A Tale of Two Cities and Many Villages: Urban Development and Social Networks in India

Citation	Xie, Kevin. 2018. A Tale of Two Cities and Many Villages: Urban Development and Social Networks in India. Bachelor's thesis, Harvard College.
Citable link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:39011598
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA;This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

1 Introduction

I examine the relationship between urbanization, proximity to urban areas, and the size and structure of social networks, using data from 66 villages in Karnataka, India that were originally collected in 2006 and 2007 by Banerjee et al. (2013). First, I test whether urbanization—as measured by the intensity of nighttime lights (NTL) in each village—varies with the density of networks for three interaction types: economic (such as borrowing money or kerosene), social (such as visiting someone's home), and advice (such as asking for general or medical opinions). Using a cross-sectional OLS regression of network density on NTL levels and distance to metropolitan areas and controlling for demographic and village characteristics, I find that there is a negative relationship between NTL levels and the density of social and economic interaction networks, as well as the overall network that includes all three interaction types. A one percent increase in the NTL level is correlated with a 0.09% decrease in the density of the network of all interactions.

Second, I examine whether NTL levels are related to variation in the transitivity, or degree of interconnectivity, in a network, as measured by two metrics: the average clustering coefficient and support. I find that NTL levels are negatively correlated with the clustering coefficient and support of the networks for economic, social, and advice interactions. A one percent increase in the NTL level is associated with a 0.11% decrease in the average clustering coefficient and a 0.09% decrease in the support for all interactions.

NTL measurements of village development may be confounded by internal growth that is not related to urbanization or urban access, so I also examine how proximity to metropolitan areas, as measured by travel distance, may influence village social networks. The villages are located near either Bangalore, a fast-growing city with a 2005 population of around 6.8 million, or Mysore, a smaller city with a 2005 population of less than 900,000. I find that villages that are closer to Bangalore have lower levels of clustering (36% less) and support (11% less), when holding NTL levels constant. This suggests that proximity to larger metropolitan areas does have a relationship with network structure.

Finally, I perform an instrumental variable regression of network statistics on NTL, using whether a village is in the 'Bangalore cohort' as an instrument. I find that differences in NTL resulting from proximity to Bangalore are negatively correlated with the density of economic and advice interactions, along with the clustering and support of all interaction types. Using this method, I find that a one percent increase in NTL levels is associated with a 0.46% decrease in the average clustering coefficient and a 0.2% decrease in the support of the union network of all interactions. All of these results provide evidence that urbanization may not only have an effect on the frequency of interactions, but could affect peoples' choices of who they interact with.

Given the nature of the data, it is not possible to conclusively determine the direction of causality. I examine two potential sources of reverse causality: community involvement in public policy and the diffusion of technology. Based on the size and demographics of the villages, I argue that it is unlikely that the villages would have been able to organize or influence policy to affect the degree of urbanization or development around them. I also find no significant relationship between various network statistics and the rate of election card ownership, electricity usage, and latrine ownership in villages, suggesting that the structure of social networks does not aid in the spread of urban technologies or political engagement.

I begin in Section 2 by reviewing the existing literature. Section 3 introduces two models that are useful for predicting the relationship between urbanization and network structure, and also discusses hypotheses about specific ways that urban development and proximity could influence networks. Section 4 describes the network and NTL data, in addition to defining some key network terms. I then discuss my empirical strategy in Section 5 before presenting results in Section 6.1. Next, I examine how to evaluate possible sources of reverse causality or omitted variable bias in Section 6.2, before concluding.

2 Literature Review

Urbanization has played a major role in human history and development. For thousands of years, cities have served as locations for individuals to live, work, and engage in various social and economic interactions with each other. The growth and development of cities has motivated a significant body of research on the relationship between urbanization and communities' social networks.

Research in other fields, especially in sociology and anthropology, has focused on questions such as whether urbanization weakens social cohesion (White and Guest, 2003). There is a vigorous academic debate in those fields about the validity of the "community lost" hypothesis, which posits that urbanization decreases the quantity and frequency of social interactions within a community. A study of residents of 50 communities in northern California with different levels of urban development found that social networks in urban settings were less dense (Fischer, 1982). However, Fischer notes that his study was vulnerable to selection effects, since underlying social, economic, or geographical factors may contribute to differences in social networks. Indeed, when controlling for educational attainment, he finds no significant effect.

Other research suggests that urbanization may encourage specific types of social interactions, including voluntary and transactional ties (White and Guest, 2003). In this study, the authors find that while the network of interactions is less dense in urban areas, urban residents tend to have larger social networks and different types of interactions, which casts doubt on the "community lost" hypothesis. This is consistent with theoretical work by Sato and Zenou, which predicts that individuals in more urbanized areas will have more "weak ties," which link people that generally are not in the same social groups, than those in less urban areas (2015). In sufficiently sparsely-populated areas, their model predicts that no weak ties will exist.

In addition to providing further evidence for the "community lost" debate, this paper and its use of an Indian context provides a unique contribution to the literature on the relationship between urbanization and social networks, which has previously mainly dealt with communities in developed countries. Investigating these types of questions in an emerging market setting is both highly relevant and potentially more fruitful. Social networks are especially important in developing nations, as ineffective formal institutions in those countries (e.g. the absence of lenders and credit rating services) combined with vulnerability to income shocks often make informal ties between individuals vital for their economic security (Breza, 2016). Evidence also suggests that individuals' engagement in networks has impacts on their physical and mental health (Israel, 1985; Leavy, 1983; Gilchrist, 2009). Networks can also facilitate the spread of knowledge and encourage community organizing and collective action, by making it easier for ideas to propagate through a community (Gilchrist, 2009). These social networks may be affected by the rapid urbanization that is taking place in many developing countries. The number of people living in cities in the developing world is expected to increase by over 1.3 billion by 2030, accounting for 96% of total global urban population growth (World Bank, 2013). Looking at how urban development impacts social and economic interactions between people in emerging-market economies may help identify important impacts of urbanization on the provision of public services, the diffusion of learning, and quality of life.

3 Theory

In this section, I discuss two models of how urbanization and proximity to urban areas could affect the size and shape of networks, based on the work of Munshi and Rosenzweig (2016) and Jackson et al. (2012). Using these models, I predict that access to larger metropolitan areas will have a negative impact on network density and transitivity, while development within villages may have an ambiguous impact. I then discuss, in Sections 3.3 and 3.4, hypotheses about specific mechanisms that could facilitate changes in network structure (such as reduced transportation and communication costs, or a lower need for peer-based monitoring). These sections distinguish ways through which urbanization could either increase or decrease the density of networks (by changing the number of interactions between community members), versus the degree of transitivity in networks (by changing not the number of connections, but with whom those connections take place). I also discuss how it may be possible to test for these mechanisms using the data.

It is important to note that these hypotheses are not mutually exclusive; it is likely that multiple factors will affect the size and shape of social networks. However, these hypotheses are a useful lens for interpreting my results, since they can give some indication as to which theories, or combination of theories, could be consistent with the data. It is also possible that causality moves in the opposite direction, and the structure of a village's social network influences its rate of urban development. I discuss the question of reverse causality in Section 6.2.

3.1 Modeling network formation

Generating testable theories about changes in network structure is a difficult task, given the multi-dimensional nature of networks. In the interest of tractability, I discuss two separate models for how households may choose to participate in networks and their predictions for how urbanization and urban access may affect network structure. I first present a basic model of network formation, originally developed by Munshi and Rosenzweig (2016), in order to examine how urban development and urban proximity might affect network density. In this model, I focus on individuals' choices of whether to participate in the network, rather than the specific structure of the network.

To begin, assume that households have logarithmic utility functions, such that they have an expected utility that consists of mean income M and normalized risk V/M^2:

$$EU = \log(M) - \frac{1}{2}\frac{V}{M^2} \tag{1}$$

In this model, households can choose to either participate in a village network, or to seek connections outside of the community.[1] The network could involve social or economic interactions, such as insurance, favor exchange, or companionship. The village network involves full risk-sharing, which is always maintained *ex post*. This is a simplifying assumption that makes it possible to generate a closed-form solution to the participation problem. These two options are mutually exclusive; it is not possible to engage in the network and also maintain external connections. This could be because having external connections makes households less accountable if they decide to cheat or sever ties within the village network, or introduces some other asymmetries in information or commitment; as a result, members of the network may refuse to allow households with external connections to participate (Munshi and Rosenzweig, 2016).

Define M_A, V_A to be the mean and variance of the household's income under autarky, when it neither participates in the village network nor forms external connections. M_N and V_N are the mean and variance, respectively, if the household participates in the village network. For a household i, choosing to form external connections increases the household's mean income to $M_A(1 + \bar{\epsilon}_i)$. This is because connections with others outside of the community could allow for a household to gain financial or social support, such as loans, job referrals, or gifts.

Different households stand to gain differently from forming external connections, and so $\bar{\epsilon}_i \sim F(\bar{\epsilon})$, with the following constraints on the distribution:

1. $F(\bar{\epsilon})$ can take any value between $[0, \infty)$.

[1] Consistent with other models that involve full risk-sharing, there is no option for households to save or invest.

2. The density of the distribution, $f(\bar{\epsilon})$, is decreasing in $\bar{\epsilon}$; it is reasonable to assume that the percentage of 'extremely helpful' outside connections is relatively low compared to 'moderately helpful' ones.

External connections also reduce income risk to $\beta V_A/M_A^2$, with $\beta < 1$; this reflects how access to an urban area may allow for alternative forms of income smoothing over various states (e.g. through access to formal institutions or proximity to other opportunities).

Note that under these assumptions, every household will either participate in the village network or elect to form external connections. To see this, note that under full risk-sharing, each household's income is a fraction of total income $\sum_{i\in n} y_i$ generated by the village. Given that mean income is the same across all households, this means that network participants will equally divide their total income. Therefore, by taking expectations:

$$M_N = E(\frac{1}{n}\sum_{i\in n} y_i) = \frac{1}{n} * nM_A = M_A \tag{2}$$

$$V_N = \text{Var}(\frac{1}{n}\sum_{i\in n} y_i) = \frac{1}{n^2} * nV_A = V_A/n \tag{3}$$

where n is the number of households participating in the network. While the expected value of income is the same, the risk has decreased, so participation in the network is strictly better than autarky, assuming that $n \geq 2$.

