td>
<td>1,560</td>
<td>0.42</td>
<td>0.08</td>
<td>0.27</td>
<td>0.70</td>
</tr>
<tr>
<td>Team $j$’s total prototype edits</td>
<td>1,260</td>
<td>19.89</td>
<td>9.97</td>
<td>4</td>
<td>46</td>
</tr>
</tbody>
</table>

*Note:* Observation is the ego $i$-alter $j$ team-team dyad.
Table 3: Was the spatial randomization successful?

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>log(Distance between $i$ and $j$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$i$ knows $j$ pre-bootcamp</td>
<td>0.025 (0.041)</td>
</tr>
<tr>
<td>$i$ has preexisting ties</td>
<td>$-0.018$ (0.022)</td>
</tr>
<tr>
<td>Constant</td>
<td>$6.975^{***}$ (0.037)</td>
</tr>
<tr>
<td>Observations</td>
<td>12,222</td>
</tr>
</tbody>
</table>

Note:  
* $p<0.1$; ** $p<0.05$; *** $p<0.01$  
Linear regression models with standard errors in parenthesis.  
Robust SEs clustered at ego team, alter team, and team-dyad levels.  
Individual-level between-team ego $i$-alter $j$ dyads.
Table 4: Does spatial distance shape network formation?

<table>
<thead>
<tr>
<th></th>
<th>Know</th>
<th>Advice</th>
<th>Friend</th>
<th>Future Team</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>(i) knows (j) pre-bootcamp</td>
<td>3.046***</td>
<td>2.709***</td>
<td>3.048***</td>
<td>2.379***</td>
<td>0.699***</td>
</tr>
<tr>
<td></td>
<td>(0.286)</td>
<td>(0.233)</td>
<td>(0.237)</td>
<td>(0.194)</td>
<td>(0.208)</td>
</tr>
<tr>
<td>log(Distance between (i) and (j))</td>
<td>−0.144**</td>
<td>−0.133*</td>
<td>−0.240***</td>
<td>−0.146</td>
<td>−0.019</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.079)</td>
<td>(0.077)</td>
<td>(0.106)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Constant</td>
<td>−0.656</td>
<td>−2.336***</td>
<td>−2.376***</td>
<td>−2.342***</td>
<td>0.475***</td>
</tr>
<tr>
<td></td>
<td>(0.456)</td>
<td>(0.566)</td>
<td>(0.597)</td>
<td>(0.747)</td>
<td>(0.167)</td>
</tr>
<tr>
<td>Observations</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>−5,394.187</td>
<td>−2,086.799</td>
<td>−1,220.773</td>
<td>−1,936.390</td>
<td></td>
</tr>
</tbody>
</table>

*\(p<0.1; \ ^{**}p<0.05; \ ^{***}p<0.01\)

Note:  
Models 1-4 logistic regression. Model 5 quasi-poisson regression.  
Robust SEs clustered at ego team, alter team, and team-dyad levels.  
Individual-level between-team ego i-alter j dyads
Table 5: Do teams perform better (worse) when they are spatially closer to other high (low) performing teams?

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team $i$’s performance</td>
<td>$-0.026^{***}$</td>
<td>$-0.090^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.032)</td>
<td></td>
</tr>
<tr>
<td>Team $j$’s performance / log(Distance)</td>
<td>0.218*</td>
<td>0.387**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.176)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.858^{***}</td>
<td>2.698^{***}</td>
<td>2.881^{***}</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.062)</td>
<td>(0.044)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,560</td>
<td>1,560</td>
<td>1,560</td>
</tr>
</tbody>
</table>

*Note:* *p<0.1; **p<0.05; ***p<0.01

OLS with standard errors in parenthesis.
Robust SEs clustered at ego team, alter team, and team-dyad levels.
Unit of analysis is the ego $i$-alter $j$ team-dyad.
Table 6: Do preexisting ties moderate the effect of spatial distance on network formation?