Using the utility functions in Equation 1, household i will decide to participate in the network if

$$\log(M_N) - \frac{1}{2}\frac{V_N}{M_N^2} \geq \log((1+\bar{\epsilon}_i)M_A) - \frac{1}{2}\beta\frac{V_A}{M_A^2} \tag{4}$$

Therefore, household i is indifferent between participating in the network and forming external connections if the added value of forming an external connection $\epsilon_i = \log(1+\bar{\epsilon}_i)$

satisfies the following:

$$\epsilon_i = \log(M_N/M_A) + \frac{1}{2}\beta\frac{V_A}{M_A^2} - \frac{1}{2}\frac{V_N}{M_N^2} \tag{5}$$

$$\rightarrow \bar{\epsilon}_i = \frac{M_N}{M_A}\left(e^{\frac{1}{2}\beta\frac{V_A}{M_A^2} - \frac{1}{2}\frac{V_N}{M_N^2}}\right) - 1$$

Note that the households' decision to participate in the network is a binary one; there is no option to form additional links to the network. This means that density in this context is the same as the participation rate in the village network. Intuitively, a more complicated model, in which households could choose their degree of investment in the network, would yield similar results of participation decreasing as the opportunity cost of community engagement rises, because the general trade-off between insurance and benefits from external connections remains. Therefore, using the assumed distribution of $\bar{\epsilon}$, the fraction of households who choose to participate in the village network is:

$$n/p = F\left(\frac{M_N}{M_A}\left(e^{\frac{1}{2}\beta\frac{V_A}{M_A^2} - \frac{1}{2}\frac{V_N}{M_N^2}}\right) - 1\right) \tag{6}$$

where n is the number of participants and p is the total number of households in the village. Given these results, we can now perform comparative statics:

1. n/p is increasing in M_N. For example, higher income growth in the village would raise the expected value of income within the network, which would incentivize households to remain.

2. n/p is increasing in β, V_A. As the external connections become riskier due to higher variance in income realization, they become less attractive.

3. n/p is decreasing in V_N and M_A. Riskier conditions within the village make diversification via external connections a more appealing option. Similarly, higher expected income under autarky reduces the need for engaging in the network.

This model predicts that connectivity with outside opportunities, such as through proximity to urban centers, would decrease network participation and density. However, it is not clear what the net effect of urbanization or proximity to urban areas would be. For instance, economic benefits from urban development within the village could increase income growth, which would encourage people to stay in the network. Similarly, urban development may facilitate the introduction of formal economic institutions and diversification of income sources, lowering V_N and encouraging network participation.

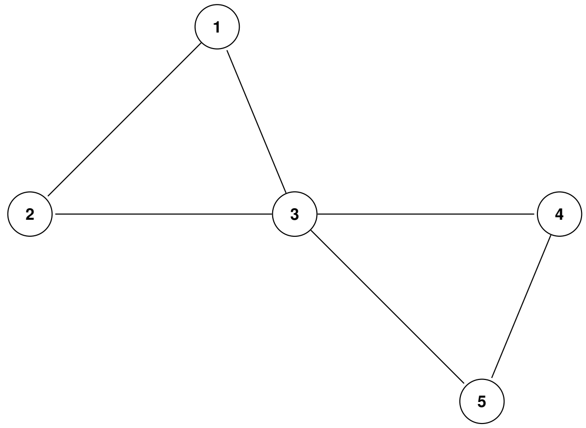
3.2 Modeling network dissolution

I now turn to using a separate model, based on theory developed by Jackson et al. to explain how urbanization may also affect transitivity in networks (2012). This 'social quilts' model also captures how a household's choice to sever its participation in the network has effects beyond their immediate connections.

First, I assume that a village has a finite number of households $N = \{1, 2, 3, ..., n\}$, who currently are linked in a network $G = (N, E)$. E represents the set of edges that connect the households. There are discrete time periods $t = \{1, 2, 3, ..\}$; in each period, there is some probability p that a household will either need a favor or will be asked to perform a favor for one of their connections. For simplicity, only one such favor will be demanded during each period. Households discount future periods by a factor of $\delta < 1$ per period.

Performing favors is costly. As a modification to the original model, I allow the costs and benefits of favor exchange to vary across different households. For any household $i \in N$, choosing to perform a favor carries a cost of c_i, while receiving a favor gives a benefit of v_i, with $v_i > c_i$. Favors could include a variety of economic and non-economic interactions, including the lending of money and household supplies or social visits to another household. Extending a network into perpetuity, household i would expect to receive a discounted benefit of $\frac{p(v_i - c_i)}{1-\delta}$ from each connection that it has.

Figure 1: Equilibrium



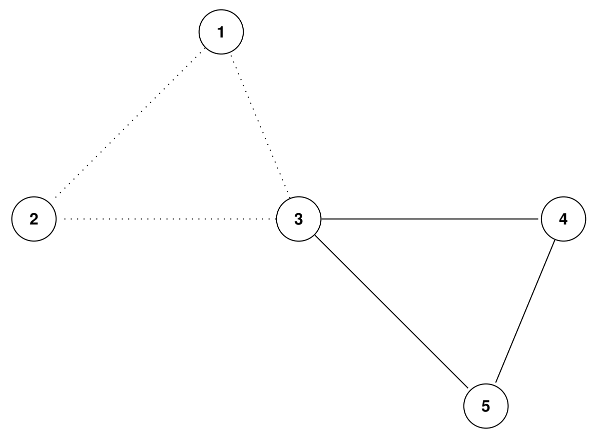
At the beginning of each period, each household decides which links to keep, and severs the ties that they no longer wish to maintain. Then, the requested favor for that period is announced, and the household who is requested to perform the favor decides whether to do it. If a household refuses to perform a favor with one of its connections, that link is destroyed. This structure allows households to respond if they do not receive a requested favor, by severing ties with unhelpful or untrustworthy friends.

In the examples in this section, I assume that $c_i > \frac{p(v_i - c_i)}{1-\delta}$, so that the cost of providing a favor outweighs the benefits of keeping a single connection. However, links can still exist since households can connect to, and obtain benefits from, multiple households. For example, consider a network with five households, all connected to each other.

Assuming that $\frac{2p(v_i - c_i)}{1-\delta} > c_i > \frac{p(v_i - c_i)}{1-\delta}$ for all $i \in N$, this network (as shown in Figure 1) is in equilibrium, since each household is better off remaining in the network. This network could change, however, if an individual receives a shock to their valuation of network benefits or to their costs. For example, increased access to formal institutions or to external social circles in a nearby metropolitan area could lower the value of v. The value of c could also increase as the opportunity cost of performing favors is higher when there are other opportunities to invest time and money in an urban setting.

Assume that household 1 now can access external options for socialization and favor exchange. Then either v_i decreases or c_i rises, since the opportunity cost of participating in this network is now higher (by preventing further engagement with the outside world). If this change is sufficiently strong such that now $c_1 > \frac{2p(v_1-c_1)}{1-\delta}$, household 1 will elect to sever its ties to households 2 and 3 the next time it is called to perform a favor. However, this also means that the connection between households 2 and 3 is unsupported, and given that $c_2 > \frac{p(v_2-c_2)}{1-\delta}$, household 2 will elect to sever the connection. Figure 2 illustrates this process.

Figure 2: Disrupting the network equilibrium



Dotted lines indicate broken connections.

These results indicate that initial removals from the network can have a cascading effect, since they can reduce the monitoring ability of other transactions in the community. This finding has two main implications. First, it suggests that minor changes for a select group of individuals in a network could create social contagion that disrupts the broader network. Second, it indicates that access to outside options could also affect the degree of clustering, since link deletion can propagate through cliques (sections of the graph where individuals are highly connected with each other).

3.3 Urbanization and network density

Given the predictions obtained from the models in the previous two sections, I now turn to listing specific ways that urbanization or proximity to urban areas may affect households' choices to participate in village networks. First, I discuss hypotheses for how urbanization could affect the density of networks.[2]

Hypothesis 1: Higher levels of urbanization correspond to a higher density of social and economic interactions, due to changes in infrastructure and the built environment. Urban development is often accompanied by improvements in infrastructure. This could include the expansion of road, electrical, and communications networks. These developments could affect individuals' ability and willingness to engage with their neighbors. Evidence from Kenya suggests that electrification in rural communities is tied to the expansion of markets, along with increased communication and connectivity (Jacobson, 2007). Similarly, evidence from a Rwandan context suggests that access to mobile phones allows entrepreneurs to find and connect with new contacts, even in low-density areas (Donner, 2006).

Other physical changes associated with urbanization could also increase social ties between villagers. For instance, the opening of new shops, markets, or community centers could create new locations for individuals to interact and increase the likelihood of new relationships forming. Increased electrification could allow villagers to stay out of their homes later, giving them more time to visit neighbors.

If infrastructure and the physical environment have a significant effect on networks, I would expect that the density of all types of interactions (social, economic, and advice) would increase, since there is an overall increase in the available time and space to engage with others. This could be considered an increase in M_N in the Munshi and Rosenzweig model in section 3.1.

[2] For a more rigorous definition of network density, refer to Section 4.3.

However, it is also possible that infrastructure could also lead to a decrease in interactions within a village, since it could allow villagers to form relationships with outsiders instead of fellow villagers. This is discussed in Hypothesis 3.

Urbanization could also potentially contribute to a decrease in social network density:

Hypothesis 2: Urbanization and urban proximity are negatively correlated with the density of economic interactions, due to reduced needs for favor exchange caused by the presence of formal institutions and economic development. As villages develop or become closer to urban areas, villagers may get access to formal institutions, such as banks, insurance firms, and welfare offices. This could reduce the need for informal lending networks and other types of favor exchange, since they often take place in the absence of formal institutions (Breza, 2016).

Growth and industrialization in urban centers could lead villagers to leave the village for work and remove incentives for villagers to cooperate. Miguel et al. show, using data from Indonesia, that districts located near areas of rapid industrialization saw increased out-migration and lower levels of reported "mutual cooperation" (2006).

If this is the case, then urbanization would be correlated with lower levels of network density, particularly for interactions that involve economic exchanges (such as borrowing money or goods) or services (such as asking for medical advice). This is because villagers would be able to seek formal assistance elsewhere, which is represented by an increase in M_A in the Munshi and Rosenzweig model or a decrease in v in the Jackson et al. model. There will likely be little to no impact on the frequency of social interactions.

Hypothesis 3: Higher levels of urbanization are negatively correlated with the density of social and economic interactions, due to increased 'competition' with outsiders. Proximity to urban areas with high population density could widen the options for social interaction. For instance, villagers may choose to interact with residents of a neighboring city, leaving them with less time and willingness to interact with fellow villagers.

If this is the case, the frequency of all types of interactions would decrease, as villagers find other people from outside of the community to interact with and obtain favors from. This change could be interpreted as an increase in M_A in the Munshi and Rosenzweig model.

3.4 Urbanization and transitivity

Urbanization might not only affect the number of interactions that take place within a network, but could also change patterns of interaction. In particular, urbanization could possibly affect the level of transitivity in a network. Transitivity refers to the frequency in which two connected people share a mutual 'friend,' and can be quantified in a number of different ways.[3]

Transitivity in networks is not a purely theoretical concept. High levels of clustering could inhibit the spread of knowledge, since information could get trapped within a group of highly interconnected friends with no method of reaching other such cliques. The existence of "weak ties," which lower the degree of clustering by connecting otherwise-isolated groups, has implications for the diffusion of knowledge (Granovetter, 1973). Other research has drawn connections between weak ties and higher levels of civic engagement and community participation (Kavanaugh et al., 2005).

Additionally, the degree of clustering can affect how easy it is for a community to maintain accountability in various interactions. For example, clustering in networks may raise the costs of cheating, since it is more likely for negative information about the cheater to spread to their other acquaintances. This has important implications for informal favor exchange networks (Jackson, 2011).

Previous research has found a relationship between urbanization and transitivity. White and Guest find that urbanization, in an American setting, is associated with higher levels of

[3] See Section 4.3 for a more detailed discussion of network statistics that measure transitivity: the average clustering coefficient and the support.

“segmentation” and “interconnectedness” (2003). However, they do not pose the question of what mechanisms contribute to this relationship.

There are two possible mechanisms through which urbanization could reduce the amount of clustering in a network: by increasing the probability of individuals encountering others outside of their main circle of acquaintances, or by reducing the need to monitor others’ behavior.

Hypothesis 4: Urbanization is correlated with a decrease in clustering for social and economic interactions, due to an improved ability to meet new acquaintances. Urbanization could create additional opportunities to encounter individuals outside of one’s immediate social circle. This could be a result of increased population density, which could create a higher likelihood of meeting new people in a given area. It could also be due to there simply being more options for who to interact with.

If this mechanism were to exist, I would expect that the amount of clustering for all types of social and economic interactions should see a negative relationship with urbanization. Note that while this is similar to Hypothesis 3, it is possible for the amount of clustering in a network to decrease while the density remains the same, if individuals substitute out of interactions with people in their immediate circle in exchange for weak ties.

Hypothesis 5: Urbanization is correlated with a decrease in clustering for economic interactions, due to a reduced need for peer monitoring. One proposed reason for why clustering exists in social networks is that engaging in highly-clustered groups helps to ensure access to trustworthy exchange partners (Levine and Kurzban, 2006). Empirically, Jackson et al. find, using the same village data as this paper, that links between individuals that involve favor exchange have a higher support, which is a similar measure of transitivity in relationships (2012). Therefore, it is possible that urbanization could reduce the need for favor exchange via improvements in quality of life and the introduction of formal institutions. This could lead to a reduced need for involvement in clusters, since individuals would be more protected from bad favor exchange partners.

If this holds true, I would expect to see a decrease in the amount of clustering for economic interactions. Additionally, there would likely be a decrease in the number of economic interactions (e.g. a less dense network for favor exchange), since this hypothesis is predicated on the presence of economic institutions reducing the need for informal favors, similar to Hypothesis 2.

4 Data

In this section, I first describe the research setting and why it is useful for determining potential effects of urban development and access on network structure. I then describe the network data and provide definitions for the network statistics used in this paper. I then discuss the nighttime lights data, and present summary statistics on various village characteristics.

4.1 Setting

The data include 66 villages in Karnataka, a state in southwest India. Figure 3 shows the locations of the villages, relative to the two main cities in the state: Bangalore and Mysore. The two cities have diverged in terms of population growth over the past 60 years. While Bangalore has historically been the more populous city for much of the 20th century, it has grown exponentially. In 1950, Bangalore had a population roughly twice that of Mysore's; today, its population is more than eight times larger. This is due to a variety of changes, including an influx of migrants, foreign direct investment, and the emergence of Bangalore as a major Indian tech hub. Figure 4 shows the divergence in population growth from 1950 to 2010.

On average, the villages are a two hour drive away from Bangalore. However, seven of the 66 villages are closer to Mysore than to Bangalore, in terms of travel distance.[4] This fact, along with the relatively recent divergence between the two cities, provides a useful setting for analyzing how proximity to a large metropolitan area is related to the size and structure of networks. Assuming that the villages have existed prior to 1950, the rapid growth of Bangalore, which was primarily fueled by sectors such as technology that were not driven by rural developments, could serve as a sort of 'shock' that induced changes in village networks.[5]

4.2 Network data

The network data were originally collected in 2006 and 2007 by Banerjee et al. (2013).[6] The data include a full household-level census for 75 different villages in Karnataka, India. The census contained questions about demographic information (e.g. the head of household's religion and caste) and physical characteristics (e.g. the number of rooms in the house, whether the household has electricity and/or a latrine). Households were also asked if they had a village 'leader,' which includes *gram panchayat* (local council) members, *anganwadi* (mother and childcare center) teachers, shopkeepers, and self-help group leaders.

In total, 14,904 households are included in the data. An individual-level survey was also administered to a subset of villagers in each of the 75 villages. Responses were collected from 16,984 villagers. The individual respondents come from 48% of all households (Banerjee et al., 2013). This allows for a high sampling rate for interactions in the network.[7] Each

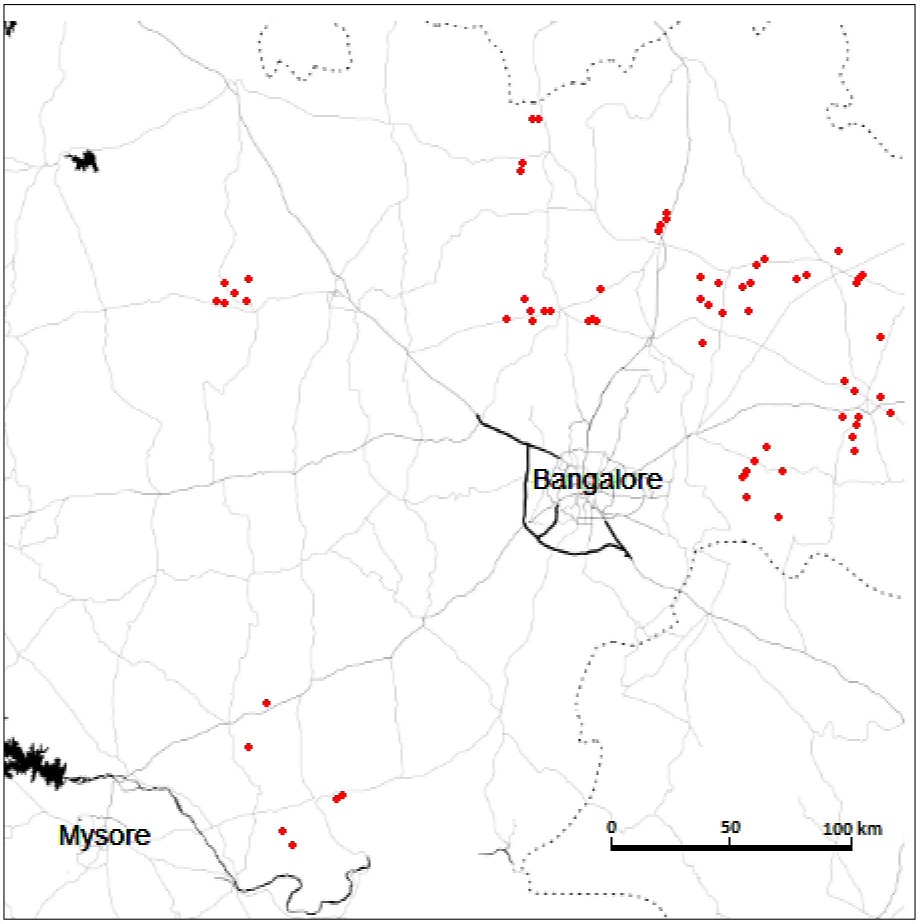
[4] Travel distances were calculated using the GPS coordinates of the villages along with the Google Maps API, which can be found here: `https://developers.google.com/maps/`

[5] This does not rule out potential sources of reverse causality. For example, some aspect of village networks might facilitate development or make villages receptive to changes in nearby cities. For a more detailed discussion of how I address the issue of reverse causality, please see section 6.2.

[6] The dataset is available online at `https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/21538`.

[7] For instance, if Household A did not have a respondent to the individual survey, but an individual from

Figure 3: Map of villages in sample



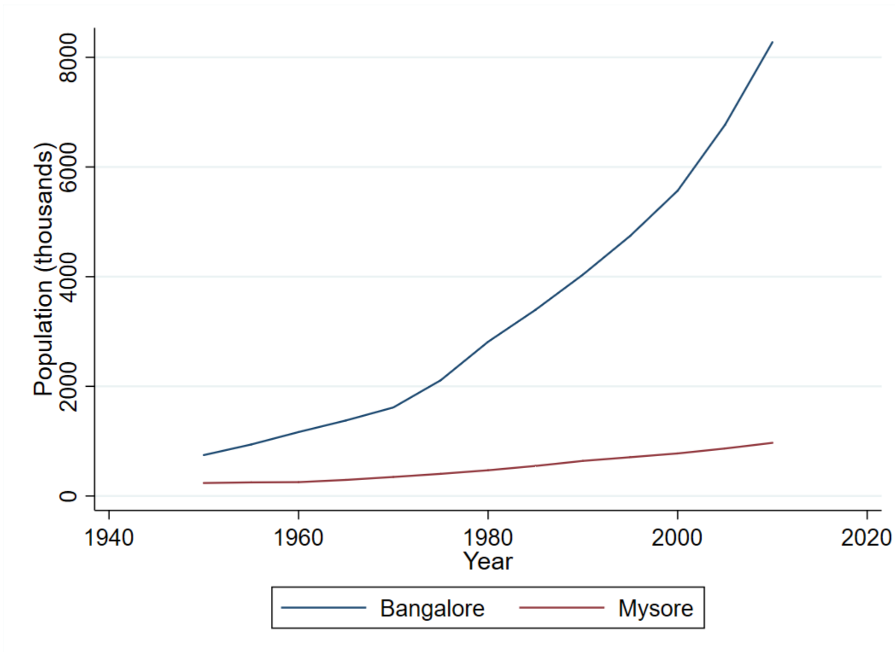
Map includes the 66 villages with available NTL data that were included in regressions. Map tiles from Stamen Design.

respondent to the individual survey was asked to identify which other villagers they interacted with, for 11 different types of social interactions:

1. Borrowing money
2. Lending money
3. Borrowing kerosene and/or rice
4. Lending kerosene and/or rice
5. Obtaining medical advice

Household B listed a member of Household A as someone they interact with, there would be a connection between households A and B.