<table>
<thead>
<tr>
<th></th>
<th>Know</th>
<th>Advice</th>
<th>Friend</th>
<th>Future Team</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>i knows j pre-bootcamp</td>
<td>3.096***</td>
<td>2.911***</td>
<td>3.310***</td>
<td>2.432***</td>
<td>0.722***</td>
</tr>
<tr>
<td></td>
<td>(0.291)</td>
<td>(0.240)</td>
<td>(0.275)</td>
<td>(0.201)</td>
<td>(0.211)</td>
</tr>
<tr>
<td>log(Distance)</td>
<td>-0.191**</td>
<td>-0.287***</td>
<td>-0.365***</td>
<td>-0.331***</td>
<td>-0.060**</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.084)</td>
<td>(0.105)</td>
<td>(0.110)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>i has preexisting ties</td>
<td>-0.773*</td>
<td>-2.720***</td>
<td>-2.258**</td>
<td>-2.628***</td>
<td>-0.629***</td>
</tr>
<tr>
<td></td>
<td>(0.467)</td>
<td>(0.896)</td>
<td>(0.932)</td>
<td>(0.858)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>Preexisting × log(Distance)</td>
<td>0.094</td>
<td>0.332***</td>
<td>0.250*</td>
<td>0.365***</td>
<td>0.084**</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.128)</td>
<td>(0.131)</td>
<td>(0.113)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.275</td>
<td>-1.088*</td>
<td>-1.291*</td>
<td>-1.021</td>
<td>0.788***</td>
</tr>
<tr>
<td></td>
<td>(0.554)</td>
<td>(0.574)</td>
<td>(0.744)</td>
<td>(0.762)</td>
<td>(0.218)</td>
</tr>
<tr>
<td>Observations</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
<td>12,222</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-5,390.127</td>
<td>-2,072.903</td>
<td>-1,211.387</td>
<td>-1,931.188</td>
<td>-1,931.188</td>
</tr>
</tbody>
</table>

Note: *p<0.1; **p<0.05; ***p<0.01

Models 1-4 logistic regression. Model 5 quasi-poisson regression.
Robust SEs clustered at ego team, alter team, and team-dyad levels.
Individual-level between-team ego i-alter j dyads.
Table 7: Do preexisting ties moderate spatial correlation in team outcomes and team work patterns?

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Team i’s performance</th>
<th>Team i’s prototype score</th>
<th>Team i’s splash page score</th>
<th>Team i’s edited prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Team i has preexisting ties (mean)</td>
<td>0.134</td>
<td>0.148</td>
<td>0.303</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.213)</td>
<td>(0.251)</td>
<td></td>
</tr>
<tr>
<td>Team j’s performance</td>
<td>-0.088***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j’s performance / log(Distance)</td>
<td>0.467**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.185)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preexisting × (Team j’s performance / log(·))</td>
<td>-0.174**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.079)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j’s prototype score</td>
<td>-0.119**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j’s prototype score / log(Distance)</td>
<td>0.757***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.282)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preexisting × (Team j’s prototype score / log(·))</td>
<td>-0.334***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j’s splash page score</td>
<td>-0.151**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.069)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j’s splash page score / log(Distance)</td>
<td>1.037**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.475)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preexisting × (Team j’s splash page score / log(·))</td>
<td>-0.445***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j edited prototype</td>
<td>-0.044</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team j edited prototype / log(Distance)</td>
<td>0.561*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.315)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preexisting × (Team j edited prototype / log(·))</td>
<td>-0.769***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2.811***</td>
<td>3.031***</td>
<td>2.789***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.089)</td>
<td>(0.120)</td>
<td>(0.159)</td>
<td></td>
</tr>
</tbody>
</table>

| Time Period FE | - | - | - | Yes |
| Ego Team FE    | - | - | - | Yes |
| Observations   | 1,560 | 1,560 | 1,560 | 191,520 |

Note: *p<0.1; **p<0.05; ***p<0.01
OLS with standard errors in parenthesis.
Robust SEs clustered at ego team, alter team, and team-dyad levels.
For Models 1-3, unit of analysis is the ego i-alter j team-dyad.
For Model 4, unit of analysis is a 30-minute period for each team-dyad.