Figure 4: Population growth in Bangalore and Mysore, 1950-2010



Population figures are from the Bangalore and Mysore urban agglomerations, which include surrounding suburbs. Data is presented in five-year intervals from 1950 to 2010. Source: United Nations Department of Economic and Social Affairs, Population Division (2014).

6. Giving advice
7. Helping with a decision
8. Engaging socially with
9. Visiting someone at their house
10. Inviting someone to their house
11. Attending temple with

Interactions 1-4 involve the exchanges of money, goods, or services, while interactions 5-7 involve advice. Interactions 8-11 are primarily social in nature. I will refer to interactions 1-4 as ‘economic interactions,’ interactions 5-7 as ‘advice interactions,’ and interactions 8-11 as ‘social interactions.’ Creating these categories allows me to test whether urbanization and urban proximity have different relationships with different types of interactions.

Individuals' responses were used to construct undirected graphs at the household level, where each node represents a household and a connection between two nodes indicates that the two households engaged in a particular interaction.

I construct graphs for each of the interaction types from 1-11. Since the graph data is undirected, I created a union graph for the following pairs: borrowing/lending money, borrowing/lending kerosene and rice, visiting/inviting to homes.[8] I then created unions of interactions 1-4, 5-7, and 8-11 as the economic, advice, and social networks. Finally, I create a union of interactions 1-11 as the union of all interactions.

I create both weighted and unweighted graphs for the economic, advice, social, and 'all interactions' networks. In the weighted network, every connection between households has an integer weight, which corresponds to the number of interaction types (1-11) that the households share with each other. The unweighted network better captures how the number of connected households may vary with urbanization, while the weighted network also gives information about the number of different types of interactions that take place between connected households. I primarily focus on the unweighted network statistics, but also include some results for the weighted statistics in the appendix, as a robustness check.

I use the census data to construct indicator variables for various demographic controls, including for religion, caste, and leader status. I also include variables on home characteristics, such as the number of rooms, number of beds, and whether the house has electricity and a latrine. Table 1 presents summary statistics on the demographics for the households included in the data. I also use additional village-level data from the 2001 Indian Village Census in order to obtain information on other village amenities such as schools, clinics and total village area.

[8] For example, if Household A reported borrowing money from Household B, but Household B was not sampled, there would only be a connection between A and B for the borrowing network and not the lending network. The union of the borrowing and lending networks captures all of the relationships.

Table 1: Village-level summary statistics

	Mean	SD
Number of households	198.12	62.83
Has leader	0.13	0.03
Has election card	0.94	0.05
Religion		
Hindu	0.96	
Muslim	0.04	
Christian	< 0.01	
Caste		
General	0.13	
Minority	0.02	
OBC	0.53	
Scheduled Caste/Tribe	0.32	
Caste HHI	0.49	0.16
House characteristics		
Has electricity	0.92	0.04
Has latrine	0.26	0.11
Number of beds	0.86	0.44
Number of rooms	2.39	0.27
Filled out individual survey	0.46	0.03

Note: Sample includes 66 villages with 12,285 total households. Caste HHI is the Herfindahl index calculated using the proportions of caste type in each village.

4.3 Network statistics

To test the predictions generated by the two models in Sections 3.1 and 3.2, I calculate three network statistics: network density, the average clustering coefficient, and support. Density reflects households' willingness to participate in the village network, and therefore is a useful metric to test the conclusions of the Munshi and Rosenzweig model, while the average clustering coefficient and support are measures of transitivity, which can be used to test the Jackson et al. 'social quilts' model.

Network density. Given a graph $G = (N, E)$, where N is the set of nodes (e.g. households) and E is the set of edges (e.g. interactions between households), the network density is the number of edges divided by the number of all possible connections between nodes:

$$\text{density} = \frac{|E|}{\#\text{ of possible connections}} = \frac{2|E|}{|N|(|N|-1)} \tag{7}$$

In the context of the data, the network density measures how many social interactions take place in a given village. It is important to note that network density does not capture all information about the frequency of social interactions; for example, it does not measure how often two connected households interact with each other. However, it is useful as an aggregate measure of how well-connected a village is.

The weighted network density is the sum of all of the edge weights, divided by the total number of possible connections $|N|(|N|-1)/2$.

Clustering coefficient. Using the same notation as before, let $G = (N, E)$. For two nodes $u, v \in N$, $uv \in E$ if there exists an edge connecting u with v.[9] The clustering coefficient of a node v is defined as the number of pairs of neighbors of v that share an edge divided by the total number of pairs of neighbors.

[9] In an undirected graph, uv is equivalent to vu. I focus on definitions for undirected graphs because the data I use is undirected.

$$\text{cluster}_v = \frac{\left|\left((j,k)\;\middle|\; vj, vk \in E \text{ and } jk \in E\right)\right|}{\left|\left((i,h)\;\middle|\; vi, vh \in E\right)\right|} \tag{8}$$

A weighted clustering coefficient can be calculated by taking the sum of the average weights of each node's connection with connected neighbors. Saramäki et al. present a more detailed discussion of how to generalize the clustering coefficient to weighted networks (2007).

The average clustering coefficient is simply the mean of the clustering coefficients of each node in the graph.

Support. The support of a network is another measure of transitivity, proposed by Jackson et al. (2012). Using the same terminology as above, it is defined as:

$$\text{support} = \frac{\left|\left(jk\;\middle|\; ij, ik \in E \text{ for some } i \neq j,k\right)\right|}{|E|} \tag{9}$$

Intuitively, it is the fraction of edges in the network that take place between individuals that share a mutual acquaintance. The weighted support can be calculated by summing the weights of supported links, instead of just counting them. Note that it is possible for the support and average clustering coefficient of a network to differ substantially from each other. This is because the average clustering coefficient is a node-based measurement, while support is an edge-based one. I calculate both using the network data, and use both in my analyses, as a robustness check.

Figure A1 (see appendix) provides a simplified example of a social network, as well as an example of how network statistics are calculated.

4.4 Nighttime lights data

I use the presence and intensity of nighttime lights (NTL) as a proxy for the degree of urban development. There is a substantial literature that investigates whether NTL intensity is positively correlated with urbanization, economic growth, and quality of life. Mellander et al. determine, using data on light intensity in Sweden, that NTL levels are strongly correlated with measures of population density and only weakly correlated with wages (2015). Nighttime lights data has been used previously in development settings; Villa uses NTL as a measure of municipal-level economic growth in order to evaluate effects from a cash transfer experiment in Colombia (2014). Nordhaus and Chen argue that on a macro level, NTL intensity can be a useful measure of economic performance in developing countries, especially when there is otherwise a lack of reliable standard economic data (2014). Additionally, time series analysis of NTL data demonstrates that light intensity and coverage is positively correlated with population and negatively correlated with poverty levels (Proville et al., 2017). Other research has also established that NTL can be a useful measure of income growth on the subnational level, provided that there is sufficient variation in measurements across observations (Henderson et al., 2012; Jean et al., 2016)

Unsurprisingly, much of the literature establishes a correlation between NTL intensity and both population density and economic growth. Therefore, using NTL as a measure of urbanization makes it difficult to disentangle the effects of urbanization-led 'population effects' (e.g. an increased pool of potential friends or a greater diversity of neighbors) from 'growth effects' (e.g. the presence of financial and commercial institutions or access to large marketplaces). However, both are valid mechanisms through which urbanization can affect social dynamics.

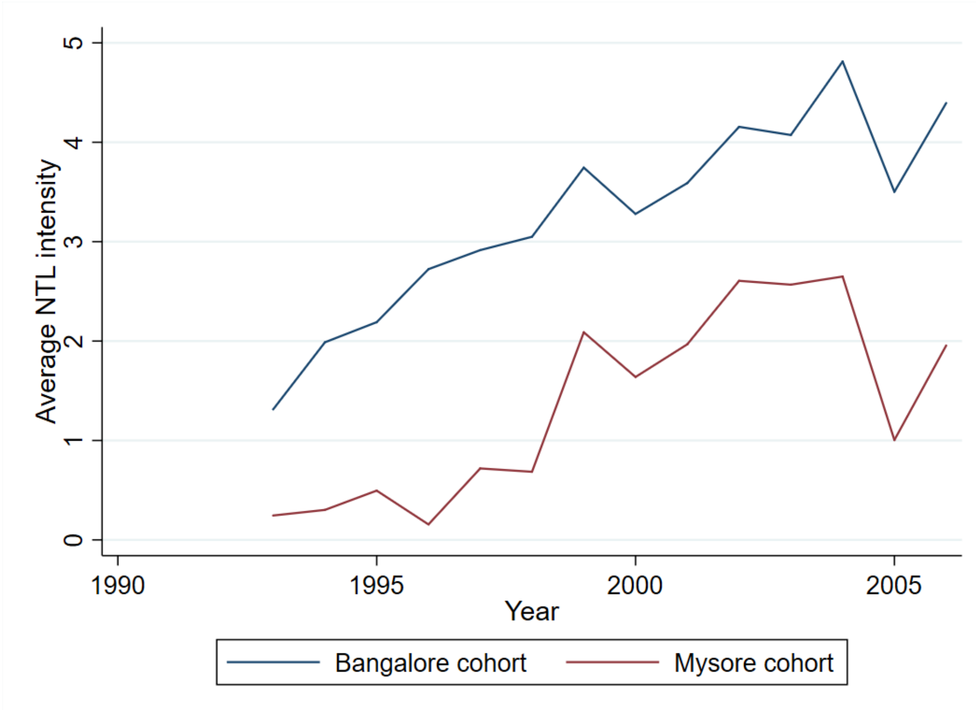
I use NTL time series data from the India Lights Project, which contains nighttime light output over a 20 year period (from 1993 to 2013) for over 600,000 villages in India (Gaba et al., 2016). I used the provided GPS coordinates for each village, and manually matched their locations to village IDs included in the 2001 Indian Census (Office of the Registrar General, 2001). I was able to identify and obtain NTL data for 66 of the 75 villages.[10]

The NTL value for a given village is a number that ranges from 0 (no light output) to 63 (maximum light output). For each village, I take the mean NTL measurement in each month of 2006 and average them. The averaged NTL measurements range from 0.99 to 23.3, with a mean of 4.3 (standard deviation of 3.39). Figure 5 shows the growth in average NTL levels for the 66 villages from 1993 to 2006, divided between villages closer to Bangalore (the ‘Bangalore cohort’) and those closer to Mysore (the ‘Mysore cohort’); the average NTL level nearly quadruples over this time period, which provides evidence of the rapid development that occurred within that timeframe.

NTL levels are weakly correlated with the electrification rate (the percentage of households with electricity) in the village. The correlation coefficient between the log NTL values and the electrification rate is 0.16. The correlation between log NTL values and the number of households in each village is 0.009. This indicates that NTL levels are capturing some aspect of urbanization outside of electrification and village size, such as growth and improved economic prospects.

[10] The other villages either did not have available NTL data, or could not be identified using their GPS coordinates (due to there being multiple villages with the same name in the same district).

Figure 5: Nighttime lights growth near Bangalore and Mysore, 1993-2006



Notes: NTL intensity is the mean of all NTL observations during the year for each village. The Bangalore cohort consists of all villages that have a shorter travel distance to Bangalore than Mysore. The Mysore cohort consists of the other villages. Source: Gaba et al. (2016).

5 Empirical Strategy

In this section, I discuss the two main identification strategies that I use: OLS and instrumental variable regressions of network statistics on NTL levels and travel distance data.

5.1 OLS Regression

I conduct a cross-sectional analysis on the relationship between urbanization, as measured by NTL levels, and network statistics for economic, social, and advice interactions across the 66 villages. I use the following form for an OLS regression:

$$\delta_{v,s} = \alpha + \beta_1 \text{LogNTL}_v + \beta_2 \text{LogMetroDist}_v + \gamma X_v + \epsilon_{v,s} \tag{10}$$

where $\delta_{v,s}$ is the network statistic (either network density, the average clustering coefficient, or the support) for village v based on the interaction type s. LogNTL_v is the mean log NTL level for village v. I use log NTL values because the relative strength of NTL levels is more relevant than the raw numbers and the distribution of NTL values is skewed to the right. LogMetroDist_v is log of the kilometer travel distance between the village and the nearest metropolitan center (either Bangalore or Mysore), as given by Google Maps. X_v is a vector of covariates (distance to district seat, fraction of leader households, religion, caste, number of beds in household, number of rooms in household, village electrification rate, village size, and household and village amenities).

Since interactions between two households that did not participate in the individual survey would not be recorded, non-surveyed households would likely have a smaller number of recorded interactions. Therefore, I also include a control for the fraction of households in each village that participated in the individual-level survey.

β_1 is the estimate of primary interest, denoting the relationship between within-village development and network statistics. I also report the estimates to β_2, since they indicate the relationship between proximity to urban areas and village network statistics, holding village NTL levels constant.

However, while NTL could be measuring the 'prosperity effects' of a village being closer to a metropolitan area, it could also pick up internal growth or other idiosyncratic changes in a village's economic prospects that are not related to urbanization or urban proximity. This could confound the results, since these changes might affect the structure of networks.

To deal with this potential issue, I use a second functional form in order to examine the relationship between distance to metropolitan areas, proximity to Bangalore, and network statistics:

$$\delta_{v,s} = \alpha + \beta_1 \text{LogMetroDist}_v + \beta_2 \text{BangaloreCohort}_v + \gamma X_v + \epsilon_{v,s} \tag{11}$$

where BangaloreCohort_v is an indicator variable that takes a value of 1 if village v has a shorter travel distance to Bangalore than to Mysore, and 0 otherwise.

This regression is designed to capture any relationship between network structure and proximity to urban areas, and does not account for any 'growth' effects from within-village development. β_1 is the relationship between distance to the nearest city and network statistics. This can be interpreted as the effect of 'ease of access' to an urban area on the network. In contrast, β_2 is the estimate for whether access to a larger urban area, with more economic and social opportunities, has any relationship with village networks (which can be thought of as the 'size of access').

Finally, I perform a third set of OLS regressions that includes the NTL, distance, and city cohort variables. Using the same notation as before, I use the following functional form:

$$\delta_{v,s} = \alpha + \beta_1 \text{LogNTL}_v + \beta_2 \text{LogMetroDist}_v + \beta_3 \text{BangaloreCohort}_v + \gamma X_v + \epsilon_{v,s} \tag{12}$$

This set of regressions estimates the relationship between network structure and NTL, distance to the nearest city, and proximity to Bangalore. β_1 is the relationship between within-village development, as measured by NTL intensity, and various network statistics. β_2 is now the correlation between the village's network statistics and the distance to a city, and β_3 is the estimated impact of proximity to the larger city of Bangalore. Under this form, β_2 and β_3 are the primary coefficients of interest, since they are the effect of proximity to urban areas, holding within-village development constant.

5.2 Instrumental Variable Regression

Variation in NTL levels is likely due to a number of factors; some of these are connected to proximity to urban areas or urban development, while others are due to unrelated internal factors. Since I am primarily interested in the former, I also use an instrumental variable estimation of the relationship between NTL and network statistics. I use whether a village is in the 'Bangalore cohort' (closer to Bangalore than Mysore) as the instrument, since the villages' geographical locations were fixed a long time ago, when they were originally founded.

I conduct a two-stage least squares regression, starting with the first-stage equation that examines the relationship between NTL levels and being in the 'Bangalore cohort':

$$\text{LogNTL}_v = \alpha + \theta_1 \text{BangaloreCohort}_v + \gamma X_v + \epsilon_v \tag{13}$$

Table 2 shows the results from the first stage regression. It appears that being in the Bangalore cohort is a strong instrument for NTL; controlling for demographics and distance to the nearest metro area, villages closer to Bangalore have a 83% higher NTL level than those closer to Mysore.

I cannot conclusively determine whether this specification satisfies the exclusion restriction. For instance, it is possible that being close to Bangalore gives villagers a wider range of external people to interact with, without affecting their economic prospects. This effect would not register as an effect on NTL levels but would affect network structure. However, given that most of the villages are on average two hours away from Bangalore by car (and the villages in the 'Mysore cohort' are roughly the same distance away from Mysore), it seems unlikely that villagers would participate in such long commutes for minor social visits, but instead would be more likely to pursue more meaningful or economically beneficial interactions. Therefore, proximity to Bangalore should mainly influence network structure through changes that would be reflected in NTL levels.

I then use these first-stage results to estimate the following:

$$\delta_{v,s} = \alpha + \beta_1 \text{LogNTL}_v + \gamma X_v + \epsilon_{v,s} \tag{14}$$

Using this strategy, β_1 is the main estimate that I am interested in. It can be interpreted as the effect of development, driven by proximity to Bangalore, on network structure.

Table 2: First stage of IV regression

	(1)
VARIABLES	Log NTL
Bangalore cohort	0.829**
	(0.400)
Log distance to metro area	-1.069***
	(0.384)
Log distance to district seat	-0.744***
	(0.175)
Observations	66
R-squared	0.697
Demographic controls	Yes

Results from the first stage regression of NTL on Bangalore cohort. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

6 Discussion

6.1 Results

First, I present results from regressing the network density of economic, social, and advice interactions on log NTL levels and distance to the nearest city, as detailed in Equation 10. Table 3 shows the estimates; there is a significant negative relationship between NTL and the densities for the union of all interactions, as well as the economic and social interaction networks. A one percent higher level of nighttime lights is associated with a 0.09% decrease in the density of the overall network. Estimates for the relationship between distance to the metropolitan area are negative but extremely noisy.

I also present results for the average clustering coefficient and support regressions, in Tables 4 and 5. There is a statistically significant negative relationship between NTL levels and the clustering of economic and social interaction networks ($p < 0.1$), as well as the union of all interactions ($p < 0.05$). NTL levels are also negatively correlated with the support of all of the three interaction types, as well as their union ($p < 0.1$ for the advice network, $p < 0.05$ otherwise). A one percent increase in NTL levels is tied to a 0.11% decrease in the average clustering coefficient and a 0.09% decrease in the support, in the overall network.

It appears that urban development, as captured by NTL, is negatively correlated with the frequency of social and economic ties. While I cannot rule out the existence of positive effects (such as those discussed in Hypothesis 1), the direction of the estimates seems to indicate that urbanization corresponds with a net decrease in network density. These results appear to be consistent with Hypothesis 3, where development creates greater outside options for interaction, which decreases the density of ties within the village. However, there is no significant relationship between NTL and the density of the advice network. Perhaps this is because advice interactions take place between more closely-knit households; for instance, individuals may choose to ask village elders for advice, which would not change based on exposure to urbanization.

The results for the average clustering coefficient and support provide evidence that higher urbanization levels are associated with lower levels of transitivity in networks of social, economic, and advice interactions. The fact that this relationship holds for all types of interactions suggests that Hypothesis 4, where proximity to others allows people to form connections outside of their immediate social circles, is the most plausible mechanism through which urbanization affects clustering.

Tables 6-8 show the results from the distance regressions, which do not include NTL (defined in Equation 11). Estimates are noisy for the relationship between network density and both distance to metropolitan area and Bangalore cohort. The only significant relationship is that villages that are closer to Bangalore have a 15% lower density of advice interactions ($p < 0.1$). However, there is stronger evidence of a relationship between the distance and Bangalore cohort variables and clustering and support. A one percent increase in distance from the nearest metropolitan area is associated with a 0.16% higher average clustering coefficient and a 0.12% higher level of support ($p < 0.1$). Urban proximity therefore seems to be associated with lower levels of transitivity, which is consistent with Hypothesis 4. Additionally, villages that were closer to Bangalore had a 41.2% lower average clustering coefficient and a 16.6% lower support than those closer to Mysore ($p < 0.01$). Results are also robust across all three interaction types. This suggests that being located next to a larger metropolitan area, compared to a smaller one, may have some effect on the degree of network transitivity within the village. Together, these findings suggest that both 'ease of access' and the 'size of access' to cities are important factors that may affect village network structure.

Table 3: OLS regression of network density on NTL and city distance

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0869**	-0.0745*	-0.0772**	-0.0398
	(0.0397)	(0.0369)	(0.0330)	(0.0447)
Log distance to metro area	-0.0930	-0.0209	-0.0815	-0.0893
	(0.0922)	(0.0987)	(0.0800)	(0.118)
Observations	66	66	66	66
R-squared	0.919	0.928	0.933	0.898
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted network density on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4: OLS regression of average clustering coefficient on NTL and city distance

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.115**	-0.0867*	-0.135*	-0.118
	(0.0468)	(0.0492)	(0.0701)	(0.0786)
Log distance to metro area	-0.146	-0.0917	-0.0146	-0.0873
	(0.114)	(0.120)	(0.145)	(0.148)
Observations	66	66	66	66
R-squared	0.806	0.809	0.748	0.748
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the unweighted average clustering coefficient on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: OLS regression of support on NTL and city distance

VARIABLES	(1) All interactions	(2) Economic	(3) Social	(4) Advice
Log NTL	-0.0876**	-0.106**	-0.0953**	-0.0951*
	(0.0324)	(0.0449)	(0.0469)	(0.0525)
Log distance to metro area	-0.0318	0.00945	0.00234	-0.00180
	(0.0621)	(0.107)	(0.118)	(0.112)
Observations	66	66	66	66
R-squared	0.802	0.804	0.776	0.754
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted support on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: OLS regression of network density on distance variables

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	-0.0485	0.0664	-0.0441	0.0127
	(0.107)	(0.110)	(0.0975)	(0.121)
Bangalore cohort	0.0179	-0.109	-0.0218	-0.153*
	(0.0959)	(0.0773)	(0.0788)	(0.0815)
Observations	68	68	68	68
R-squared	0.893	0.909	0.908	0.891
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted network density on whether the village is in the Bangalore cohort and their distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: OLS regression of average clustering coefficient on distance variables

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	0.160*	0.194**	0.390***	0.332**
	(0.0906)	(0.0827)	(0.104)	(0.158)
Bangalore cohort	-0.412***	-0.426***	-0.582***	-0.596***
	(0.0630)	(0.0728)	(0.0794)	(0.121)
Observations	68	68	68	68
R-squared	0.833	0.853	0.830	0.788
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the unweighted average clustering coefficient on whether the village is in the Bangalore cohort and their distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 8: OLS regression of support on distance variables

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	0.118*	0.243***	0.216**	0.275**
	(0.0665)	(0.0879)	(0.100)	(0.119)
Bangalore cohort	-0.166**	-0.325***	-0.305***	-0.353***
	(0.0614)	(0.0726)	(0.0743)	(0.0776)
Observations	68	68	68	68
R-squared	0.727	0.806	0.769	0.739
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted support on whether the village is in the Bangalore cohort and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: OLS regression of network density on NTL and distance variables

VARIABLES	(1) All interactions	(2) Economic	(3) Social	(4) Advice
Log NTL	-0.101**	-0.0635	-0.0816*	-0.0166
	(0.0435)	(0.0421)	(0.0410)	(0.0512)
Log distance to metro area	-0.144	0.0193	-0.0975	-0.00425
	(0.110)	(0.115)	(0.101)	(0.135)
Bangalore cohort	0.0899	-0.0710	0.0283	-0.150
	(0.0965)	(0.0867)	(0.0887)	(0.0989)
Observations	66	66	66	66
R-squared	0.920	0.929	0.933	0.901
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted network density on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 10: OLS regression of average clustering coefficient on NTL and distance variables

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0594	-0.0239	-0.0540	-0.0304
	(0.0421)	(0.0433)	(0.0638)	(0.0679)
Log distance to metro area	0.0584	0.139	0.284*	0.235
	(0.117)	(0.103)	(0.148)	(0.179)
Bangalore cohort	-0.361***	-0.406***	-0.526***	-0.568***
	(0.0878)	(0.0989)	(0.129)	(0.145)
Observations	66	66	66	66
R-squared	0.866	0.864	0.839	0.817
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the unweighted average clustering coefficient on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 11: OLS regression of support on NTL and distance variables

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0703**	-0.0636	-0.0538	-0.0463
	(0.0332)	(0.0414)	(0.0487)	(0.0503)
Log distance to metro area	0.0318	0.167*	0.155	0.177
	(0.0680)	(0.0984)	(0.121)	(0.115)
Bangalore cohort	-0.112*	-0.278***	-0.269***	-0.316***
	(0.0636)	(0.0838)	(0.0980)	(0.0942)
Observations	66	66	66	66
R-squared	0.815	0.843	0.813	0.791
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of unweighted support on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Tables 9-11 show results from regressing network statistics on log NTL, distance to the nearest city, and city cohort, as shown in Equation 12. When including the city cohort variables, the negative relationship between log NTL and network density and support still holds (at $p < 0.05$ for the union network of all interactions), but not the relationship between the average clustering coefficient and log NTL. Villages in the Bangalore cohort have a lower level of clustering and support for all types of interactions, significant at the $p < 0.01$ level. On average, a village that is closer to Bangalore has a 36% lower average clustering coefficient and a 11% lower support than one that is closer to Mysore.

Tables A1-A9 in the appendix present results from the regressions in Equations 10, 11, and 12 using the weighted network statistics (density, average clustering coefficient, and support). It is harder to interpret these results by themselves, since the weighted statistics do not just capture the degree of connectivity in the network but also some aspect of frequency or intensity of links; however, they are useful as a robustness check. The results are generally similar to the regressions with unweighted network statistics. Estimates are noisier for the regression of network statistics on NTL and distance to the nearest city, but the coefficients for NTL remain negative (see Tables A1-A3). Additionally, there is a negative relationship between the 'Bangalore cohort' variable and all three of the weighted network statistics, compared to just clustering and support for the regressions with unweighted network statistics.

Given that all interaction types have significant relationships with NTL and distance variables, it appears that Hypotheses 3 and 4 in Sections 3.3 and 3.4 are the most plausible explanations for how urbanization could affect the density and shape of social networks; urbanization creates opportunities for villagers to interact with outsiders, and with other villagers outside their immediate social circles. However, since the density and transitivity of economic interactions also appears to vary with NTL and distance to the nearest city, I cannot rule out the possibility that Hypotheses 2 and 5 are also correct.

Tables 12-14 show the results from the IV regression described in Section 5.2. There is a negative relationship between network density and log NTL levels for economic and advice interactions ($p < 0.05$). However, the result for the union network is not significant.

Table 12: IV regression of network density on NTL, Bangalore instrument

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	0.00111	-0.144**	-0.0495	-0.187**
	(0.0808)	(0.0664)	(0.0589)	(0.0892)
Observations	66	66	66	66
R-squared	0.909	0.922	0.932	0.873
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log IV regression of unweighted network density on NTL, using Bangalore cohort as an instrument for NTL. NTL refers to log mean NTL values in 2006. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 13: IV regression of average clustering coefficient on NTL, Bangalore instrument

VARIABLES	(1) All interactions	(2) Economic	(3) Social	(4) Advice
Log NTL	-0.468***	-0.484***	-0.650***	-0.674***
	(0.155)	(0.166)	(0.230)	(0.246)
Observations	66	66	66	66
R-squared	0.427	0.460	0.172	0.315
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log IV regression of unweighted average clustering coefficient and NTL, using Bangalore cohort as an instrument for NTL. NTL refers to log mean NTL values in 2006. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 14: IV regression of support on NTL, Bangalore instrument

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.197***	-0.379***	-0.358***	-0.404***
	(0.0648)	(0.118)	(0.121)	(0.137)
Observations	66	66	66	66
R-squared	0.717	0.553	0.536	0.520
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log IV regression of unweighted support and NTL, using Bangalore cohort as an instrument for NTL. NTL refers to log mean NTL values in 2006. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Additionally, higher log NTL levels appear to be very significantly correlated with lower levels of clustering and support; a one percent increase in NTL (explained by proximity to Bangalore) corresponds with a 0.47% decrease in the average clustering coefficient and a 0.2% decrease in the support of all interactions. As discussed in Section 5.2, being in the Bangalore cohort is likely not an ideal instrument for NTL levels. However, these results provide further evidence that some relationship exists between NTL, urban proximity, and network statistics.

Across all of the different regressions, the results for clustering and support appear to be more robust than the results for density. This may indicate that the Jackson et al. 'social quilts' model may be a good source of predictions for network responses to urban development, and therefore warrants further investigation. These findings also complicate the "community lost" hypothesis; while access to urban areas may reduce the density of social networks, it may also facilitate a restructuring of connections so that communities are less clustered and have more "weak ties." This is consistent with previous work, such as by White and Guest (2003).

6.2 Causality

The identification strategy that I use cannot be used to establish causality. In this section I examine whether there is a plausible case for reverse causality (where the structure of a village's social networks affects its rate of urbanization or access to urban areas), or for the presence of an omitted variable.

First, given the relatively small size of the villages (an average of 198 households per village), it is unlikely that household or village-level actors have had much influence on patterns of urbanization in the areas surrounding Bangalore. Furthermore, the majority of villagers who responded to the individual survey work in agricultural or resource extraction activities, with only 13 out of 16,984 (0.077%) respondents employed in a village administration or gram panchayat (council) position. As a result, the ability for communities in the villages to shape policies that would affect urbanization should be limited.

It is possible that social network characteristics could make it easier for villagers to organize for collective action. While I do not have detailed data on political participation in the villages, I use data on the fraction of individuals in each village that have an election card for the following regression:

$$\text{Electioncard}_v = \alpha + \beta_1 \text{statistic}_v + \beta_2 \text{LogMetroDist}_v + \beta_3 \text{BangaloreCohort}_v + \gamma X_v + \epsilon_v \tag{15}$$

where Electioncard_v is the standardized fraction of individuals that have an election card, and statistic_v is a standardized network statistic for the union of all interactions (either density, average clustering, or support), and the other variables kept the same as in Equations 10-12.

Table 15 presents the results for this regression. I do not find any significant relationship between any of the network statistics and the prevalence of election card ownership. This casts some doubt on the hypothesis that network structure enables villagers to mobilize themselves politically. However, rates of election card ownership are relatively high (94% across the entire sample), and owning a card does not mean that the individuals voted or are otherwise politically engaged, so I cannot categorically rule out this explanation.

Another way through which the structure of social networks could have affected urbanization is by encouraging or discouraging the adoption of technology. The structure of a social network could facilitate the spread of urban technology (such as electricity or plumbing), via the diffusion of learning and potential for lending. Related research, that uses the same data as this paper, has demonstrated that certain network characteristics, such as diffusion centrality, can affect the spread of microfinance usage in villages (Banerjee et al., 2013). It is plausible that less clustering (and the existence of more distant ties) may allow for technology to spread beyond small close-knit groups and reach a larger portion of the community. While I do not have detailed data on the prevalence of many urban technologies, I examine whether electricity usage or latrine ownership are related to network characteristics, by using the following regression form:

$$\mathrm{Technology}_v = \alpha + \beta_1 \mathrm{statistic}_v + \beta_2 \mathrm{LogMetroDist}_v + \beta_3 \mathrm{BangaloreCohort}_v + \gamma X_v + \epsilon_v \tag{16}$$

where $\mathrm{Technology}_v$ is either the fraction of households with electricity or latrines in village v, and $\mathrm{statistic}_v$ is either the network density, average clustering coefficient, or support for the union of all interactions for the village.

Tables 16 and 17 show the results of the regressions for electrification and latrine ownership, respectively. There is no significant relationship between any of the network statistics and both of the technology variables, and estimates are generally small. Distance from the nearest city is negatively correlated with rates of latrine ownership. However, this is likely due to structural issues with latrine installation rather than a network-based effect.

Table 15: OLS regression of election card ownership on network statistics

	(1)	(2)	(3)
VARIABLES	Election card	Election card	Election card
Log interaction density, std	0.414		
	(0.289)		
Log metro distance, std	-0.0751	-0.102	-0.136
	(0.134)	(0.153)	(0.137)
Log interaction clustering, std		0.0656	
		(0.249)	
Log interaction support, std			0.217
			(0.164)
Bangalore cohort	-0.395	-0.254	-0.129
	(0.549)	(0.720)	(0.586)
Observations	68	68	68
R-squared	0.629	0.612	0.624
Demographic controls	Yes	Yes	Yes

Results from regressing the frequency of election card ownership on unweighted network statistics from the union of all interactions and distance to the nearest city (Bangalore or Mysore). Distance to metro area is measured in kilometers. All displayed variables are standardized. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 16: OLS regression of electrification rate on network statistics

VARIABLES	(1) Power	(2) Power	(3) Power
Log interaction density, std	0.0821		
	(0.276)		
Log metro distance, std	-0.0608	-0.0427	-0.0852
	(0.141)	(0.149)	(0.148)
Log interaction clustering, std		-0.114	
		(0.225)	
Log interaction support, std			0.104
			(0.176)
Bangalore cohort	-0.160	-0.363	-0.0383
	(0.667)	(0.742)	(0.698)
Observations	68	68	68
R-squared	0.601	0.603	0.603
Demographic controls	Yes	Yes	Yes

Results from regressing the village electrification rate on unweighted network statistics from the union of all interactions and distance to the nearest city (Bangalore or Mysore). Distance to metro area is measured in kilometers. All displayed variables are standardized. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 17: OLS regression of latrine ownership on network statistics

	(1)	(2)	(3)
VARIABLES	Latrine	Latrine	Latrines
Log interaction density, std	-0.181		
	(0.242)		
Log metro distance, std	-0.454***	-0.460***	-0.437***
	(0.121)	(0.125)	(0.125)
Log interaction clustering, std		0.0494	
		(0.204)	
Log interaction support, std			-0.0629
			(0.166)
Bangalore cohort	0.967	1.061*	0.899
	(0.579)	(0.614)	(0.575)
Observations	68	68	68
R-squared	0.688	0.685	0.686
Demographic controls	Yes	Yes	Yes

Results from regressing the rate of latrine ownership on unweighted network statistics from the union of all interactions and distance to the nearest city (Bangalore or Mysore). Distance to metro area is measured in kilometers. All displayed variables are standardized. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

These results cast doubt on the likelihood that network characteristics affect the take-up of new technologies. However, further investigation of the relationship between network characteristics and the diffusion of technology would be welcome.

It is also possible that some omitted variable may exist. For example, a group of geographically close villages may share increased access to a city or social and cultural conditions which could affect both urbanization and social network structure. In an attempt to investigate this possibility, I run the same specifications as in Equations 10, 11, and 12, but add district level fixed effects:

$$\delta_{v,s} = \alpha + \beta_1 \text{LogNTL}_v + \beta_2 \text{LogMetroDist}_v + \gamma_1 X_v + \gamma_2 D + \epsilon_{v,s} \tag{17}$$

$$\delta_{v,s} = \alpha + \beta_1 \text{LogMetroDist}_v + \beta_2 \text{BangaloreCohort}_v + \gamma_1 X_v + \gamma_2 D + \epsilon_{v,s} \tag{18}$$

$$\delta_{v,s} = \alpha + \beta_1 \text{LogNTL}_v + \beta_2 \text{LogMetroDist}_v + \beta_3 \text{BangaloreCohort}_v + \gamma_1 X_v + \gamma_2 D + \epsilon_{v,s} \tag{19}$$

where D is a vector of dummy variables corresponding to the districts in the data. Tables 18-20 present results for the above specifications. Some of the significant results drop out when including the district level controls. When regressing network statistics on NTL and distance to the nearest metropolitan area, there is now a negative relationship between distance to metropolitan area and network density. There is still a negative relationship between log NTL levels and the average clustering coefficient ($p < 0.1$). When including the Bangalore cohort variable, the negative relationship between distance to the metropolitan area and network density also appears. However, there is a negative relationship between the average clustering coefficient and both NTL and the Bangalore cohort. For the regression with just the distance and Bangalore cohort variables, villages closer to Bangalore appear to have higher network density. In general, it appears that villages that are closer to a city have higher densities. On the other hand, being closer to Bangalore is associated with a lower average clustering coefficient. The reversal in signs for some of the estimates may be because the district controls are capturing some aspect of distance as well.

Additionally, given the relatively small sample of villages, group sizes may not be large enough to ensure unbiased estimates when using fixed effects (Maas and Hox, 2005). It is also unclear what types of geography-based variation among village clusters is associated with urbanization and therefore should have an effect on NTL (e.g. proximity to neighboring towns and other urban centers) and what variation is due to factors that are not related to urbanization (e.g. district-level regulations or cultural traditions). Significant heterogeneity within districts could also bias the results when using these specifications. Future work should investigate whether other instruments for urbanization could be used, in order to capture the true effect of urbanization on social networks.

Table 18: OLS regression of network statistics on NTL, district fixed effects

VARIABLES	(1) Density	(2) Clustering	(3) Support
Log NTL	-0.0858	-0.0770*	-0.0660
	(0.0515)	(0.0445)	(0.0415)
Log distance to metro area	-0.321**	0.00597	-0.0317
	(0.119)	(0.126)	(0.0846)
Observations	66	66	66
R-squared	0.935	0.877	0.826
Demographic controls	Yes	Yes	Yes
District FEs	Yes	Yes	Yes

Results from the log-log regression of unweighted network statistics on NTL and distance to the nearest city, with fixed effects at the district level. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table 19: OLS regression of network statistics on distance variables, district fixed effects

	(1)	(2)	(3)
VARIABLES	Density	Clustering	Support
Log distance to metro area	-0.264*	0.101	0.0266
	(0.130)	(0.110)	(0.0940)
Bangalore cohort	0.460***	-0.467**	0.0636
	(0.165)	(0.200)	(0.146)
Observations	68	68	68
R-squared	0.917	0.841	0.751
Demographic controls	Yes	Yes	Yes
District FEs	Yes	Yes	Yes

Results from the log-log regression of unweighted network statistics on whether the village is in the Bangalore cohort and distance to the nearest city, with fixed effects at the district level. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 20: OLS regression of network statistics on NTL and distance variables, district fixed effects

	(1)	(2)	(3)
VARIABLES	Density	Clustering	Support
Log NTL	-0.0858	-0.0770*	-0.0660
	(0.0515)	(0.0445)	(0.0415)
Log distance to metro area	-0.321**	0.00597	-0.0317
	(0.119)	(0.126)	(0.0846)
Bangalore cohort	0.291	-0.565**	-0.0520
	(0.193)	(0.235)	(0.172)
Observations	66	66	66
R-squared	0.935	0.877	0.826
Demographic controls	Yes	Yes	Yes
District FEs	Yes	Yes	Yes

Results from the log-log regression of unweighted network statistics on NTL, city cohort, and distance to the nearest city, with fixed effects at the district level. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

7 Conclusion

Network data from 66 villages in Karnataka, India provide evidence that there is a negative relationship between urbanization, as measured by nighttime lights, and the density of networks for social and economic interactions. There is also a negative relationship between NTL levels and the average clustering coefficient and support in village networks.

I find some evidence that villages closer to urban centers have lower levels of clustering and support. I also discover that villages that are closer to Bangalore have a lower level of clustering and support, compared to those that are closer to Mysore. These results provide evidence that both urban development and proximity to major urban areas may have a negative effect on network density and transitivity. When using the 'Bangalore cohort' as an instrument for NTL, I find that NTL (as influenced by proximity to a large metropolitan area) is also negatively correlated with density and transitivity of networks for all types of interactions.

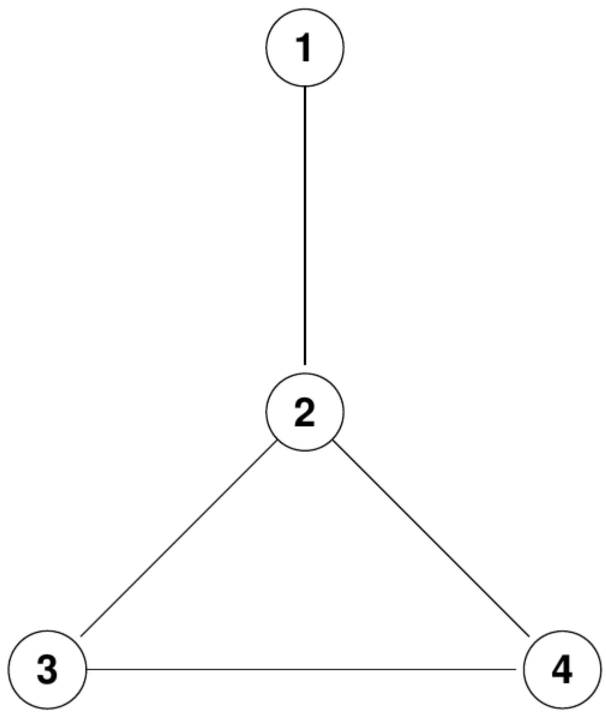
These results are consistent with the theory that urbanization or access to urban areas facilitates the formation of connections outside of the community, at the expense of ties within the community. They are also consistent with the theory that urbanization facilitates the formation of "weak ties" and leads to less clustering by giving more opportunities for individuals to encounter others outside of their immediate social circles. These findings also add nuance to the debate over the "community lost" hypothesis, by indicating that urban proximity is not just associated with a simple decrease in network density, but a broader change in the structure and shape of networks. The empirical results for clustering and support appear to be more robust than the result for network density, which suggests that the 'social quilts' model originally developed by Jackson et al. may be a useful theoretical basis for investigating how households' access to urban areas could have widespread effects on network structure. More work remains to be done on how to identify and quantify how urban development affects these specific mechanisms.

While I present some results that challenge the existence of reverse causality, the data that I use cannot be used to conclusively rule it out. Future work should attempt to identify methods to effectively determine whether urbanization affects social networks, or vice versa. For example, future research could test how social networks change when formal services (e.g. microfinance loans or cash transfers) or communications technology (e.g. mobile phones) are introduced to community members. While these experimental treatments are not explicitly tied to urbanization, they could help reveal how access to formal institutions or easier access to external connections could affect networks.

In 1992, the United Nations declared that “the future is urban” (United Nations Population Division, 1992). This prediction appears to be borne out by the experiences of developing countries over the past two and a half decades. The continued urbanization of these countries will not just affect city dwellers, but also rural residents who will increasingly come to interact with and rely upon urban communities and institutions. As such, the importance of investigating the relationship between urbanization, urban access, and social networks will only increase.

8 Appendix

Figure A1: Network statistics example



A simple network with four nodes and four edges is displayed above. The density of the network is equal to $4/\binom{4}{2} = 0.6\bar{6}$. The average clustering coefficient of the network is equal to $\frac{0+1+1+1/3}{4} = 0.58\bar{3}$. The support of the network is equal to $\frac{3}{4} = 0.75$.

Table A1: Relationship between NTL, city distance, and weighted network density

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0469	-0.0610*	-0.0793**	-0.00656
	(0.0352)	(0.0337)	(0.0327)	(0.0506)
Log distance to metro area	-0.0555	0.00623	-0.0680	-0.113
	(0.0985)	(0.0981)	(0.0829)	(0.139)
Observations	66	66	66	66
R-squared	0.931	0.934	0.935	0.883
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of weighted network density on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table A2: Relationship between NTL, city distance, and weighted clustering coefficient

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0880	-0.0681	-0.138*	-0.113
	(0.0639)	(0.0597)	(0.0738)	(0.0850)
Log distance to metro area	-0.120	-0.0763	-0.00673	-0.126
	(0.146)	(0.134)	(0.159)	(0.174)
Observations	66	66	66	66
R-squared	0.755	0.790	0.721	0.744
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted average clustering coefficient on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Relationship between NTL, city distance, and weighted support

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0553	-0.0948*	-0.0962*	-0.0779
	(0.0396)	(0.0491)	(0.0491)	(0.0604)
Log distance to metro area	-0.0214	0.0237	0.00650	-0.00836
	(0.0974)	(0.120)	(0.126)	(0.135)
Observations	66	66	66	66
R-squared	0.754	0.790	0.732	0.758
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted support on NTL and distance to the nearest city (Bangalore or Mysore). NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A4: Relationship between distance variables and weighted network density

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	0.0373	0.130	-0.00600	-0.0264
	(0.102)	(0.0998)	(0.100)	(0.134)
Bangalore cohort	-0.153**	-0.190***	-0.0596	-0.191**
	(0.0640)	(0.0677)	(0.0732)	(0.0817)
Observations	68	68	68	68
R-squared	0.924	0.929	0.913	0.885
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of weighted network density on whether the village is in the Bangalore cohort and their distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A5: Relationship between distance variables and weighted clustering coefficient

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	0.257***	0.254***	0.433***	0.310*
	(0.0884)	(0.0836)	(0.115)	(0.170)
Bangalore cohort	-0.605***	-0.516***	-0.633***	-0.643***
	(0.0622)	(0.0737)	(0.0821)	(0.124)
Observations	68	68	68	68
R-squared	0.876	0.865	0.817	0.796
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted average clustering coefficient on whether the village is in the Bangalore cohort and their distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table A6: Relationship between distance variables and weighted support

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log distance to metro area	0.188***	0.306***	0.250**	0.272**
	(0.0658)	(0.0838)	(0.103)	(0.119)
Bangalore cohort	-0.346***	-0.413***	-0.353***	-0.407***
	(0.0450)	(0.0658)	(0.0733)	(0.0781)
Observations	68	68	68	68
R-squared	0.837	0.843	0.751	0.784
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted support on whether the village is in the Bangalore cohort and their distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** p<0.01, ** p<0.05, * p<0.1.

Table A7: Relationship between NTL, distance variables, and weighted network density

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.0243	-0.0343	-0.0776*	0.0285
	(0.0414)	(0.0379)	(0.0406)	(0.0566)
Log distance to metro area	0.0274	0.104	-0.0616	0.0160
	(0.113)	(0.109)	(0.105)	(0.153)
Bangalore cohort	-0.146*	-0.173**	-0.0112	-0.227**
	(0.0836)	(0.0842)	(0.0872)	(0.104)
Observations	66	66	66	66
R-squared	0.935	0.940	0.935	0.891
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of weighted network density on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A8: Relationship between NTL, distance variables, and weighted clustering coefficient

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	0.00534	0.0123	-0.0491	-0.0171
	(0.0452)	(0.0476)	(0.0662)	(0.0730)
Log distance to metro area	0.223*	0.219**	0.321*	0.227
	(0.117)	(0.105)	(0.160)	(0.201)
Bangalore cohort	-0.605***	-0.521***	-0.579***	-0.624***
	(0.0939)	(0.104)	(0.134)	(0.151)
Observations	66	66	66	66
R-squared	0.881	0.873	0.827	0.814
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted average clustering coefficient on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A9: Relationship between NTL, distance variables, and weighted support

	(1)	(2)	(3)	(4)
VARIABLES	All interactions	Economic	Social	Advice
Log NTL	-0.00155	-0.0357	-0.0469	-0.0173
	(0.0288)	(0.0392)	(0.0477)	(0.0534)
Log distance to metro area	0.176**	0.241**	0.187	0.214
	(0.0747)	(0.0983)	(0.121)	(0.129)
Bangalore cohort	-0.348***	-0.383***	-0.319***	-0.393***
	(0.0614)	(0.0847)	(0.0977)	(0.0957)
Observations	66	66	66	66
R-squared	0.851	0.860	0.787	0.804
Demographic controls	Yes	Yes	Yes	Yes

Results from the log-log regression of the weighted support on NTL, city cohort, and distance to the nearest city. NTL refers to log mean NTL values in 2006. Distance to metro area is measured in kilometers. All regressions include demographic controls for religion, caste, village size, mean household and home characteristics, distance to district seat, and village amenities. Robust standard errors are displayed. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

References

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013). The Diffusion of Microfinance. *Science 341*(6144).

Breza, E. (2016). Field Experiments, Social Networks, and Development. In *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Donner, J. (2006). The Use of Mobile Phones by Microentrepreneurs in Kigali, Rwanda: Changes to Social and Business Networks. *Information Technologies & International Development 3*(2).

Fischer, C. (1982). *To Dwell Among Friends: Personal Networks in Town and City*. University of Chicago Press.

Gaba, K., B. Min, A. Thakker, and C. Elvidge (2016). nightlights.io: Twenty Years of India Lights.

Gilchrist, A. (2009). *The Well-connected Community: A Networking Approach to Community Development*. Policy Press.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology 78*(6), 1360–1380.

Henderson, J., A. Storeygard, and D. Weil (2012). Measuring Economic Growth from Outer Space. *American Economic Review 102*, 994–1028.

Israel, B. (1985). Social Networks and Social Support: Implications for Natural Helper and Community Level Interventions. *Health Education Quarterly 12*(1).

Jackson, M., T. Rodriguez-Barraquer, and X. Tan (2012). Social Capital and Social Quilts: Network Patterns of Favor Exchange. *American Economic Review 102*(5), 1857–1897.

Jackson, M. O. (2011). An Overview of Social Networks and Economic Applications. In *Handbook of Social Economics*. North-Holland.

Jacobson, A. (2007). Connective Power: Solar Electrification and Social Change in Kenya. *World Development 35*(1), 144–162.

Jean, N., M. Burke, M. Xie, W. Davis, D. Lobell, and S. Ermon (2016). Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science 353*, 790–794.

Kavanaugh, A., D. Reese, J. Carroll, and M. Rosson (2005). Weak Ties in Networked Communities. *The Information Society 21*(2), 119–131.

Leavy, R. (1983). Social Support and Psychological Disorder: A Review. *Journal of Community Psychology 11*(1).

Levine, S. and R. Kurzban (2006). Explaining Clustering in Social Networks: Towards an Evolutionary Theory of Cascading Benefits. *Managerial and Decision Economics 27*(2-3), 173–187.

Maas, C. and J. Hox (2005). Sufficient Sample Sizes for Multilevel Modeling. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences 1*(3), 86–92.

Mellander, C., J. Lobo, K. Stolarick, and Z. Matheson (2015). Night-Time Light Data: A Good Proxy Measure for Economic Activity? *PLoS One 10*(10).

Miguel, E., P. Gertler, and D. Levine (2006). Does Industrialization Build or Destroy Social Networks? *Economic Development and Cultural Change 54*(2), 287–317.

Munshi, K. and M. Rosenzweig (2016). Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap. *American Economic Review 106*(1).

Nordhaus, W. and X. Chen (2014). A Sharper Image? Estimates of the Precision of Nighttime Lights as a Proxy for Economic Statistics. *Journal of Economic Geography 15*(1), 217–246.

Office of the Registrar General, I. (2001). Code List for Land Regions (State, District, Sub-District, Town, Village).

Proville, J., D. Zavala-Araiza, and G. Wagner (2017). Night-time Lights: A Global, Long Term Look at Links to Socio-economic Trends. *PLoS One 12*(3).

Saramäki, J., M. Kivelä, J. Onnela, K. Kaski, and J. Kertesz (2007). Generalizations of the Clustering Coefficient to Weighted Complex Networks. *Physical Review E 75*(2).

Sato, Y. and Y. Zenou (2015). How Urbanization Affects Employment and Social Interactions. *European Economic Review 75*, 131–155.

United Nations Department of Economic and Social Affairs, Population Division (2014). World Urbanization Prospects: The 2014 Revision. CD-ROM.

United Nations Population Division (1992). The Future is Urban.

Villa, J. (2014). Social Transfers and Growth: The Missing Evidence from Luminosity Data. Technical report, WIDER Working Paper.

White, K. and A. Guest (2003). Community Lost or Transformed? Urbanization and Social Ties. *City & Community 2*(3), 239–259.

World Bank (2013). Developing Countries Need to Harness Urbanization to Achieve the MDGs: IMF-World Bank report.